Psychophysical and Physiological Evidence for Viewer-centered Object Representations in the Primate

N. K. Logothetis and J. Pauls

Division of Neuroscience, Baylor College of Medicine, Houston, Texas 77030

A key question concerning the perception of 3D objects is the spatial reference frame used by the brain to represent them. The celerity of the recognition process could be explained by the visual system's ability to quickly transform stored models of familiar 3D objects, or by its ability to specify the relationship among viewpoint-invariant features or volumetric primitives that can be used to accomplish a structural description of an image. Alternatively, viewpoint-invariant recognition could be realized by a system endowed with the ability to perform an interpolation between a set of stored 2D templates, created for each *experienced* viewpoint.

In the present study we set out to examine the nature of object representation in the primate in combined psychophysical-electrophysiological experiments. Monkeys were trained to recognize novel objects from a given viewpoint and subsequently were tested for their ability to generalize recognition for views generated by mathematically rotating the objects around any arbitrary axis.

The perception of 3D novel objects was found to be a function of the object's retinal projection at the time of the recognition encounter. Recognition became increasingly difficult for the monkeys as the stimulus was rotated away from its familiar attitude. The generalization field for novel wire-like and spheroidal objects extended to about $\pm 40°$ around an experienced viewpoint. When the animals were trained with as few as three views of the object, 120° apart, they could often interpolate recognition for all views resulting from rotations around the same axis.

Recordings from inferotemporal cortex during the psychophysical testing showed a number of neurons with remarkable selectivity for individual views of those objects that the monkey had learned to recognize. Plotting the response of neurons as a function of rotation angle revealed systematic view-tuning curves for rotations in depth. A small percentage of the view-selective cells responded strongly for a particular view and its mirror-symmetrical view. For some of the tested objects, different neurons were found to be tuned to different views of the same object; the peaks of the view-tuning curves were 40–50° apart. Neurons were also found that responded to the sight of unfamiliar objects or distractors. Such cells, however, gave nonspecific responses to a variety of other patterns presented while the monkey performed a simple fixation task.

The ability to recognize objects is, in many ways, a remarkable accomplishment for biological systems. Familiar objects can be readily recognized from their shape, color, or texture, and even when partially occluded, they can be "surmised" based on some sort of reasoning processes driven by contextual information.

In striking contrast, recognition has proved to be very difficult to achieve in artificial systems. This is partly because we know very little about what constitutes an object. There is nothing special about objects, at least not in the way they are represented in the input of the visual system. The shape of an object cannot always be determined by a predictable combination of visual primitives. Any given 2D image can be parsed into an arbitrary set of objects, each of which can be recursively decomposed into smaller objects. Moreover, what we consider to be an object depends on the visual input, yet it is also determined by the task at hand. The neural representation of objects is a mystery even when considering simple geometrical objects, such as a cube, a cone, or a cylinder, seen in isolation.

Most theories of object recognition posit that the visual system stores a representation of an object and that recognition occurs when the stored representation is matched to its corresponding sensory representation generated from the viewed object. This assumption, however, raises two obvious questions: what are these representations and how is matching achieved? Are objects represented explicitly in the visual cortex, say, by the activation of a set of selective neurons on the top of a visual processing hierarchy, or are they implicitly represented by the activity of large populations of cells, each of which might have little selectivity for any of the complex features of an object? Furthermore, are the stored representations object-centered, 3D descriptions of the objects, or are they viewer-centered descriptions corresponding to 2D perspective views?

In attempting to determine a possible reference system for object representation it is useful to consider first the different taxonomic levels of abstraction at which object recognition can occur. Objects are usually recognized first at the *basic level* (Rosch et al., 1976). This level refers to the initial classification of individual visual entities, such as a piano or a horse. When detailed distinctions between objects of the same category are required, for instance, when discriminating different horse breeds, then recognition is said to occur at the *subordinate* level. Subordinate categories share a great number of object attributes with other subordinate categories, and to a large extent have similar shape (Rosch, 1975; Rosch et al., 1976; Jolicoeur et al., 1984). Nonetheless, atypical exemplars of basic-level categories can be occasionally classified faster at the subordinate than at the basic level (Jolicoeur et al., 1984). For example, the image of a penguin is more likely to be initially identified as penguin before it is determined to be a bird. Since the notion of "basic level" was defined for entire categories based on the degree of inclusiveness of perceptual and functional attributes (Rosch et al., 1976), the term "entry point level" was coined by Jolicoeur et al. (1984) to denote the abstraction level at which stored information can be fastest accessed, regardless of what the taxonomic level may be. Interestingly, clinical studies indicate that recognition at different categorization levels may involve different neural circuitry (Tranel et al., 1988; Damasio, 1990).

Object representations may vary for different classification tasks. Object-centered representations imply the existence of a complete 3D description of an object (Ullman, 1989), or of a structural description of the image specifying the relationships among viewpoint-invariant volumetric primitives (Marr, 1982; Biederman, 1987). A prediction of object-centered representations is uniform performance (in terms of error rate) regardless of viewpoint, provided that the information needed to access the correct model is present in the image, that is, provided that the image is not the result of an *accidental*

alignment of the eye and the object (Biederman, 1987). For example, imagine a flat disk seen from the side. If the gaze is perfectly parallel to the disk's surface (a rare event indeed), the disk's profile will be simply a thick, straight line.

Viewer-centered representations model 3D objects as a set of 2D views, or templates, and recognition consists of matching image features against the views in this set. Viewpoint invariance in systems based on viewer-centered representations may, therefore, rely on object familiarity, and performances may be progressively worse for views that are far from those previously experienced.

When tested against human behavior, the verification of the predicted performance of object-centered representations appears to depend on the object classification level. While humans can recognize familiar objects or objects of the "entry point level" in a viewpoint-independent fashion (Biederman, 1987), they fail to do so at the subordinate level, at which fine, shape-based discriminations are required for identifying an individual entity (Rock et al., 1981; Rock and DiVita, 1987; Tarr and Pinker, 1989, 1990; Buelthoff and Edelman, 1992; Edelman and Buelthoff, 1992; Logothetis et al., 1994).

Viewer-centered representations, on the other hand, can explain human recognition performance at any taxonomic level, but they have been often considered implausible because of the amount of memory required to store all views that a 3D object can generate when viewed from different distances and orientations. Yet, recent theoretical work indicates that a viewer-centered representation system may accomplish viewpoint invariance relying on a small number of 2D views. For example, it has been shown that under conditions of orthographic projection all possible views of an object can be expressed simply as the linear combination of as few as three distinct 2D views, given that the same features remain visible in all three views (Ullman and Basri, 1991). The model of linear combinations of views, however, relies only on geometrical features, and fails to predict human behavior for subordinate level recognition (Rock et al., 1981; Rock and DiVita, 1987; Logothetis et al., 1994).

Alternatively, generalization could be accomplished by nonlinear interpolation among stored orthographic or perspective views, which can be determined on the basis of geometric features or material properties of the object. Indeed, it has been shown that a simple network can achieve viewpoint invariance by interpolating between a small number of stored views (Poggio and Edelman, 1990). Computationally, such a network uses a small set of sparse data corresponding to an object's training views to synthesize an approximation to a multivariate function representing the object. The approximation technique, known as generalized radial basis functions (GRBFs), is mathematically equivalent to a multilayer network (Poggio and Girosi, 1990). A special case of such a network is that of the radial basis functions (RBFs), which can be conceived of as "hidden-layer" units, the activity of which is a radial function of the disparity between a novel view and a template stored in the unit's memory (see Vetter et al., 1995). Such an interpolation-based network makes psychophysical predictions (Poggio, 1990) that have been supported by human psychophysical work (Rock et al., 1981; Rock and DiVita, 1987) and can be directly tested against monkey recognition performance. It also predicts that learning a novel object from example-views may rely on the formation of new, bell-shaped receptive fields tuned to the trained views. Combined activity of such units could then be one possible mechanism for achieving view-independent recognition. The inferior temporal cortex (IT) of monkeys trained to recognize novel 3D objects is an obvious brain area in which to explore the existence of such view selectivity for novel objects. The IT [roughly coextensive with area TE and TEO (Von Bonin and Bailey, 1947; see also Fig. 1A)] has been shown to be essential for object vision. Patients undergoing unilateral anterior temporal lobectomy for the relief of focal epilepsy exhibit specific visuoperceptual deficits (Milner, 1958, 1968, 1980; Kimura, 1963; Lansdell, 1968) and significant impairment in remembering complex visual patterns (Milner, 1968, 1980; Kimura, 1963; Taylor, 1969). Similarly, lesions to area TEO and to area TE yield disruptions of pattern perception and recognition (Iwai and Mishkin, 1969; Gross, 1973) while leaving thresholds for low-level visual tasks unaffected.

Electrophysiological research has provided further evidence regarding the role of IT in object recognition. Charles Gross and his colleagues were the first to obtain visually evoked responses in this area using both macro- and microelectrodes in anesthetized and unanesthetized monkeys (Vaughan and Gross, 1966; Gross et al., 1967, 1969; Gerstein et al., 1968).

The posterior part of IT (approximately area TEO) has a rough visuotopy while the anterior part is not visuotopically organized. IT neurons have large, ipsilateral, contralateral, or bilateral receptive fields that almost always include the fovea, and most are selective for stimulus attributes, such as size, shape, color, orientation, or direction of movement (Gross et al., 1969, 1972). The response of the neurons to stimuli presented at the fovea tends to be more vigorous than elsewhere in the receptive field. Some neurons respond best to complex shapes, including hands, trees, and human or monkey faces. A large number of investigations confirmed and extended these initial findings. As it stands, we know that IT neurons respond only to visual stimuli, which can be bars or spots of light, simple geometrical entities, or complex patterns (Gross et al., 1967, 1972; Desimone and Gross, 1979; Desimone et al., 1984). Selectivity has been also reported for material properties of objects such as texture or color (Mikami and Kubota, 1980; Desimone et al., 1984; Komatsu et al., 1992). In general, neurons responsive to similar features seem to be organized in columns that span most of the IT cortical layers (Gochin et al., 1991; Tanaka et al., 1991; Fujita et al., 1992). Neurons recorded on the same electrode tend to have similar stimulus selectivity and are more likely to show functional interactions than those recorded on different electrodes spaced farther apart (Gochin et al., 1991).

Shape is clearly a prevailing stimulus feature in the IT cortex. IT neurons respond in a selective manner to the shape of various natural or man-made objects (Desimone et al., 1984), parametric shape descriptors (Schwartz et al., 1983), or 2D functions that can be made to synthesize any visual pattern to a required degree of accuracy (Richmond et al., 1987). Pattern-tuned neurons maintain their selectivity even when the stimuli are defined by visual cues other than luminance or color contrast, such as motion or texture differences (Sry et al., 1993). The shape selectivity of IT neurons is also maintained over changes in stimulus position and size (Gross and Mishkin, 1977). Although such changes usually alter the absolute firing rate of the neurons (Schwartz et al., 1983), the relative preference for a particular stimulus is maintained; to this extent IT neurons exhibit size and position invariance.

The most striking class of highly selective cells in IT are those responding to the sight of faces (Bruce et al., 1981; Perrett et al., 1982; Hasselmo et al., 1986; Yamane et al., 1987). Face neurons have been recorded in both adult and infant monkeys (Rodman et al., 1993). They are usually found deep in the lower bank and fundus of the superior temporal sulcus (STS) and in the polysensory area located dorsal to IT in the fundus and upper bank of the STS (Perrett et al., 1982; Desimone et al., 1984). Most face-selective neurons are 2–10 times more sensitive to faces than to other complex patterns, simple geometrical stimuli, or real 3D objects (Perrett et al., 1979,

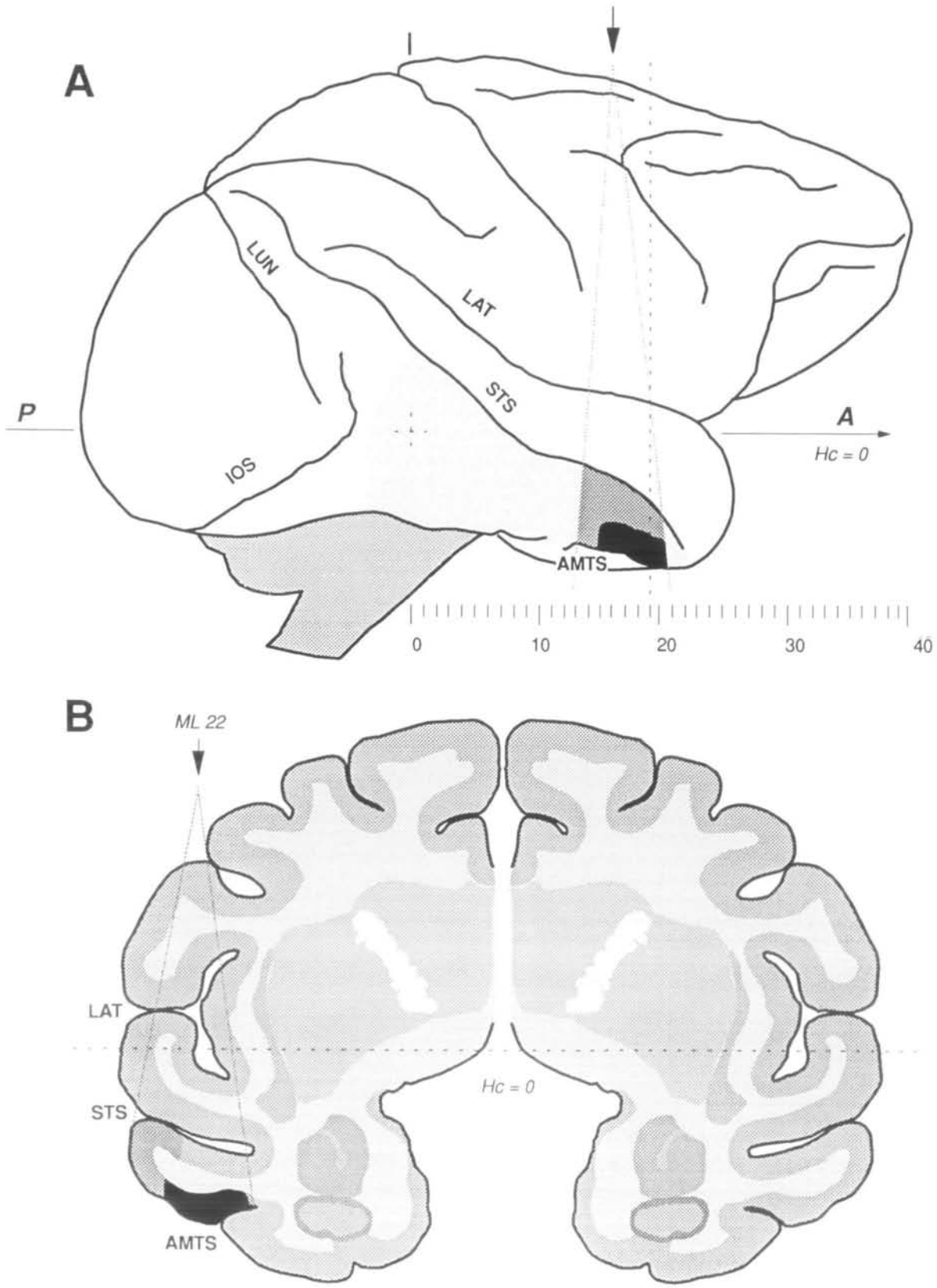


Figure 1. Anatomical location of the recording site estimated from the stereotaxic coordinates. *A*, Lateral view of a macaque brain. Labeled are the lunate sulcus (*LUN*), the inferior occipital sulcus (*IOS*), the lateral fissure (*LAT*), the superior temporal sulcus (*STS*), and the anterior medial temporal sulcus (*AMTS*). The *dashed cross* represents Horsley-Clarke zero. The *dashed vertical line* marks 19 mm anterior, which is the location of the frontal section presented in *B*. The *thin gray lines* (indicated by the *vertical arrow* in both *A* and *B*) represent the roughly conical volume accessible for recording using a ball-and-socket electrode drive. The anterior-posterior and medial-lateral extent of the primary recording site (*dark gray* and *black* in *A* and *B*) were from 14 to 21 mm anterior and 16 to 24 mm lateral. The *black area* in both drawings denotes the estimated location in which the data presented in this article were collected. *B*, Drawing of a frontal section showing the recording site as estimated from the stereotaxic coordinates and the depth measurements during the recordings. We concentrated our recording in the dorsal lip, dorsal bank, and fundus of the AMTS

1982). Presenting different views or parts of a face in isolation revealed that the neurons may respond selectively to face views (Desimone et al., 1984; Perrett et al., 1985), features, or subsets of features (Perrett et al., 1989; Young and Yamane, 1992a,b). Thus, face neurons do have properties reminiscent of an RBF network. Is such view selectivity specific to faces? Could one expect to find neurons in this area that are tuned to views (or parts thereof) of *nonsensical* objects that the monkey just learns to recognize?

To address these questions we have first examined whether the performance of monkeys is view invariant or is a function of the disparity between a novel view and the view that the animal experienced in short training sessions. Specifically, we trained monkeys to recognize objects presented from one view and subsequently tested their ability to generalize recognition for views generated by mathematically rotating the objects around arbitrary axes. We then examined whether neurons in IT respond selectively to novel objects that the monkey learns to recognize, and whether or not those cells that might be selective show view-dependent activity.

Brief reports of these experiments have been published previously (Logothetis et al., 1992, 1993).

Materials and Methods

Subjects and Surgical Procedures

Three juvenile rhesus monkeys (*Macaca mulatta*) weighing 7–9 kg were used in these experiments. The animals were cared for in accordance with the National Institutes of Health Guide and the guidelines of the Animal Protocol Review Committee of the Baylor College of Medicine. Each monkey was first trained to sit in a primate chair. After familiarization with the laboratory environment and the experimenter, the animal underwent a surgery for the placement of the head restraint post and the scleral search eye coil (Judge et al., 1980). The surgical procedure was carried out under strictly aseptic conditions using isoflurane anesthesia (3 5% induction and 1.2–1.5% maintenance, at 0.8 liter/min oxygen). Throughout the surgical procedure the heart rate, blood pressure, and respiration were monitored constantly and recorded every 15 min. Body temperature was kept at 37.4°C using a heating pad. Postoperatively, the monkey was administered an opioid analgesic (buprenorphine hydrochloride, 0.02 mg/kg, i.m.) every 6 hr for 1 d. Tylenol (10 mg/kg) and antibiotics (Tribrissen, 30 mg/kg) were given to the animal for 3–5 d after the operation. At the end of the training period another sterile surgery was performed to implant a ball-and-socket chamber for the electrophysiological recordings.

Visual Stimuli

The stimuli, examples of which are shown in Figure 2, were similar to those used by Edelman and Buelthoff (1992) in human psychophysical experiments (Fig. 2*a–d*), or a variety of other 2D or 3D patterns, including commonplace objects, scenes, or body parts (Fig 2*e–h*). Some of the objects were generated mathematically and others were simply digitized using a standard (RS170) CCD camera. All stimuli were presented on a monitor situated at a distance of 114 cm from the animal.

The view generated by the selection of the appropriate parameters was arbitrarily named the *zero view* of the object. The viewpoint coordinates of the observer with respect to the object were defined as the longitude and the latitude of the eye on an imaginary sphere centered on the object (Fig. 3*A*). We have used a right-handed coordinate system for the object transformations.

All objects were rendered using a visualization system (Application Visualization System, Stardent Computer Inc.) on a DEC5000 work station, and transferred to an IBM-compatible AT486 computer (GATEWAY 486/33C). Display of the images was accomplished by means of a graphics card (Number Nine Computer, SGT board) of 640 × 480 resolution, at 60 Hz refresh rate.

Animal Training

The monkeys were trained to recognize objects irrespective of position or orientation. They were first allowed to inspect an object, the *target,* presented from a given viewpoint, and subsequently were tested for recognizing views of the same object generated by ± 10° to ± 180° rotations around the vertical axis. In some experiments the animals were tested for recognizing views around either the vertical or the horizontal axis, and in some others the animals were tested for views around all four axes. The images were presented sequentially, with the target views dispersed among a large number of other objects, the *distractors.* Two levers were attached to the front panel of the chair, the reinforcement was contingent upon pressing the right lever each time the target was presented. Pressing the left lever was required upon presentation of a distractor. Correct responses were rewarded with fruit juice.

Initially, the animals were trained to recognize the target's zero view among a large set of distractors, and subsequently were trained to recognize additional target views resulting from progressively larger rotations around one axis. After the monkey learned to recognize a given object from any viewpoint in the range of ± 90°, the procedure was repeated with a new object. On average, the monkey required 4 months of training to learn to generalize the task across different objects of the class, and about 6 months to generalize across different object classes. Within an object class the similarity of the targets to the distractors was gradually increased, and in the final stage of the experiments distractor wire-objects were generated by

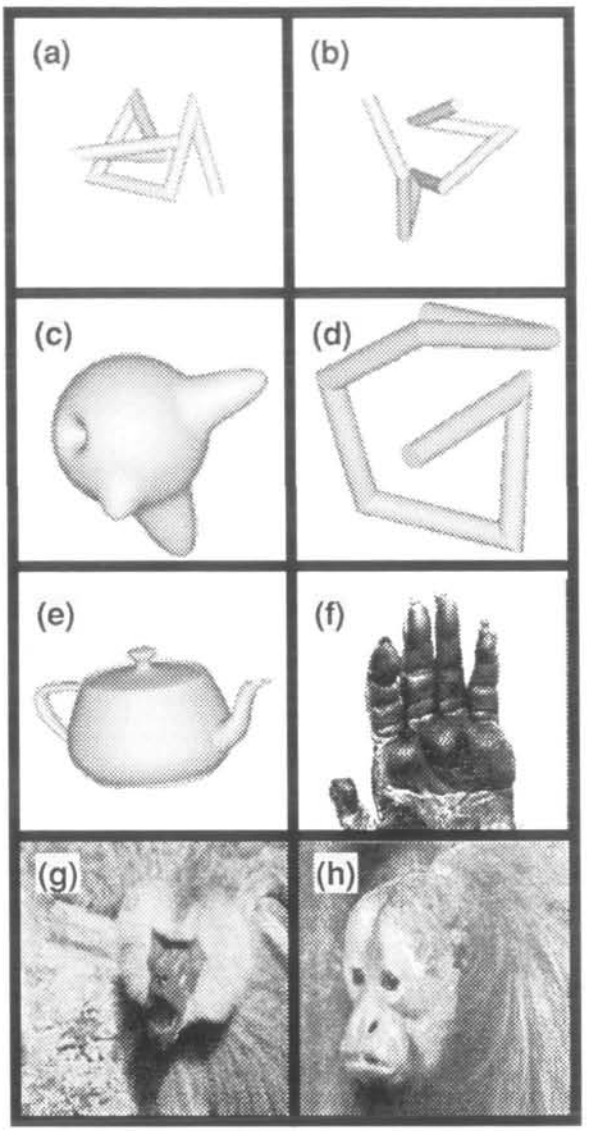


Figure 2. Examples of stimuli used in experiments on object recognition. Wire-like, amoeboid, and common-type objects were created mathematically and rendered by a computer. Pictures of various natural objects such as hands and faces were digitized using a camera and a standard PC-based frame-grabber.

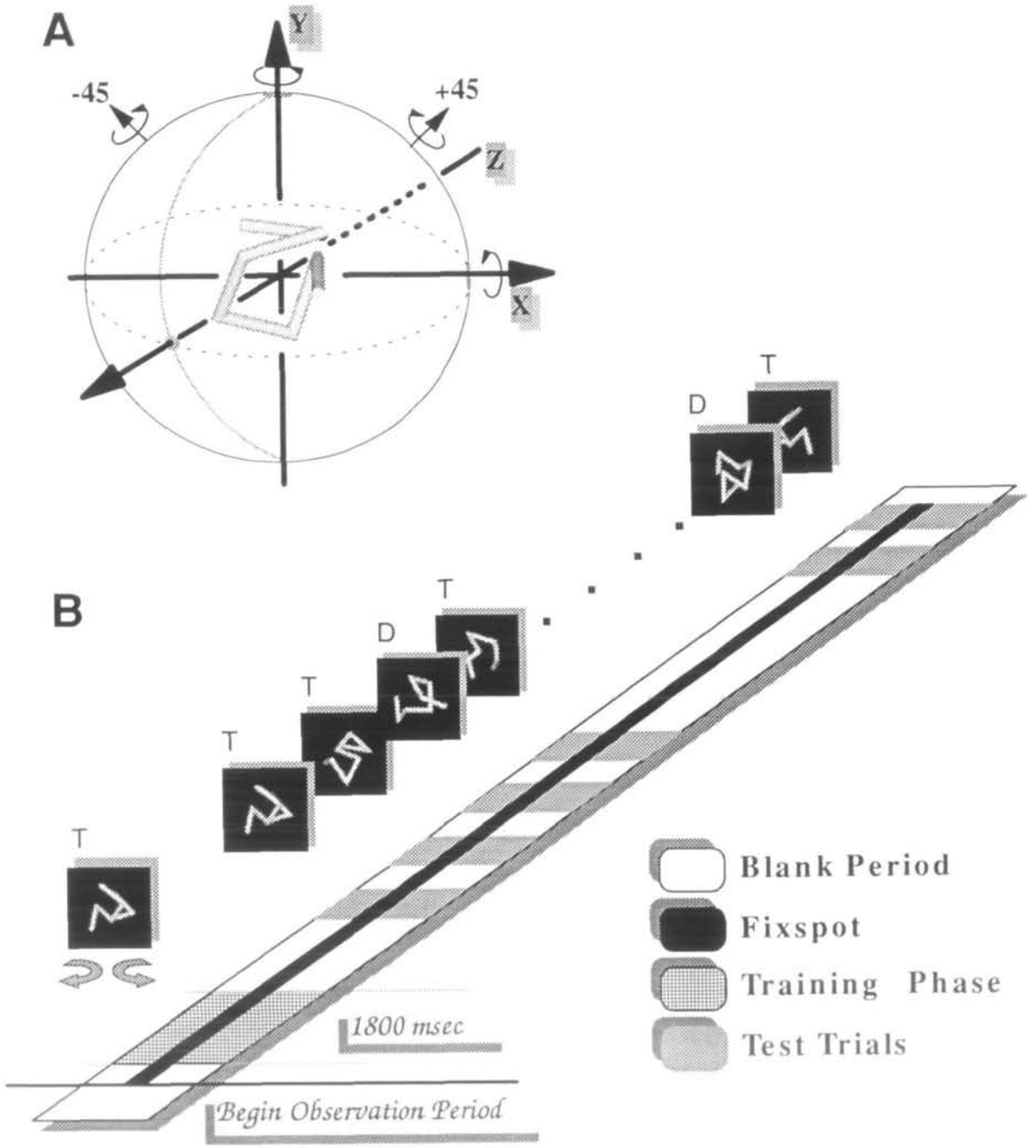


Figure 3. The viewing sphere and the experimental paradigm. *A*, The novel objects used here were created and subjected to transformations with respect to a right-handed reference system. The presentation of various views of the objects was achieved by mathematically calculating the appearance of the object after it underwent arbitrary rotation Recognition was tested for views generated by rotations around the vertical (*Y*), horizontal (*X*), and the two oblique ($\pm 45°$) axes lying on the x–y plane. Thus, rotations around the x-, y-, and z-axes resulted in vertical, horizontal, and plane rotations, respectively. Any arbitrary view of the object could be achieved by the appropriate combination of rotations around the three axes. *B*, An observation period consisted of a *Training Phase* during which the target object was presented oscillating $\pm 10°$ around a fixed axis, and a set of *Test Trials* during which the subjects were presented with up to 10 single static views of either the target or the distractors. Training and testing, as well as individual test trials, were separated by brief *Blank Periods*. Subjects were required to maintain fixation throughout the entire observation period, as indicated by the persistence of the *Fixspot*. The subject had to respond by pressing one of two levers, right for the target and left for the distractors. Feedback was not given under testing conditions.

adding different degrees of positional or orientation noise to the target objects. A criterion of 95% correct for several objects was required to proceed with the psychophysical data collection.

In the early phase of the training a juice reward followed each correct response. In later stages of the training the animals were reinforced on a variable-ratio schedule, within which reward was administered after a specified average number of consecutive correct responses. Finally, in the last stage of the training the monkey was rewarded only for 10 consecutive correct responses. The end of the observation period was signaled by an increased juice reward and a green flash that filled the screen.

During the training, irrespective of reinforcement schedule, the monkey always received feedback as to the correctness of its response since one incorrect report aborted the entire observation period. In contrast, no feedback was given to the monkey during the psychophysical data collection. The behavior of the animals was continuously monitored during the data collection by computing on-line hit rate and false alarms. To discourage arbitrary performance or the

development of hand preferences, for example, giving only right-hand responses, sessions of data collection were randomly interleaved with sessions with novel objects, in which incorrect responses aborted the trial.

Task Description and Data Collection

Figure 3*B* describes the sequence of events in a single observation period. Successful fixation was followed by the *learning phase.* In this phase the target was inspected for 2–4 sec from one or two viewpoints, called the *training views.* To provide the subject with 3D structure information, the target was presented as a motion sequence of 10 adjacent, Gouraud-shaded views, 2° apart, centered around the zero view. The animation was accomplished at a temporal rate of two frames per view; that is, each view lasted 33.3 msec, yielding the impression of an object oscillating slowly ± 10° around a fixed axis.

The learning phase was followed by a short fixation period after which the *testing phase* started. Each testing phase consisted of up to 10 trials. The beginning of a trial was indicated by a low-pitched tone, immediately followed by the presentation of the test stimulus, a shaded, static view of either the target or a *distractor.* The duration of stimulus presentation was 500–800 msec. The monkeys were given up to 1500 msec to respond by pressing one of the two levers. Typical reaction times were below 1000 msec for all the animals. An experimental session consisted of a sequence of 60 observation periods, each of which lasted about 25 sec.

During each observation period the animals' eye movements were measured using the scleral search coil technique (Robinson, 1963), and digitized at 200 Hz. Manual responses were recorded at 200 Hz through a digital I/O interface. Recording of single-unit activity was done using platinum-iridium electrodes of 2–3 MΩ impedance. The electrodes were advanced into the brain through a guide tube mounted into a ball-and-socket positioner. By swiveling the guide tube different sites could be accessed within an approximately 10 × 10 mm cortical region. Action potentials were amplified (Bak Electronics, model 1A-B) and routed to an audio monitor (Grass AM-8) and to a time–amplitude window discriminator (Bak model DIS-1). The output of the window discriminator was used to trigger the real-time clock interface (KWV11) of the computer (PDP11/83).

Results

Viewpoint-Dependent Recognition Performance

The monkeys were trained to recognize any given object viewed on one occasion in one orientation, when presented on a second occasion in a different orientation. Technically, this is a typical "old–new" recognition memory task, whereby the subject's ability to remember stimuli to which it has been exposed is tested by presenting those stimuli intermixed with other objects never before encountered. The subject is required to state for each stimulus whether it is familiar (old, target) or unfamiliar (new, distractor). The probability of the subject reporting "familiar" when presented a target determines the *hit rate,* while the probability of reporting "familiar" when presented a distractor determines the *false alarm rate.* Both measures are plotted in the figures shown below.

Figure 4 shows the average performance of two monkeys tested for the recognition of the same wire-like object. Both the target and the distractors were generated using the same constraints; that is, they had a similar moment of inertia, similar variability in segment orientation, and identical segment length and thickness. Thirty target views generated from rotations around the vertical axis and 60 distractor objects were used during testing. On the abscissa of the graph is plotted the rotation angle and on the ordinate the experimental hit rate. The small rectangles show mean performance for each tested view for four sessions of 40 presentations each. Two monkeys were tested each for all four sessions. The solid line was obtained by a distance-weighted least-squares smoothing of the data using the McLain algorithm (McLain, 1974). The monkeys could correctly identify the views of the target around the trained, zero view, while their performance dropped below chance levels for disparities larger than ±45°. Performance below chance level is probably the result of the large number of distractors used within a session, which limited learning of the distractors per se. Therefore, an object that was not perceived as a target view was readily classified as a distractor. The lower plot in Figure 4 shows the false alarm rate.

To exclude the possibility that the observed view dependency was specific to nonopaque structures lacking extended surface, we have also tested recognition performance using spheroidal, amoeba-like objects with characteristic protrusions and concavities. Thirty-six views of a target amoeba and 120 distractors were used in any given session. As illustrated in Figure 5, the monkey was able to generalize only for a limited number of novel views clustered around the views presented in the training phase.

Interpolating between Familiar Views

The ability of the monkeys to generalize recognition to novel views was also examined after training the animals with two successively presented views of the target 75°, 120°, and 160° apart. The monkey was initially trained to identify two views of an object among 60 distractor objects of the same class. During this period the animal did receive feedback as to the correctness of the response. Note, however, that this "familiarization" phase was the only period in which the monkey was given feedback. No feedback was given during the testing and data collection. Training was considered complete when the monkey's hit rate for the two target views was consistently above 95%, false alarm rate remained below 10%, and the dispersion coefficient of reaction times was minimized. Usually a total of 500–600 presentations were required to achieve the above conditions, after which testing and data collection began.

Interpolation was found complete (above 95% performance) for training views 75° apart. Error rate increased for views 120° apart, and for a disparity of 160° the monkey was unable to interpolate recognition. Figure 6 shows the results of the experiment in which training was done with two views of a wire-like object, 120° apart.

During testing, the monkey was first shown briefly the two familiar views of the object, and then shown 10 stimuli in succession that could be either target or distractor views. The stimuli were pseudorandomly selected from a set of 36 targets and 60 distractors. Within one experimental session each of the 36 tested target views was presented 30 times. As can be seen in Figure 6, the performance of the animal remains above criterion (75%) for all views between and around the trained views. Note the somewhat increased hit rate for views around the −120° view

The experiment with two views, 120° apart, was repeated after briefly training the monkey to recognize the 60° view of the object. During the second "training period" the animal was simply given feedback as to the correctness of the response for the 60° view of the target. The filled circles in Figure 6 show the performance of the monkey in this second experiment. The animal was now able to recognize all views generated by rotating the target around the vertical axis. For most wire-like objects training with three to five views proved enough for generalizing around one great circle. Generalization was often facilitated by the generation of "virtual familiar views" (see below).

Generation of "Virtual Familiar Views"

For some objects the monkeys showed the typical view-dependent performance described above; however, they could also recognize the target from views resulting from approximately 180° rotations of the training view. This type of be-

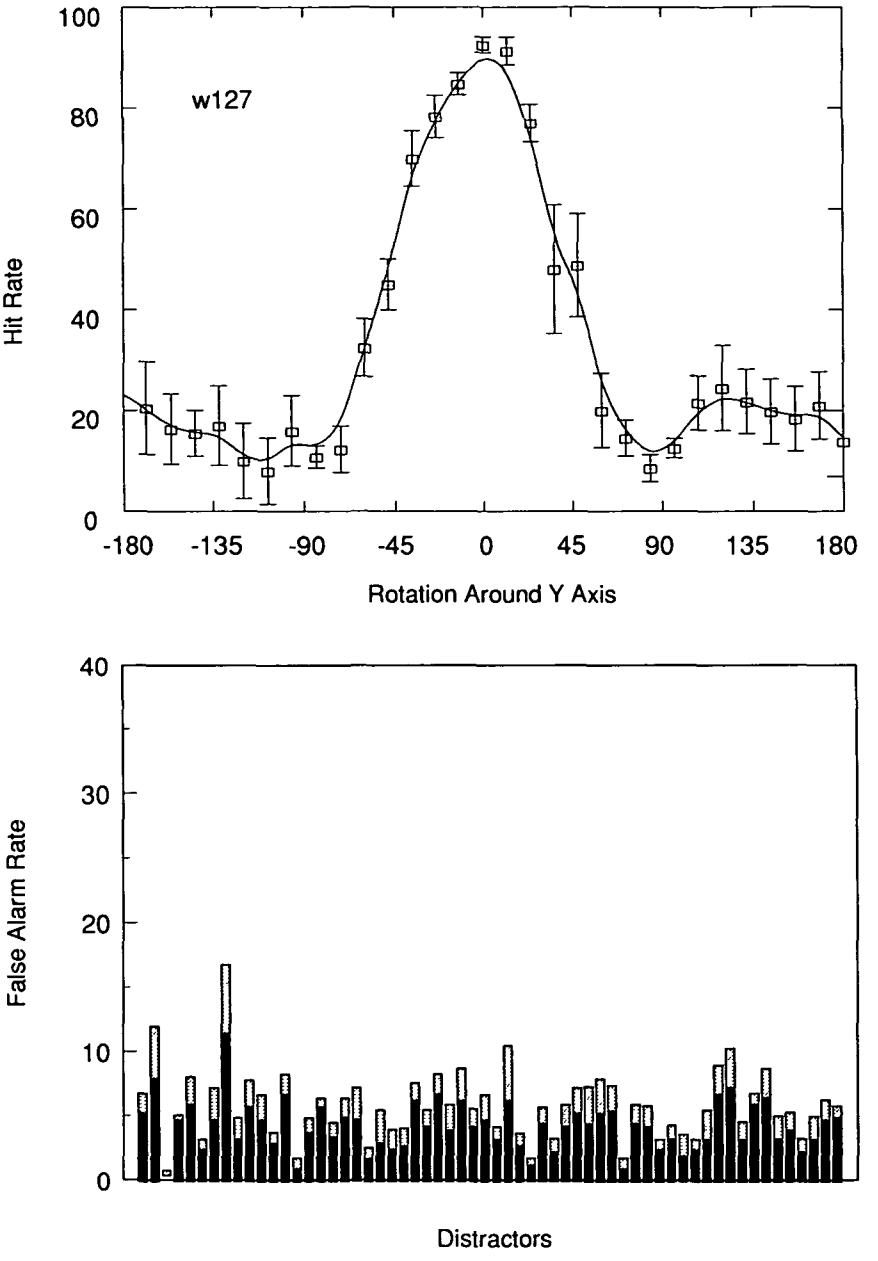


Figure 4. Recognition performance as a function of rotation in depth for a wire-like object (w127). The *abscissa* of the *upper graph* shows the rotating angle, and the *ordinate*, the hit rate. The *rectangles* represent recognition performance for 12° increments around the horizontal meridian, and the error bars are the standard error of the mean for each view. The *solid line* was obtained by a distance-weighted least-squares smoothing of the data using the McLain algorithm. When the object is rotated more than about 30–40° away from zero, the training view, performance falls below 50%. The *lower graph* plots false alarm rates for the 60 different distractor objects used during testing. *Black bars* indicated the mean rate of false alarms for each distractor, and *gray bars* represent the standard deviation.

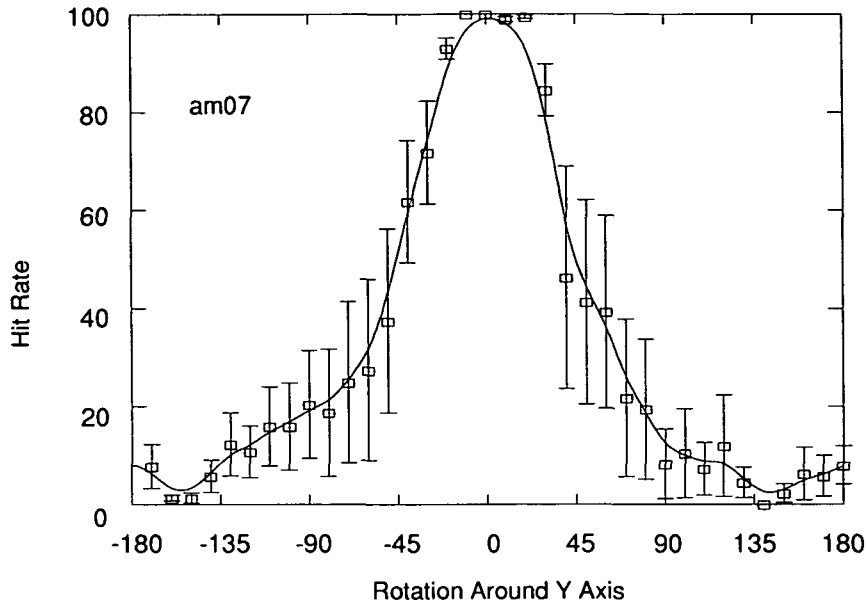
havior is evident in Figure 7 for one of the monkeys. As can be seen in the figure, performance drops for views farther than 30° but it resumes as the unfamiliar views of the target approach the 180° view of the target. This behavior was specific to those wire-like objects for which the zero and 180° views appeared as mirror-symmetrical images of each other, due to accidental minimal self-occlusion. We call such views *pseudo-mirror-symmetrical.* In this respect, the improvement in performance parallels the reflectional invariance observed in human psychophysical experiments (Biederman and Cooper, 1991).

Such reflectional invariance may also partly explain the observation that information about bilateral symmetry simplifies the task of 3D recognition by reducing the number of views required to achieve object constancy (Vetter et al., 1994). Not surprisingly, performance around the 180° view of an object did not improve for any of the opaque, spheroidal objects used in these experiments.

Recognition of "Basic-Level" Objects

Performance was found to be viewpoint invariant when the animals were tested for basic-level classifications, or when they were trained with multiple views of wire-like or amoeba-like objects. It should be noted that the term "basic level" is used here simply to denote that the objects were largely different in shape from the distractors.

Figure 8 shows the mean performance of two monkeys for one object (different views of the starship *Enterprise*). Each curve was generated by averaging individual hit-rate measurements obtained from two animals for the target shown in Figure 8 (upper plot). The lower plot of Figure 8 shows the false alarm rate in the same experiment. Distractors were selected from a set of 120 objects, including geometrical constructs, wires, spheroidals, plane models, or fractal objects (see insets in the lower plot of Fig. 8). Since all animals were already trained to perform the task, independent of the object type used as a target, no familiarization with the object's zero



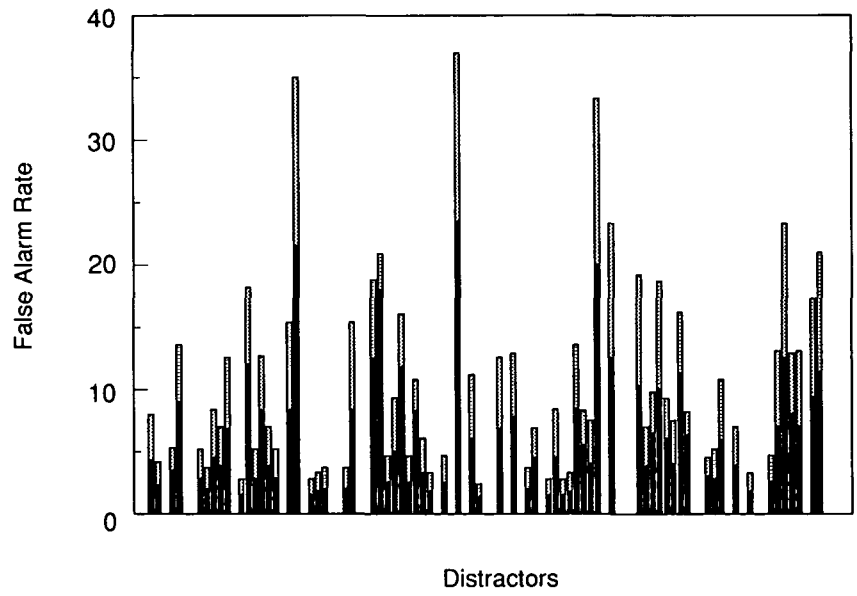


Figure 5. Recognition performance as a function of rotation in depth for a spheroidal object. Data are from one monkey. Conventions are as in Figure 4.

view preceded the data collection in these experiments. Yet, the animals generalized recognition for all tested novel views.

Responses of IT Neurons to Novel 3D Objects

The activity of neurons in the IT has been examined in a simple fixation task and in the experimental paradigm described above. We have collected data from over 700 isolated IT neurons in two macaque monkeys. Figure 1 shows the recording sites as estimated from the stereotaxic coordinates. Since the animals are being used in further experiments on object recognition, no histological reconstructions are currently available.

Isolated units were tested with a variety of simple or complex patterns while the animal was involved in a fixation task. The animal was trained to maintain fixation within a 1° × 1° window. The activity of the cells that responded to either the wire or the amoeboid objects was further examined while the animal performed the recognition task.

We found a number of units that showed a remarkable selectivity for individual views of wire objects that the monkey had learned to recognize. An example of a highly selective neuron is shown in Figure 9. The responses of this cell were studied for 30 different views of a wire-like object that the monkey was trained to recognize. The monkey's performance was above 95% for all the views of this object. The cell was highly selective for views located around −72°. Its activity decreased considerably with even a 12° deviation from the preferred view. Individual wire segments with the same orientation as the segments of the preferred view did not elicit any response when presented in the same location of the receptive field. Nonetheless, the view 180° away from the preferred view (see also Fig. 12) did elicit firing of the neuron (albeit weaker). Figure 10 shows the responses of the same cell for 15 wire-like and 15 nonwire distractor objects; none of these objects elicited any response of the neuron. The responses of this unit to all 120 tested distractors were very similar to those shown in Figure 10.

A second example of a view-tuned neuron is shown in

Figure 11. In this example, the monkey was trained to recognize the views of the wire object labeled 0° and 160°, and subsequently was tested for 36 views around the same axis with no feedback as to the correctness of responses. As mentioned above, when the training views were this far apart no interpolation was observed during the testing period. The monkeys appear to learn the two views in a nonassociative manner, just like learning two different wire objects. The cell responded for views around the zero view of the object, but it did not discharge for the other training view.

Many other cells showed a tuning that was likewise independent of the animal's behavioral performance. These observations provide strong evidence that the activity of these neurons is not simply the manifestation of arousal, attention, or a sensation of familiarity, but rather relates to the object's characteristic features or views. To date, 71 of the 773 (9%) analyzed cells showed view-selective responses similar to those illustrated in Figure 11. In their majority, the rest of the neurons were visually active when plotted with other simple or complex stimuli, including faces. A small percentage of neurons, although frequently firing with a rate of 5-20 Hz, could not be driven by any of the stimuli used in these experiments. A detailed, quantitative description of the response types is currently in preparation.

No selective responses were ever encountered for views that the animal systematically failed to recognize. A small percentage of the view-selective cells (5 of the 71) responded strongly for a particular view and its pseudo-mirror-symmetrical view An example of a neuron responding to pseudo-mirror-symmetrical views is shown in Figure 12. This cell was most responsive for a set of views around 100°; however, it also gave a large response for a set of views around the −80° view of the target. The monkey's performance was 100% for almost all the tested views. The high performance is the result of giving feedback to the animal for multiple views of the object. Neither of the neuron's preferred views, however, represents the view which the monkey was shown in the training period.

A small percentage of cells (8 of 773) responded to wire-like objects presented from any viewpoint, thereby showing view-invariant response characteristics. An example of such a neuron is shown in Figure 13. The upper plot shows the monkey's hit rate and the middle plot the neuron's average spike rate. The cell fires with a rate of about 40 Hz for all target views. The lower plot shows the responses of the same cell to 120 distractors. With four exceptions, activity was uniformly low for all distractor objects presented. In all cases, even the best response to a distractor, however, remains about one-half of the worst response to a target view.

View-selective neurons were also tested for invariance across changes in position or size. Most neurons preserved their selectivity independent of changes in size Responses, however, varied according to the stimulus position in the receptive field, with stronger responses usually elicited in the foveal region. Two exceptional cells that showed almost complete position invariance are described by Pauls et al. (1995).

Finally, for three objects more than one neuron was found to be tuned to different views of the same object. Figure 14 shows the responses of three units that were found to respond selectively to different views of a wire object (wire 71). The distance between the peaks of the tuning curves of individual neurons averaged about 60°. The neuron identified as 216 is the same shown in Figure 12. The animal had been exposed repeatedly to this object, and its psychophysical performance remains above 95% for all tested views, as can be seen in the lower plot of Figure 14.

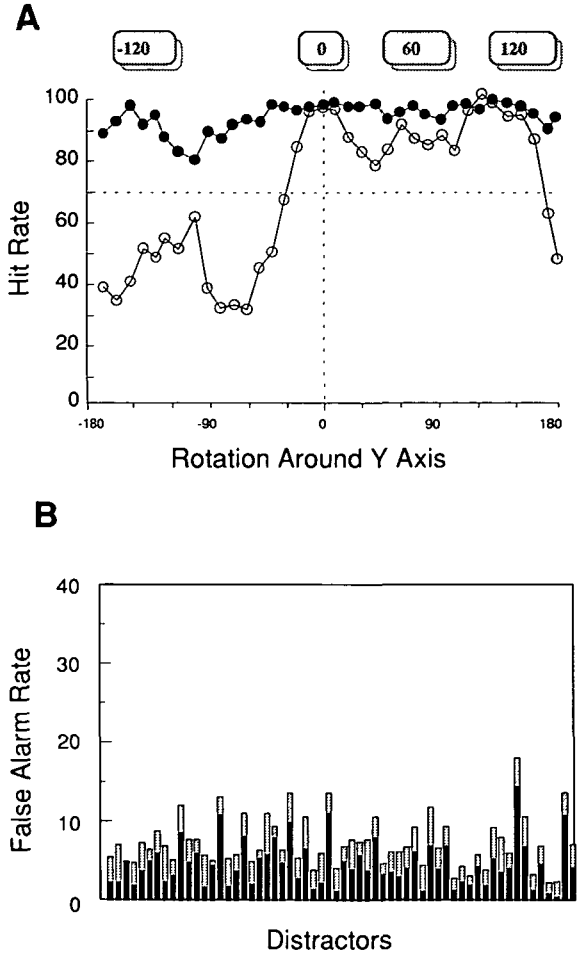


Figure 6. Training with multiple views and performance with objects that belong to different classes. *A*, In the learning phase the monkey was presented sequentially with the 0° and 120° views of a wire-like object and was subsequently tested with 36 views around the vertical axis. The *open circles* show the performance of the animal. *Filled circles* show performance when the monkey was tested with the same views, but after brief training with the 60° view of the wire object. The animal can now recognize the object from any viewpoint. For most objects, training with three to five views proved enough for generalizing around the entire great circle. *B*, False alarm rates during the same experiment. *Black bars* represent the mean, and *gray bars* represent the standard deviations.

Discussion

The main findings of this study can be summarized as follows. (1) Even when complete information about the structure of an object is available to the subject, recognition at the subordinate level depends on the object's attitude. (2) A memory-based, viewer-centered recognition system is not an implausible mechanism for object constancy. Both theoretical work and the results of the experiments described here suggest that only a small number of object views need to be stored to achieve the perceptual invariance that biological visual systems exhibit in everyday life. (3) A small population of neurons in IT was found to respond selectively to individual members of the object classes tested in this study. The response of some neurons was a function of the object's view. Some of the view-selective neurons responded equally well to mirror-symmetrical views. (4) For all objects used in the combined psychophysical-electrophysiological experiments, view tuning was observed only for those views that the monkey could recognize. Several neurons were also found that

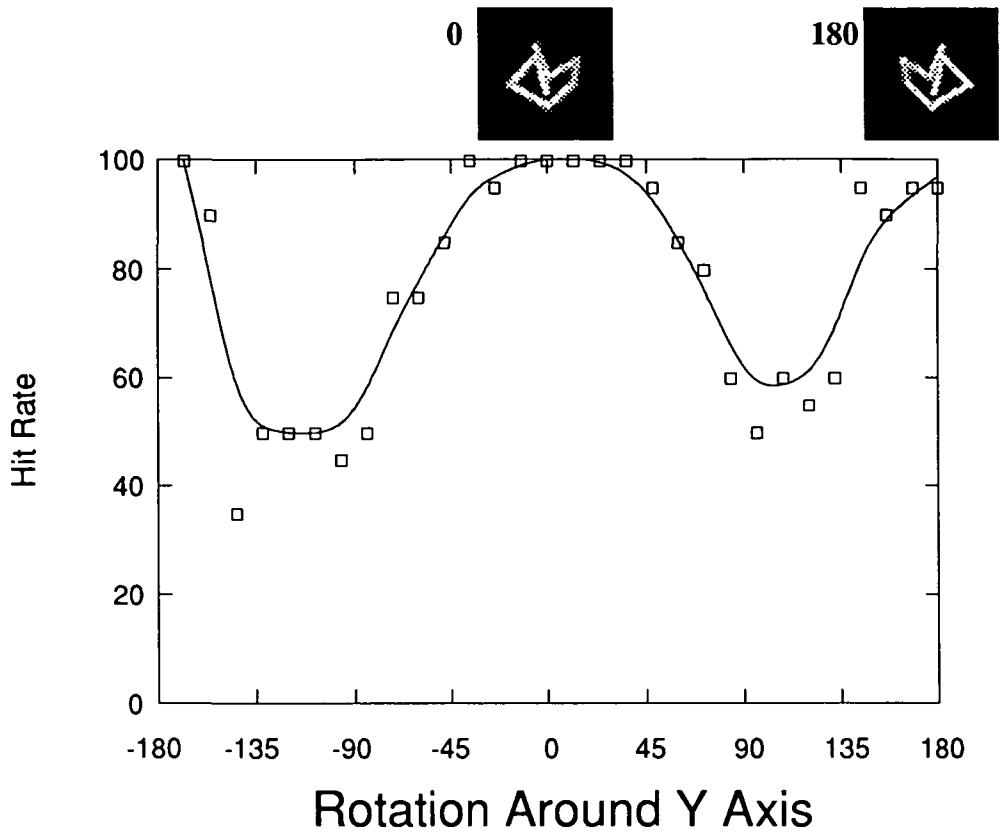


Figure 7. Responses to "pseudo-mirror-symmetrical" images: improvement of recognition performance for views generated by 180° rotations of wire-like objects. This type of performance was specific to wire-like objects that possessed zero and 180° views resembling mirror-symmetrical, 2D images due to accidental lack of self-occlusion. Conventions are as in Figure 4A.

responded to the sight of unfamiliar or distractor objects. Such cells, however, gave nonspecific responses to a variety of other patterns presented while the monkey performed a simple fixation task.

View Dependency of Subordinate-Level Recognition

The first demonstration of strong viewpoint dependence in the recognition of novel objects was that of Rock and his collaborators (Rock et al., 1981; Rock and DiVita, 1987). These investigators examined the ability of subjects to recognize 3D, smoothly curved wire-like objects experienced from one viewpoint, when presented from a novel viewpoint. In actuality, they presented real, 3D objects in one of the quadrants of the visual field (e.g., the upper right) and subsequently tested the subjects' recognition performance when the object was presented in another quadrant. Thus, viewpoint changes in their experiments were the result of changes in the position of a real, 3D object. In their study, they found that humans are unable to recognize views corresponding to object rotations as small as 30-40° around a given axis. This result was obtained even though their stimuli provided the subject with full 3D information. Furthermore, subsequent investigations showed that subjects could not even imagine what a wire-like object looks like when rotated, despite instructions for visualizing the object from another viewpoint (Rock et al., 1989). Similar results were obtained in later experiments by Edelman and Buelthoff with computer-rendered, wire-like objects presented stereoscopically or as flat images (Buelthoff and Edelman, 1992; Edelman and Buelthoff, 1992).

In this article we provide evidence of similar view dependency of recognition in the nonhuman primate. All tested monkeys were unable to recognize objects rotated more than approximately 40° of visual angle from a familiar view. These results are hard to reconcile with theories postulating object-centered representations. Such theories predict uniform performance across different object views, provided that 3D information is available to the subject at the time of the first encounter. Thus, one essential issue is whether information about the object's structure was available to the monkeys during the learning phase of these experiments.

For one, the objects were computer rendered with appropriate shading (Gouraud-shaded views) and were presented in slow oscillatory motion. The kinetic depth effect (motion parallax) produced by such motion yields vivid and accurate perception of the 3D structure of an object or surface (Braunstein, 1968; Rogers and Graham, 1979). In fact, psychometric functions showing depth modulation thresholds as a function of spatial frequency of 3D corrugations are very similar for surfaces specified through either disparity or motion parallax cues (Rogers and Graham, 1982, 1983, 1985). Experiments on monkeys have shown that nonhuman primates, too, possess the ability to see such structure from motion in random-dot kinematograms (Siegel and Andersen, 1988).

Second, wires were visible in their entirety since, unlike most opaque natural objects in the environment, regions in front do not substantially occlude regions in back. As mentioned above, when many of these objects are seen from viewpoints that are 180° apart, they resemble "mirror-symmetrical" views of a 2D pattern. Thus, during the learning phase of each observation period, information about the 3D structure of the target was available to the monkey by virtue of shading, the kinetic depth effect, and minimal self-occlusion.

Could the view-dependent behavior of the animals be a result of the monkeys' failing to understand the task? The monkey could indeed recognize a 2D pattern as such, without necessarily perceiving it as a view of an object. Correct performance around the familiar view could then be simply

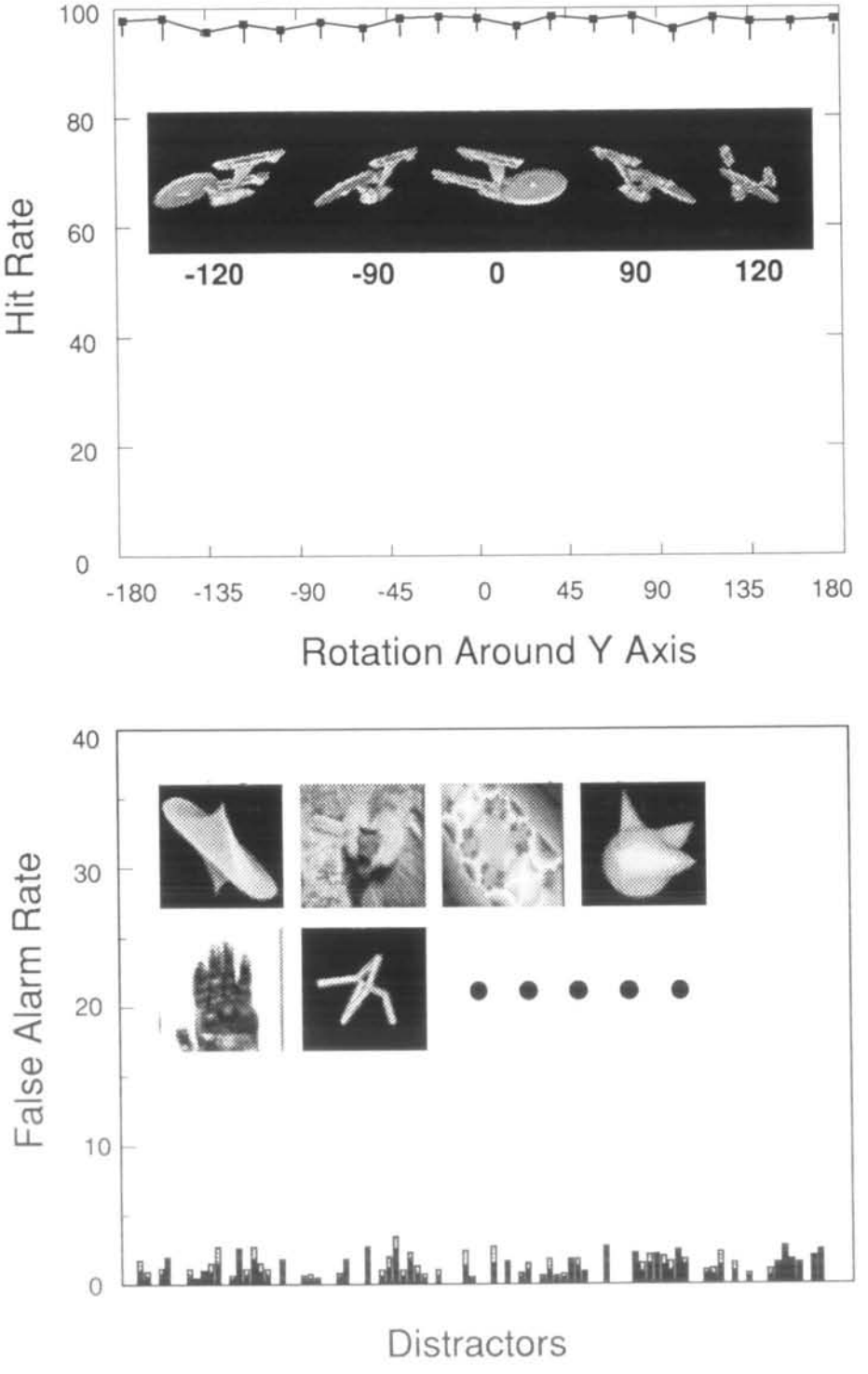


Figure 8. Viewpoint-invariant performance with "basic" objects. *A*, Performance of monkey S5396 in an experiment in which the target and the distractors belonged to different object classes (different common-type objects, such as faces, teapots, geometrical constructs, plane models, etc.). *B*, False alarm rate for the same experiment. Examples of distractors are shown as insets. Conventions are as in Figure 4.

explained as the inability of the animal to discriminate adjacent views. Several lines of argument refute such an interpretation of the obtained results. First, the animals easily generalized recognition to all novel views of common objects. Second, when the wire-like objects had prominent characteristics, such as very sharp or very wide angles, closures, or other pronounced combinations of features, the animals were able to perform in a view-invariant fashion. Evidently the visual system will use any information available to identify an object. A representation can be built based on a detailed description of shape as well as on only some characteristic features of an object, including nongeometrical properties such as its color or texture. There is no reason to expect that either humans or monkeys will rely on subordinate shape discriminations when the objects can be clearly identified otherwise. Third, when more than one view of the target was presented in the training phase, the animals successfully interpolated between the trained views, often with 100% performance. Interpolation between views could not be explained by simply combining the generalization gradients of the trained views, since performance for the views between the samples was at least two times better

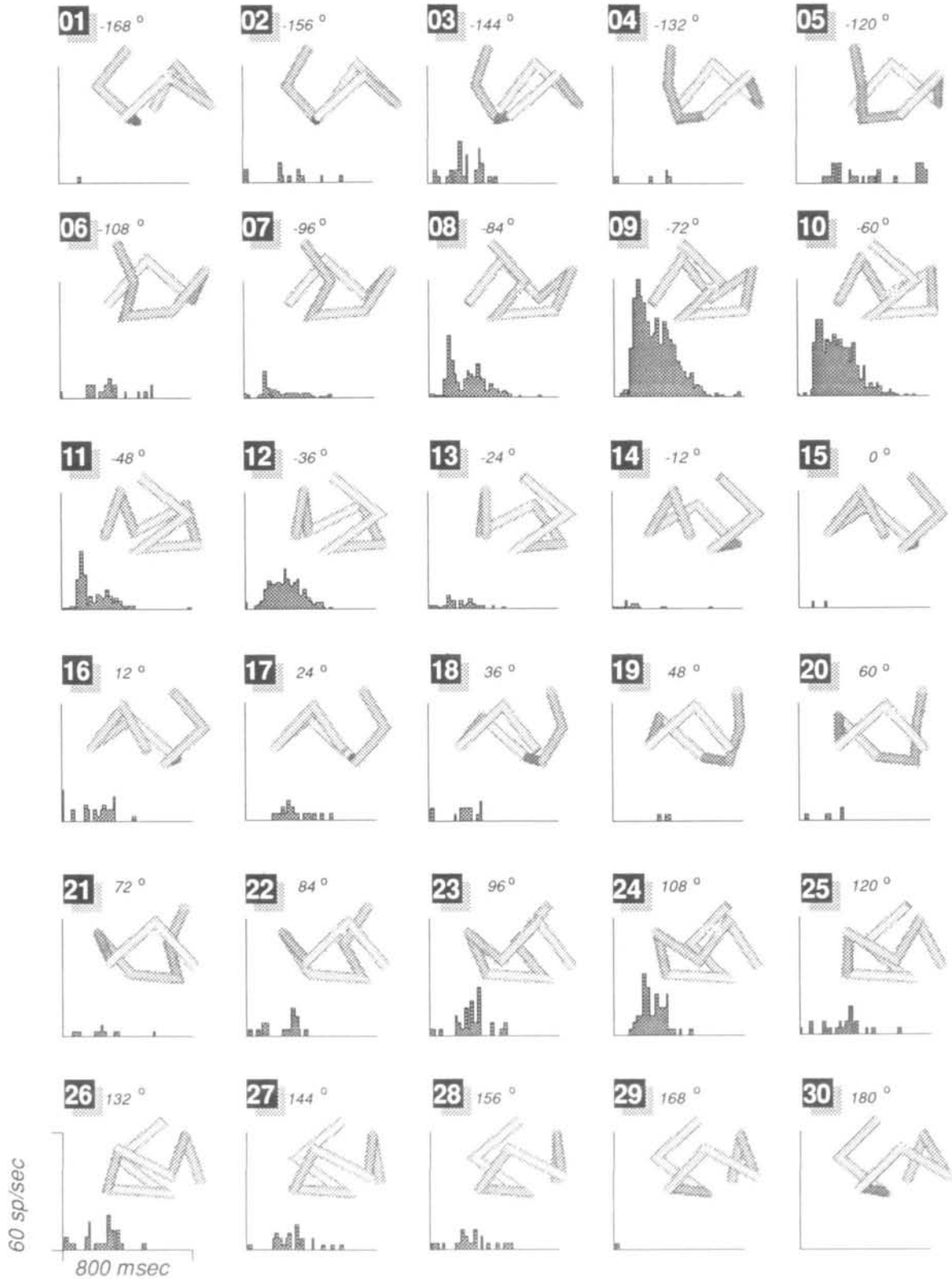


Figure 9. Example of a view-selective cell. The *abscissa* of each small plot represents time, and the *ordinate* is the spike rate. *Angles* refer to the corresponding view of the target. Performance of the monkey was above 95% for all views of the targets (not shown here). This cell was highly selective for views located around −72°. Activity decreased considerably with even a 12° deviation from the preferred view. Note the improved response for the view 180° away from the preferred view.

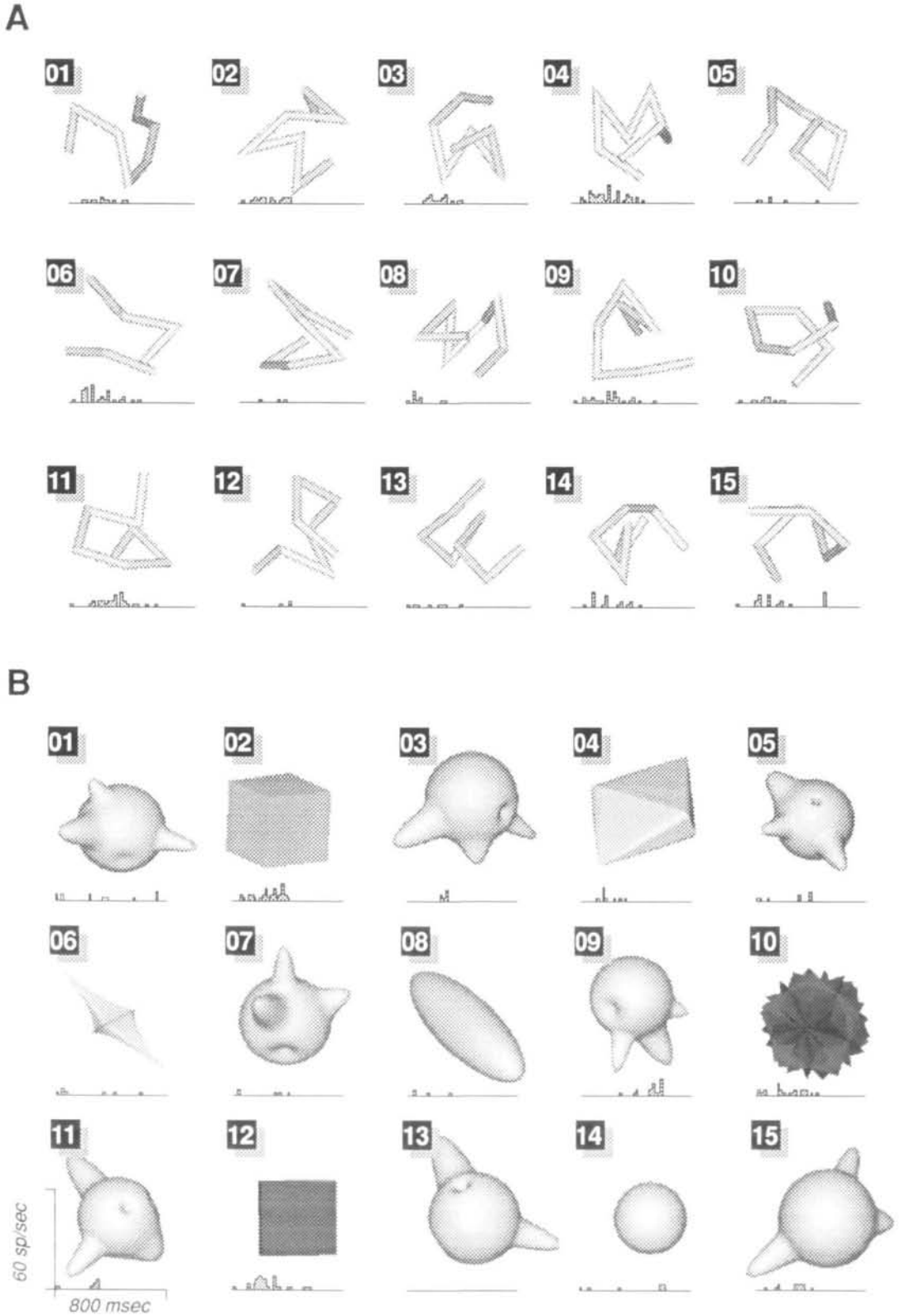


Figure 10. Responses to distractors. *A*, Responses to 15 wire-like distractor objects used in testing the view-selective cell in Figure 9. *B*, A set of 30 nonwire distractor objects, some of which had been used as targets in earlier sessions, is presented, together with the peristimulus histograms of the neuron.

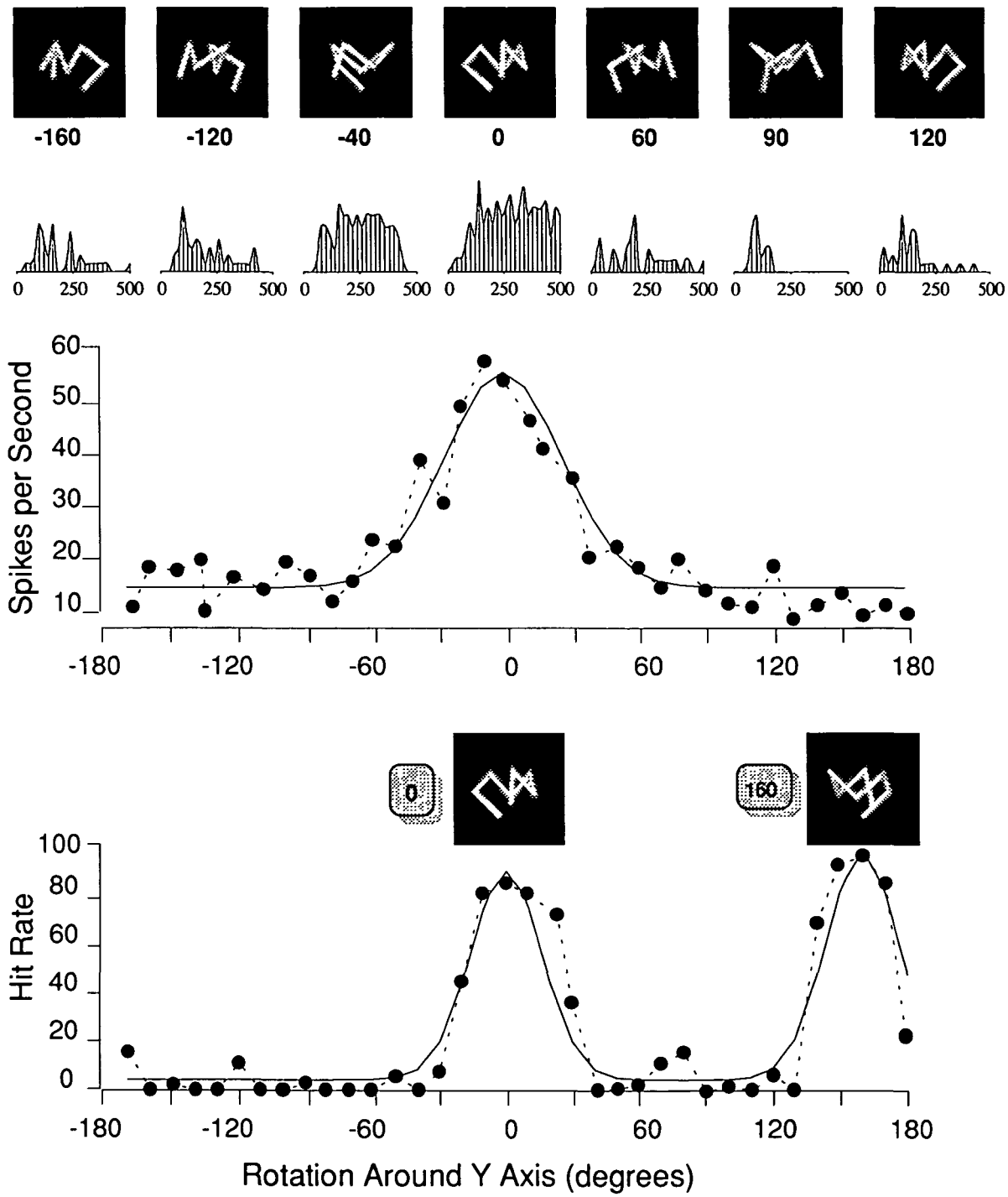


Figure 11. Hit rate and mean spike rate as a function of target view: recognition performance for a monkey (*lower plot*) as a function of rotation around the y-axis. The *two insets labeled 0 and 160* are the object views presented in the learning phase. Testing was carried out using views spaced at 10° intervals around the horizontal meridian. The *upper plot* shows the mean spike rate as a function of rotation angle for a neuron recorded during the same session. The *insets* show a subset of the tested views, and the *small plots*, the cell activity during the test-view presentation. The broken line in each plot serves only to connect the data points for easier visualization. The solid line is a fit of the data with one (upper) or two (lower) Gaussian curves determined using the quasi-Newton method.

than that expected from the conjunction of the bell-shaped performance curves (Logothetis and Pauls, 1994) obtained in experiments in which the animals were trained with only one view.

The strongly view-dependent performance at the subordinate level of recognition was not specific to the wire-like objects. Similar performance was observed within an entirely different object class, whose members had extensive surface and occlusion like many objects in daily life. Thus, it appears that monkeys, just like human subjects, show rotational invariance for familiar, basic-level objects, but they fail to generalize recognition at the subordinate level, when identification of individual entities relies on fine, shape-based discriminations. Interestingly, training with a limited number of views was sufficient for all the animals tested to achieve view-independent performance. Hence, view invariance based

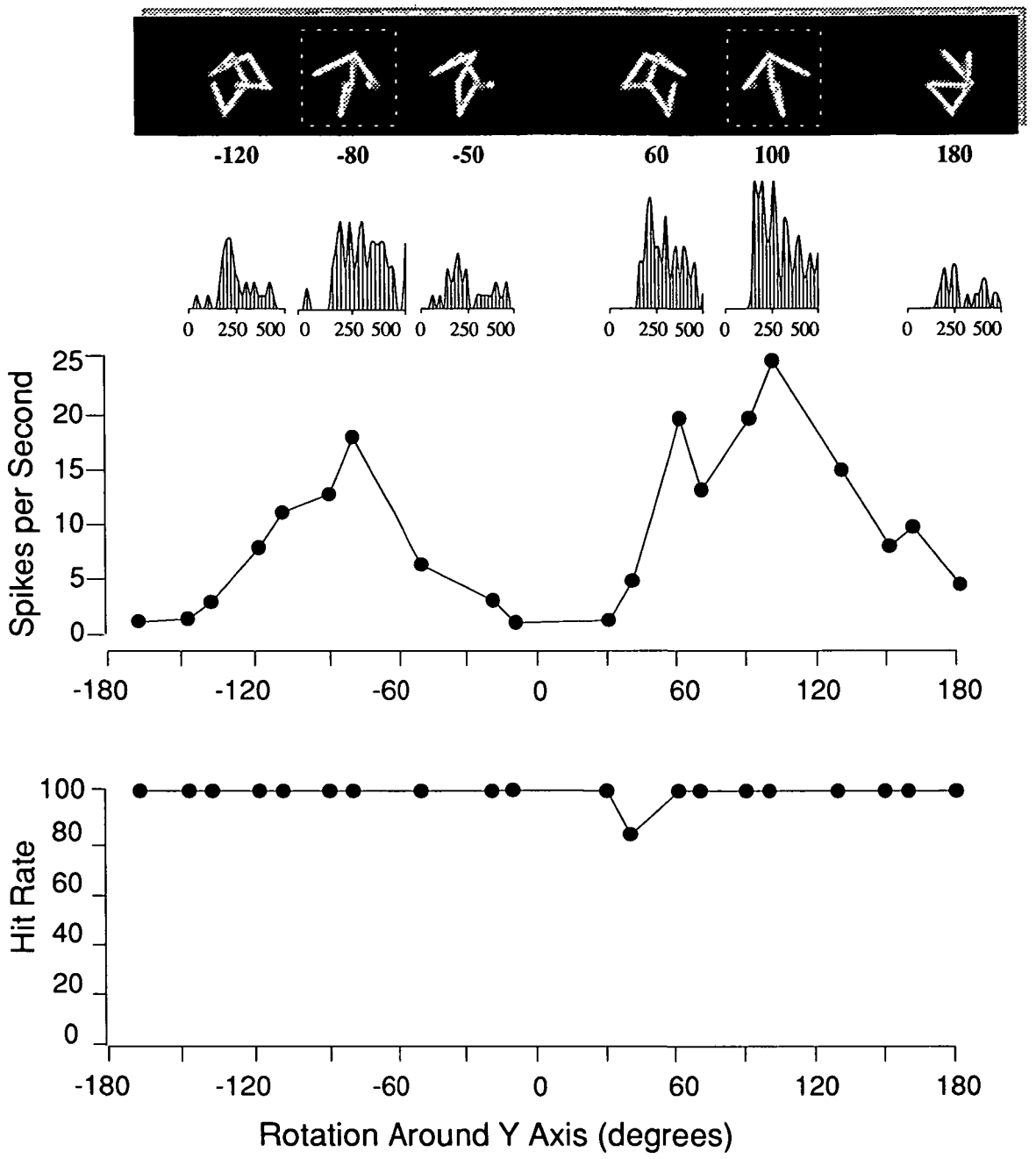


Figure 12. Cell responses for pseudo-mirror-symmetrical views. Recognition performance for a wire-like object is presented as a function of rotation angle around the y-axis (*lower graph*). The monkey was trained on one view, at 0°, and was subsequently tested at 10° intervals around one axis. Note that at the time these data were collected the recognition performance for the monkey for this object was above 90% for all views around the y-axis. The spike rate of a single neuron, as a function of rotation angle, is plotted in the *middle graph.* This particular cell was most selective for a set of views around 100°; however, it also gave a large response for the set of views around the view at −80° (compare the *two inset histograms* and their associated views, highlighted by the *dashed squares*).

on familiarity does not require the storage of a formidable number of object views.

Finally, it is worth noting that recognition based entirely on fine shape discriminations, such as those described here, is not uncommon in daily life. Face identification is a striking example of subordinate-level recognition. Despite the great structural similarity among individual faces, face recognition is an "easy task" for both humans and monkeys. We are certainly able to recognize mountains or cloud formations, as well as man-made objects like modern sculptures or different types of cars. Many of these objects are recognized based on their shape, and most of them cannot necessarily be structurally decomposed into simpler parts. Moreover, even those theories suggesting that recognition involves the indexing of a limited number of volumetric components (Biederman, 1987) and the detection of their relationships have to face the problem of learning those components that cannot be further decomposed.

View Selectivity in Inferior Temporal Cortex

Cells selective for specific patterns or object views are not rare throughout the IT. View selectivity has previously been reported for face-selective IT neurons. Desimone et al. (1984) reported cells that were sensitive to the orientation of the

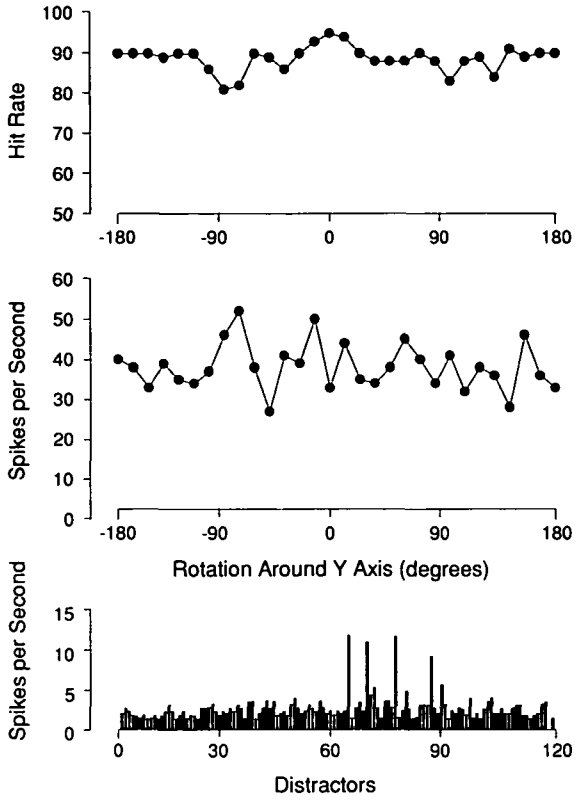


Figure 13. Example of a neuron showing view-invariant activity. The monkey's recognition performance, in terms of hit rate, is plotted as a function rotation around the y-axis (*top graph*). In this case the monkey's performance is greater than 80% for all views. The *middle graph* shows the activity of a cell for the same set of object views. This cell is selective for the entire object, showing invariance to rotation in depth. The *cell's mean response to each of 120 different distractors is presented in the lower graph.*

head in depth. Some cells were maximally sensitive to the front view of a face, and their response fell off as the head was rotated into the profile view; other cells were sensitive to the profile view, with no reaction to the front view of the face. In their example, the cell's activity fell to half its maximum when the face was rotated about 30-40° (Desimone et al., 1984, their Fig. 7), which is in close agreement with the data presented in this article using the wire-like or amoeboid objects.

A detailed investigation of these types of cells by Perrett and colleagues revealed a total of five cell types in the STS, each maximally responsive to one view of the head (Perrett et al., 1985). The five types of cells were separately tuned for full face, profile, back of the head, head up, and head down. Most of these neurons were found to be 2-10 times more sensitive to faces than to simple geometrical stimuli or 3D objects (Perrett et al., 1979, 1982).

Recent studies have shown that such selectivity appears early in the visual system of the monkey. Rodman and colleagues found cells in the IT of infant monkeys (as young as 5 weeks old) that exhibited responses selective for shape, faces, geometrical patterns, and color (Rodman et al., 1993). So, it seems that at least some of the neurons that are selective for highly complex patterns are available to the recognition system even at the earliest developmental stages of the visual system.

In the present study, we found neurons that responded selectively to novel visual objects that the monkey learned to recognize during the experiments. None of these objects had any prior meaning to the animal, and none of them resembled anything familiar in the monkey's environment. Thus, it appears that neurons in this area can develop a complex receptive field organization as a result of extensive training in the discrimination and recognition of objects.

The monkeys were trained with different types of objects, such as the wire-like, the spheroidal, and the basic types of objects shown in Figure 2. Interestingly, the frequency of encountering neurons selective to a particular object type seemed to be related to the animal's familiarity with the object class. In one of the monkeys, the wire-like objects, which were extensively used during the psychophysical experiments, were much more likely to elicit cell responses (71 selective cells) than, for example, spheroidal objects (10 selective cells), which were used to a much lesser extent. The converse was observed in another animal that was extensively trained with the amoeboid objects. Most selective neurons responded best to one view of the object, while their response decreased as the object was rotated away from the preferred view. Plotting the cell responses as a function of rotation angle revealed systematic view tuning curves similar to those obtained from striate neurons tested with lines rotated in the frontal plane.

Invariance for Reflections

Some of the neurons showed similar response magnitude for views of the wire-like objects that were 180° apart. As mentioned above, these views tend to look like mirror-symmetrical images of 2D patterns, thus implying some sort of "reflection invariance" in the response of IT neurons.

Interestingly, some face cells have also been found to respond to views of a face 180° apart, especially the left and right profiles (Perrett et al., 1989). These cells are presumably similar to those reported here as responding to the "pseudo-mirror-symmetrical" views. Reflection invariance has also been shown in the responses of some face-selective IT cells in infant monkeys (Rodman et al., 1993), a finding suggesting that reflectional invariance may be generated automatically for every learned object, and may be already present early in an individual's life.

In fact, such an invariance may be the cause of the inability of children to distinguish between mirror-symmetrical letters like "d" and "b." This type of letter confusion was studied intensively in children by Orton (1928). He observed a delay in the learning of mirror-symmetrical letters and words, as well as several other characteristics involving the establishment of "handedness," in language-handicapped children (Orton, 1928). This eventually led to the description by Orton of a disorder known as *strephosymbolia.* This confusion, observed in normal children as well, appears to be the rule during development and not the exception (Corballis and McLaren, 1984). In fact, Gross and Bornstein suggest that the confusion of mirror images may be an adaptive mode of processing visual information and not a real "confusion" (Bornstein et al., 1978; Gross and Bornstein, 1978). These authors note correctly that in the natural world there are never any mirror images that would be useful for an animal to distinguish. Even in the case of bilateral symmetry observed in most animals, the two mirror-symmetrical sides are aspects of the same thing, and it would be more adaptive to treat them as the same.

In our experiments, improvement in the recognition of pseudo-mirror-symmetrical views, as well as cell selectivity to pseudo-mirror-symmetrical views, was observed for rotations around any of the four tested rotation axes. At a first glance this might appear surprising since most of the mirror symmetry that humans experience is around the vertical axis. In fact, rotations of faces around the horizontal axis usually have a robust effect on human performance, while they have

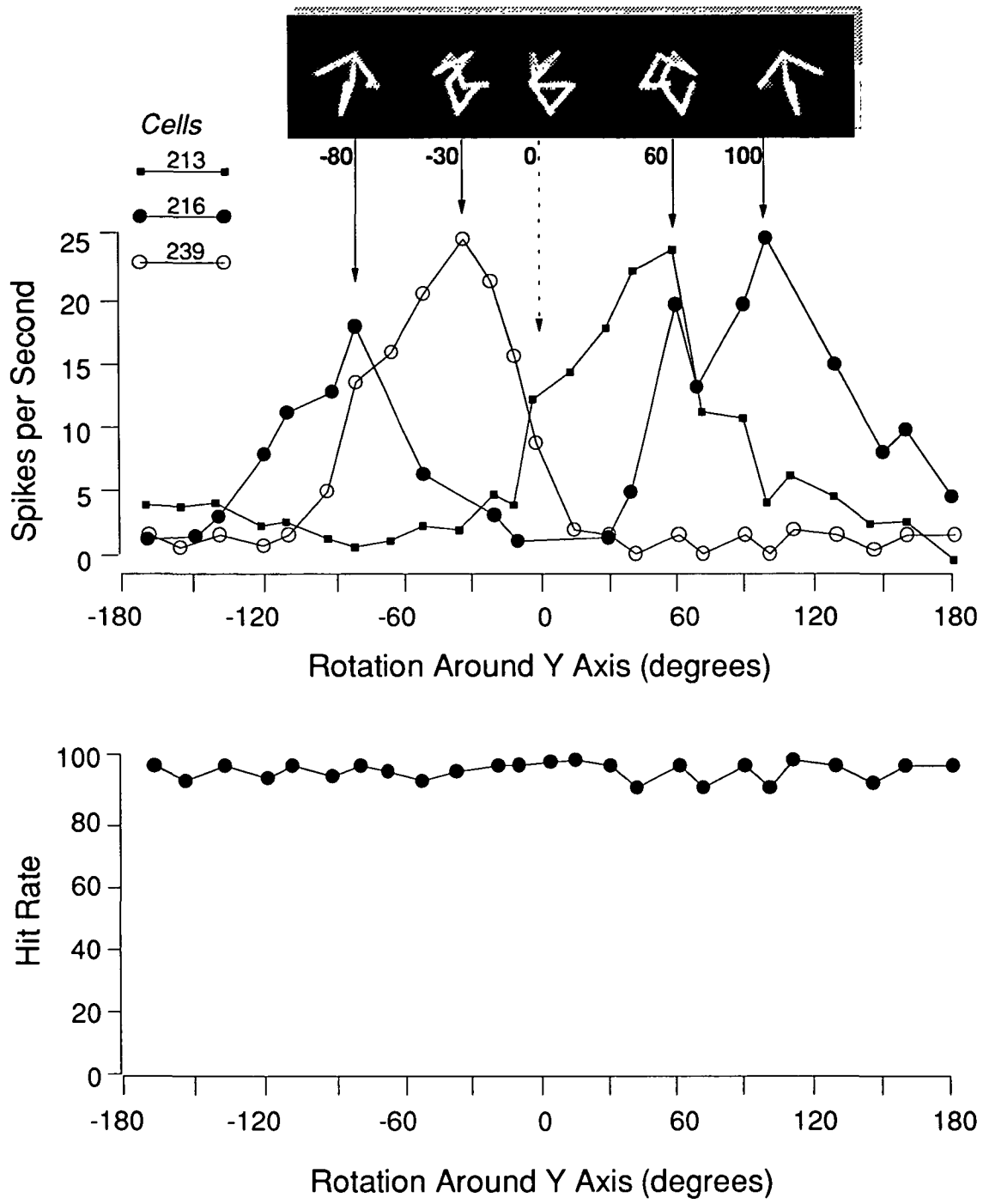


Figure 14. Example of multiple neurons tuned to the same object. Conventions are as in Figure 12.

no significant effect on the performance of the monkey (Bruce, 1982). The difference might be attributed to the development of laterality in humans, or simply to the fact that monkeys often look at each other from an upside-down attitude. Interestingly, children up to 10 years old, who may still have incomplete development of laterality, can remember faces presented upside down almost as well as they do those presented upright (Carey and Diamond, 1977).

Nonlinear Interpolation between Stored Views

The psychophysical performance of the animals seems to be consistent with the idea that view-based approximation modules synthesized during training may indeed be one of several algorithms the primate visual system uses for object recognition. Training the monkey with one single view results in a bell-shaped generalization field for both the wire-like and the amoeboid objects. Moreover, the ability of the monkey to interpolate between two familiar views depends on their distance, a finding difficult to reconcile with a recognition system based on linear interpolation (Ullman and Basri, 1991) but directly predicted by a system relying on nonlinear approximation (Poggio and Girosi, 1990). Sets of neurons that are tuned broadly to individual object views may represent the neural substrate of such approximation modules.

Notes

We thank Drs. D. Sheinberg and J. Assad for critical reading of the manuscript. This research was sponsored by grants from the National Institutes of Health (NIH 1R01EY10089-01) and the Office of Naval Research (N00014-93-1-0290) and by a McKnight Endowment Fund for Neuroscience to N.K.L.

Correspondence should be addressed to N. K. Logothetis, Division of Neuroscience, Baylor College of Medicine, One Baylor Plaza, Houston, Texas 77030.

References

Biederman I (1987) Recognition-by-components: a theory of human image understanding. Psychol Rev 94:115-147.

Biederman I, Cooper EE (1991) Evidence for complete translational and reflectional invariance in visual object priming Perception 20:585-593.

Bornstein MH, Gross CG, Wolf JZ (1978) Perceptual similarity of mirror images in infancy Cognition 6:89-116.

Braunstein ML (1968) Motion and texture as sources of slant information. J Exp Psychol 78:247-253.

Bruce CJ (1982) Face recognition by monkeys: absence of an inversion effect. Neuropsychologia 20:515-521.

Bruce CJ, Desimone R, Gross CG (1981) Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque. J Neurophysiol 46:369-384.

Buelthoff HH, Edelman S (1992) Psychophysical support for a two-dimensional view interpolation theory of object recognition. Proc Natl Acad Sci USA 89:60-64.

Carey S, Diamond R (1977) From piecemeal to configuration representation of faces. Science 195:312-313

Corballis MC, McLaren R (1984) Winding one's Ps and Qs: mental rotation and mirror-image discrimination. J Exp Psychol [Hum Percept] 10:318-327.

Damasio AR (1990) Category-related recognition defects as a clue to the neural substrates of knowledge. Trends Neurosci 13:95-99.

Desimone R, Gross CG (1979) Visual areas in the temporal cortex of the macaque. Brain Res 178:363-380.

Desimone R, Albright TD, Gross CG, Bruce CJ (1984) Stimulus-selective properties of inferior temporal neurons in the macaque. J Neurosci 4:2051-2062

Edelman S, Buelthoff HH (1992) Orientation dependence in the recognition of familiar and novel views of 3D objects. Vision Res 32: 2385-2400.

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

Gerstein GL, Gross CG, Weinstein M (1968) Inferotemporal evoked potentials during visual discrimination performance by monkeys J Comp Physiol Psychol 65:526-528.

Gochin PM, Miller EK, Gross CG, Gerstein GL (1991) Functional interactions among neurons in inferior temporal cortex of the awake macaque Exp Brain Res 84:505-516

Gross CG (1973) Visual functions of the inferotemporal cortex. In: Handbook of sensory physiology (Jung, R, ed.), pp 451-482. Berlin: Springer.

Gross CG, Bornstein MH (1978) Left and right in science and art. Leonardo 11:29-38.

Gross CG, Mishkin M (1977) The neural basis of stimulus equivalence across retinal translation. In: Lateralization in the nervous system (Harnad S, Doty RW, Jaynes J, Goldstein L, Krauthamer G, eds), pp 109-122. New York: Academic Press.

Gross CG, Schiller PH, Wells C, Gerstein GL (1967) Single-unit activity in temporal association cortex of the monkey. J Neurophysiol 30:833-843.

Gross CG, Bender DB, Rocha-Miranda CE (1969) Visual receptive fields of neurons in inferotemporal cortex of the monkey. Science 166:1303-1306.

Gross CG, Roche-Miranda CE, Bender DB (1972) Visual properties of neurons in inferotemporal cortex of the macaque. J Neurophysiol 35:96-111.

Hasselmo ME, Rolls ET, Baylis GC (1986) Object-centered encoding of faces by neurons in the cortex in the superior temporal sulcus of the monkey. Soc Neurosci Abstr 12:1369.

Iwai E, Mishkin M (1969) Further evidence on the locus of the visual area in the temporal lobe of the monkey. Exp Neurol 25:585-594.

Jolicoeur P, Gluck MA, Kosslyn SM (1984) Pictures and names: making the connection. Cognit Psychol 16:243-275.

Judge SJ, Richmond BJ, Chu FC (1980) Implantation of magnetic search coils for measurement of eye position: an improved method. Vision Res 20:535-538.

Kimura D (1963) Right temporal lobe damage. Arch Neurol 8:264-271.

Komatsu H, Ideura Y, Kaji S, Yamane S (1992) Color selectivity of neurons in the inferior temporal cortex of the awake macaque monkey. J Neurosci 12:408-424.

Lansdell H (1968) Effect of temporal lobe ablations on two lateralized deficits. Physiol Behav 3:271-273.

Logothetis NK, Pauls J, Buelthoff HH, Poggio T (1992) Evidence for recognition based on interpolation among 2D views of objects in monkeys. Invest Ophthalmol Vis Sci [Suppl] 34:1132.

Logothetis NK, Pauls J, Buelthoff HH, Poggio T (1993) Responses of inferotemporal (IT) neurons to novel wire-objects in monkeys trained in an object recognition task. Soc Neurosci Abstr 19:27.

Logothetis NK, Pauls J, Buelthoff HH, Poggio T (1994) View-dependent object recognition by monkeys. Curr Biol Vol. 4, no. 5 1994.

Marr D (1982) Vision. San Francisco: Freeman.

McLain DH (1974) Drawing contours from arbitrary data points. Comput J 17:318-324.

Mikami A, Kubota K (1980) Inferotemporal neuron activities and color discrimination with delay. Brain Res 182:65-78

Milner B (1958) Psychological defects produced by temporal-lobe excision. Res Publ Assoc Res Nerv Ment Dis 36:244-257.

Milner B (1968) Visual recognition and recall after right temporal-lobe excision in man. Neuropsychologia 6:191-209.

Milner B (1980) Complementary functional specialization of the human cerebral hemispheres. In: Nerve cells, transmitters and behaviour (Levy-Montalcini R, ed), pp 601-625. Vatican City: Pontificiae Academiae Scientiarium Scripta Varia.

Orton ST (1928) Specific reading disability—strephosymbolia. JAMA 90:1095-1099.

Pauls J, Bricolo E, Logothetis NK (1995) View invariant representations in monkey temporal cortex: Position, scale and rotational invariance. *In*: Early visual learning. New York: Oxford UP, in press.

Perrett DI, Rolls ET, Caan W (1979) Temporal lobe cells of the monkey with visual responses selective for faces. Neurosci Lett [Suppl] S3:S358.

Perrett DI, Rolls ET, Caan W (1982) Visual neurones responsive to faces in the monkey temporal cortex. Exp Brain Res 47:329-342.

Perrett DI, Smith PAJ, Potter DD, Mistlin AJ, Head AS, Milner AD, Jeeves MA (1985) Visual cells in the temporal cortex sensitive to face view and gaze direction. Proc R Soc Lond [Biol] 223.293-317.

Perrett DI, Mistlin AJ, Chitty AJ (1989) Visual neurones responsive to faces. Trends Neurosci 10:358-364.

Poggio T (1990) A theory of how the brain might work. Cold Spring Harbor Symp Quant Biol, vol. LV, 899-910.

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

Poggio T, Girosi F (1990) Regularization algorithms for learning that are equivalent to multilayer networks. Science 247:978-982.

Richmond BJ, Optican LM, Podell M, Spitzer H (1987) Temporal encoding of two-dimensional patterns by single units in primate inferior temporal cortex. I. Response characteristics. J Neurophysiol 57:132-146.

Robinson DA (1963) A method of measuring eye movement using a scleral search coil in a magnetic field. IEEE Trans Biomed Eng 101:131-145.

Rock I, DiVita J (1987) A case of viewer-centered object perception. Cognit Psychol 19:280-293.

Rock I, DiVita J, Barbeito R (1981) The effect on form perception of change of orientation in the third dimension. J Exp Psychol [Gen] 7:719-732.

Rock I, Wheeler D, Tudor L (1989) Can we imagine how objects look from other viewpoints? Cognit Psychol 21:185-210.

Rodman HR, Scalaidhe SPO, Gross CG (1993) Response properties of neurons in temporal cortical visual areas of infant monkeys. J Neurophysiol 70:1115-1136.

Rogers BJ, Graham M (1979) Motion parallax as an independent cue for depth perception. Percept Psychophys 8:125-134.

Rogers BJ, Graham M (1982) Similarities between motion parallax

and stereopsis in human depth perception. Vision Res 27:261-270.
Rogers BJ, Graham M (1983) Anisotropies in the perception of three-dimensional surfaces. Science 221:1409-1411.
Rogers BJ, Graham M (1985) Motion parallax and the perception of three-dimensional surfaces. In: Brain mechanisms and spatial vision (Ingle DJ, Jeannerod M, Lee DN, eds), pp 95-113. Dordrecht: Nijhoff.
Rosch E (1975) Cognitive representations of semantic categories. J Exp Psychol [Gen] 104:192-233.
Rosch E, Mervis CB, Gray WD, Johnson DM, Boyes-Braem P (1976) Basic objects in natural categories. Cognit Psychol 8:382-439.
Schwartz EL, Desimone R, Albright TD, Gross CG (1983) Shape recognition and inferior temporal neuron. Proc Natl Acad Sci USA 80:5776-5778.
Siegel RM, Andersen RA (1988) Perception of three-dimensional structure from motion in monkey and man. Nature 331:259-261.
Sry G, Vogels R, Orban GA (1993) Cue-invariant shape selectivity of macaque inferior temporal neurons. Science 260:995-997.
Tanaka K, Saito H-A, Fukada Y, Moriya M (1991) Coding visual images of objects in the inferotemporal cortex of the macaque monkey. J Neurophysiol 66:170-189.
Tarr M, Pinker S (1989) Mental rotation and orientation-dependence in shape recognition. Cognit Psychol 21:233-282.
Tarr M, Pinker S (1990) When does human object recognition use a viewer-centered referenceframe? Psychol Sci 1:253-256.
Taylor L (1969) Localization of cerebral lesions by psychological testing. Clin Neurosurg 16:269-287.
Tranel D, Damasio AR, Damasio H (1988) Intact recognition of facial expression, gender, and age in patients with impaired recognition of face identity. Neurology 38:690-696.
Ullman S (1989) Aligning pictorial descriptions: an approach to object recognition. Cognition 32:193-254.
Ullman S, Basri R (1991) Recognition by linear combinations of models. IEEE Trans Pattern Anal Mach Intell 13:992-1005
Vaughan HG, Gross CG (1966) Observations on visual evoked responses in unanesthetized monkeys. Electroencephalogr Clin Neurophysiol 21:405-406.
Vetter T, Poggio T, Buelthoff HH (1994) The importance of symmetry and virtual views in three-dimensional object recognition. Curr Biol 4:18-23.
Vetter T, Hurlbert AC, Poggio T (1995) View-based models of 3D object recognition: invariance to imaging transformations. Cereb Cortex 5:261-269.
Von Bonin G, Bailey P (1947) The neocortex of *Macaca mulatta*, 4th ed. Urbana: University of Illinois.
Yamane S, Kaji S, Kawano K, Hamada T (1987) Responses of single neurons in the inferotemporal cortex of the awake monkey performing human face discrimination task. Neurosci Res S5:S114.
Young MP, Yamane S (1992a) Sparse population coding of faces in the inferotemporal cortex. Science 256:1327-1331.
Young MP, Yamane S (1992b) An analysis at the population level of the processing of faces in the inferotemporal cortex. In: Brain mechanisms of perception and memory: from neuron to behaviour (Squire L, Ono T, Fukuda M, Perrett D, eds), pp 47-71. New York: Oxford UP.