STIMULUS-SELECTIVE PROPERTIES OF INFERIOR TEMPORAL NEURONS IN THE MACAQUE[1]

ROBERT DESIMONE,[*, 2] THOMAS D. ALBRIGHT,[‡] CHARLES G. GROSS,[‡] AND CHARLES BRUCE[§]

*Laboratory of Neuropsychology, National Institute of Mental Health, Bethesda, Maryland 20205, ‡Department of Psychology, Princeton University, Princeton, New Jersey 08544, and §Section of Neuroanatomy, Yale University School of Medicine, New Haven, Connecticut 06510

Abstract

Previous studies have reported that some neurons in the inferior temporal (IT) cortex respond selectively to highly specific complex objects. In the present study, we conducted the first systematic survey of the responses of IT neurons to both simple stimuli, such as edges and bars, and highly complex stimuli, such as models of flowers, snakes, hands, and faces. If a neuron responded to any of these stimuli, we attempted to isolate the critical stimulus features underlying the response. We found that many of the responsive neurons responded well to virtually every stimulus tested. The remaining, stimulus-selective cells were often selective along the dimensions of shape, color, or texture of a stimulus, and this selectivity was maintained throughout a large receptive field. Although most IT neurons do not appear to be "detectors" for complex objects, we did find a separate population of cells that responded selectively to faces. The responses of these cells were dependent on the configuration of specific face features, and their selectivity was maintained over changes in stimulus size and position. A particularly high incidence of such cells was found deep in the superior temporal sulcus. These results indicate that there may be specialized mechanisms for the analysis of faces in IT cortex.

Inferior temporal (IT) cortex plays a role in visual processing several steps beyond that of the primary visual cortex. Removal of IT cortex results in a severe impairment of visual discrimination learning and retention which exists in the absence of any changes in sensory thresholds, such as acuity (see reviews by Mishkin, 1966; Gross, 1972; Dean, 1976). IT neurons receive a converging input from visuotopically organized areas in prestriate cortex (Desimone et al., 1980; see Ungerleider, 1984), but, unlike neurons in either striate or prestriate cortex, IT neurons are not visuotopically organized. Rather, they have large receptive fields (median size 26° × 26°) that almost always include the center of gaze and usually extend into both the contralateral and ipsilateral visual fields (Gross et al., 1972; Desimone and Gross, 1979).

In previous studies, responses from IT neurons were elicited either with simple stimuli, such as white and colored bars, or more complex stimuli, including patterns, hands, or faces (Gross et al., 1972; Ridley and Ettlinger, 1973; Rolls et al., 1977; Jarvis and Mishkin, 1977; Ridley et al., 1977; Gross et al., 1979; Mikami and Kubota, 1980; Sato et al., 1980; Fuster and Jervey, 1982; Perret et al., 1982; Richmond and Optican,

1983). However, there have been no systematic attempts to determine the incidence of selectivity for particular stimulus dimensions, such as shape or color, or to analyze the apparent selectivity for complex stimuli, with the exception of the recent work of Perret et al. (1982) on face-selective cells.

In the present study, we attempted both to survey the responses of IT neurons to simple and complex stimuli and to study the properties of face-selective neurons in IT cortex in detail.

In the first part of this investigation, IT neurons were systematically tested with simple stimuli, such as bars varying in length and width, with two-dimensional shapes and patterns, and with three-dimensional objects. The objects provided a rich variety of colors, textures, shapes, and behaviorally relevant stimuli. If a neuron responded to any of the stimuli, we attempted to isolate the critical stimulus features underlying the neuron's response. In the second part of this study, we studied in detail a group of neurons in the superior temporal sulcus that appeared to respond selectively to faces.

We found that, as in other visual areas, most IT neurons respond to many different visual stimuli and, thus, cannot be narrowly tuned "detectors" for particular complex objects. Many IT neurons do appear to be selective along the dimensions of shape, color, or texture. We also found neurons that responded selectively to faces or hands. The responses of the hand- and face-selective neurons were dependent on the configuration of many specific face or hand features, and their selectivity was maintained over changes in stimulus size and position. These stimuli may be treated differently from other classes of visual patterns in IT cortex.

Materials and Methods

The methods of anesthesia, immobilization, single neuron recording, and histological localization of recording sites were identical to those previously reported in detail (Desimone and Gross, 1979) and will only be summarized here. The animals were implanted, under sterile conditions, with a recording chamber and a bolt for holding the head. They were recorded from 8 to 12 times, in sessions of 10 to 14 hr, over a 12-week period. During the recording sessions, the animals were anesthetized with N_2O and O_2 (70:30) and immobilized with pancuronium bromide. Only the contralateral eye was used. The pupil was dilated with cyclopentolate, and the cornea was covered with a contact lens chosen to focus the eye at 57 cm. The recording electrodes were made of varnish-coated tungsten with exposed tips of 10 μm or less. All recording sites were within the central portion of cytoarchitectonic area TE (Fig. 1).

Part 1: Survey of IT neurons. Two classes of stimuli were used. The first consisted of white or colored slits, edges, and dark bars, rear-projected onto a 70° × 70° tangent screen. The background luminance of the screen was 1.5 foot-lamberts (ft-L), the white stimuli had a luminance of 20 ft-L, and the dark stimuli had a luminance of 1.5 ft-L on a background of 20 ft-L. Red, yellow, green, and blue stimuli were produced by Wratten filters, numbers 29, 90, 65A, and 47B, respectively, placed in front of the projector. The white stimuli were 0.5 to 1.2 log units brighter than any of the colored stimuli. No attempt was made to equate the stimuli for luminance; however, for cells that responded better to a colored stimulus than to a (brighter) white stimulus, we tested whether the preference was for color or for a particular luminance level by varying the intensity of the white stimulus over a range of 2 log units.

The second class of stimuli that were used to test each cell for stimulus selectivity was more complex than the first and included a standard set of 14 three-dimensional objects. The set consisted of plastic models of a human head, a human hand, a red apple, a yellow banana, a yellow flower, a coiled black snake, and a large black spider as well as a fur-covered monkey doll, and black, blue, green, yellow, red, and white brushes. The objects were chosen to include objects desirable to monkeys (e.g., apple, banana), objects that elicit strong interest (e.g., face, hand), objects that elicit fear (e.g., spider, snake), and objects that had produced strong responses from IT neurons in our previous studies (e.g., brushes). All of the models except for the doll were fairly realistic.

If a cell responded selectively to one or more of the objects, the cell was then tested with color and black-and-white photographs of the objects, with variations of the objects or of their photographs, and with a variety of other complex two-dimensional patterns. The photographs

were printed at half, full, and twice the size of the objects. The variations of the objects and photographs included (*1*) similar objects (e.g., brushes with different shapes), (*2*) scrambled photographs of the objects (4 to 10 rearranged pieces), (*3*) isolated components of the photographs, and (*4*) high contrast, white or colored cutouts with the same external contours as the objects. In addition, other cutouts with complex boundaries, gratings, and textured patterns were also used.

All of the complex objects, shapes, and patterns could be shown either front-illuminated or as dark shadows on the tangent screen and could be presented either manually (held by a thin rod) or under computer control. The computer controlled either of two motorized arms (not visible to the animal) that held the stimuli. One arm moved front-illuminated stimuli in an arc in front of the tangent screen. The arc and the illumination source were centered at the animal's eye in order to minimize three-dimensional depth cues generated by the motion of the three-dimensional objects. The other arm was located behind the tangent screen and moved rear-illuminated stimuli that appeared as shadows on the screen. Each arm swept the stimuli along the horizontal meridian of the animal's visual field through a window 15° to 20° in width centered on the fovea. The stimulus velocity ranged from 1 to 5°/sec.

The procedure for testing each neuron was as follows. After the unit was isolated, it was tested with simple slits, edges, and bars with a hand-held projector. Stimuli were varied in length and width from 1° to 20° and were tested at a minimum of four orientations and four colors in addition to white and black. The optimal length, width, orientation, direction of motion, and color were determined, if possible. Since the precise borders of large IT receptive fields are often difficult to plot, we normally only verified that the receptive field included at least the central 15° of the contralateral or ipsilateral visual field. Each of the 14 standard objects was then tested repeatedly by hand, and the strength of the responses to each was estimated. If the unit responded differentially to any of the stimuli, computer-assisted testing was begun in order to confirm the selectivity and isolate the critical features for the unit's response. For example, if we initially judged that the unit responded best to the hand, it would then be tested with cutouts and shadows of the hand at different orientations and sizes, photographs of monkey hands, scrambled photographs of hands, gratings, and other stimuli. Stimuli were presented randomly interleaved, for a minimum of five trials each, in each of two directions. Histograms of the firing of the unit before, during, and after the stimulus presentation were computed. Because the responsivity of many IT units varies over time, comparisons were made only among stimuli that had been randomly interleaved with each other in the same block of trials.

Part 2: Neurons selectively responsive to faces. This experiment was directed exclusively at neurons that appeared to respond selectively to
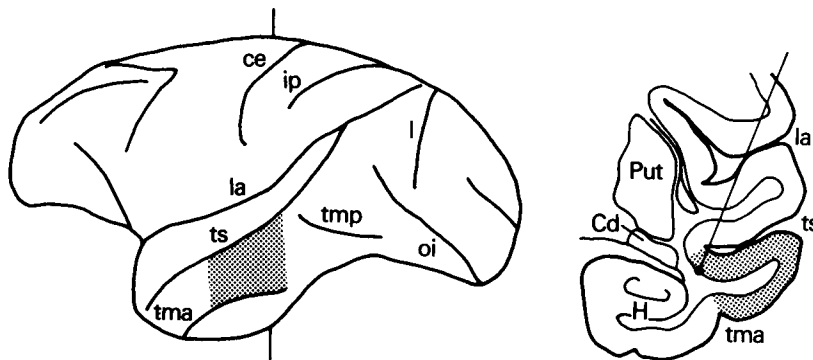

Figure 1. Location of recording sites. *Left*, Lateral view of macaque brain. The *stippled area* indicates location of recording sites on inferior temporal gyrus. Recording sites were also located in the lower bank of the superior temporal sulcus adjacent to the *stippled area*. *Right*, Section through inferior temporal cortex illustrating reconstructed electrode penetration. The location of the section is indicated by *vertical lines* on the brain drawing at *left*. The *stippled area* delimits the range of all recording sites within the superior temporal sulcus and on the inferior temporal gyrus. The electrode penetration shown is from experiment 2, in which a high proportion of cells responsive to faces was found. All other penetrations in exeriment 2 were located within 3 mm of the penetration shown. *ce*, central sulcus; *Cd*, caudate nucleus; *H*, hippocampus; *ip*, intraparietal sulcus; *l*, lunate sulcus; *la*, lateral sulcus; *oi*, inferior occipital sulcus; *Put*, putamen; *tma*, anterior middle temporal sulcus; *tmp*, posterior middle temporal sulcus; *ts*, superior temporal sulcus.

faces. All of these neurons were tested initially with the same set of complex objects and assortment of edges and bars used in the survey (part 1). In addition, all neurons were informally tested with various auditory and somesthetic stimuli. The neurons that responded strongly to the model of the face and weakly or not at all to any other stimulus initially tested were selected for further study. Except for sine-wave gratings presented on a cathode ray tube (CRT) screen, all stimuli used were photographic slides projected onto the tangent screen. The slides were prepared from photographs of human and monkey heads in orientations ranging from front view through profile to rear view. In addition, there were slides of faces with some components removed or rearranged. Neutral density filters were used to equate the stimuli for average luminance (10 ft-L). The stimuli were presented by a computer-controlled optical bench for 2.5 sec each, randomly interleaved, for a minimum of 10 trials each. All stimuli were initially presented at the center of gaze. Because many IT units recorded in paralyzed monkeys respond in a more sustained fashion to slowly moving stimuli than to flashed stimuli, the stimuli were presented with a slight oscillatory horizontal motion (1 Hz, 0.5° peak-to-peak).

Results

Experiment 1: Survey of IT neurons

One hundred and fifty-one units were sampled on 26 penetrations in three animals. Of these neurons, 13% (20 of 151) did not respond to any visual stimulus, and another 14% (21 of 151) gave responses that were so weak or inconsistent that it

TABLE I

Distribution of cell properties in experiment 1

Total Units (151)
Responsive (110)
Selective (66)
Shape only (20)
Shape + texture (2)
Color only (7)
Color + shape (2)
Texture only (1)
Texture + color (1)
Hands (2)
Faces (3)
Unclassified (28)
Nonselective (44)
Unresponsive or responses too weak to study (41)

was impossible to investigate their stimulus specificity. The remaining 73% (110 of 151) gave sufficiently strong and consistent responses to allow study of their stimulus specificity, and all further analyses in experiment 1 are based on this sample (see Table I). Eighty-eight percent of these units responded only with an increase in firing rate, 7% responded only with a decrease in rate, and 5% reponded with an increase or decrease in rate, depending on the stimulus.

A large minority of the responsive units, 41% (44 of 110), did not clearly respond selectively to any of the stimuli we tested; that is, they responded similarly to all of the complex objects and showed no clear preference for the size, shape, orientation, direction of motion, or color of slits, edges, or bars (Fig. 2). None of these units responded well to a diffuse light flash, indicating at least a small degree of underlying spatial antagonism within the receptive field. In addition, a few of the units classified as nonselective showed no preferences among the three-dimensional objects and no preferences among the projected stimuli but gave somewhat stronger responses to each of the objects than to any of the projected stimuli.

The remainder of the responsive units, 59% (66 of 110), responded selectively to the different stimuli. The selectivity of most units for a particular stimulus was relative rather than absolute. Over half (61%) of the selective units gave at least a weak response to every stimulus tested, and the remainder generally responded to most of the stimuli tested. Complex stimuli in particular were more likely to elicit a good response. Most selective units (73%) responded better to one or more of the complex objects than to any slit or bar tested. Most of the selective neurons appeared to be sensitive to stimulus shape, color, texture, or combinations of the three. Other neurons responded selectively to different complex stimuli, but extensive testing with all available stimuli did not reveal which stimulus features were critical for the responses. All classes of selective responses are described in the following sections (see Table I).

Shape. Twenty-four (36%) of the selective units were sensitive to stimulus shape, i.e., the length, width, or other boundary features of a plane closed figure. Ten of these cells were selective for the length and width of a light or dark bar swept through the receptive field. The optimal width was always narrower than the receptive field, and the optimal length was the length of the receptive field or shorter. In these respects, the cells resembled complex or hypercomplex cells in striate cortex;
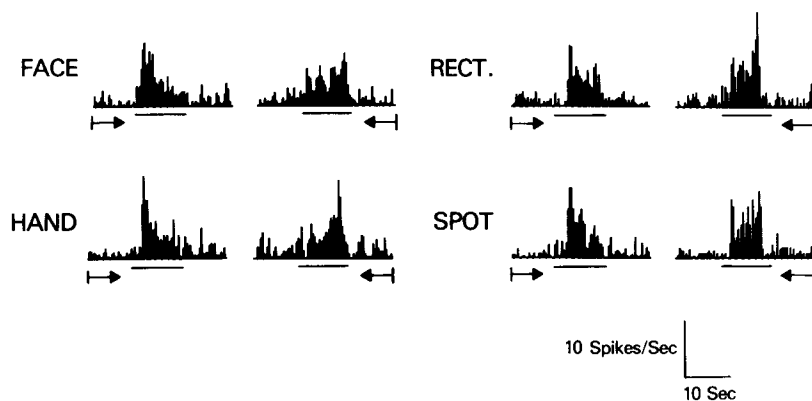

Figure 2. Responses of a unit that responded comparably to all stimuli tested, including light and dark bars, spots, and a set of complex objects. Responses shown were to a model of a human face, a model of a human hand, a 2 × 10° white rectangle, and a 5° white spot. The stimuli were swept at 1.7°/sec along the horizontal meridian and were visible to the monkey only when moving through a 15° "window" centered on the fovea. The *bar* under each *histogram* indicates when the stimulus was in the window. *Arrows* indicate the direction of stimulus motion and the direction of time in the *histograms*. The ipsilateral visual field is located on the *left* of each *histogram* and the contralateral visual field on the *right*.

however, the receptive fields were much larger than that of striate neurons, and, unlike complex cells, the IT cells showed little or no orientation or direction specificity.

The remainder of the shape-selective units (14) could not be driven with simple slits of light or bars. During the initial manual testing, all responded to one or more complex objects and would have been classified as unresponsive if the complex objects had not been used. After we determined the best object for each unit, we tested the unit's responses both to a plain cutout whose outer boundary had the same shape as the best object and to other cutouts with different shapes. The cutouts were presented either front-illuminated or as dark shadows. Units were classified as shape-selective if their response varied with the shape of the cutouts. The responses of a typical shape-selective unit are shown in Figure 3. The receptive field of the unit was at least 20° × 20°. It did not respond to any projected edge or bar tested. During the initial manual testing, the unit responded to the objects with irregular edges, such as the hand and brush, but not to the objects with smooth edges, such as the apple. A brush and a variety of cutouts with different shapes were then placed on the computer-controlled track and were presented as rear-illuminated shadows on the tangent screen. The unit responded best to the shadow of the brush
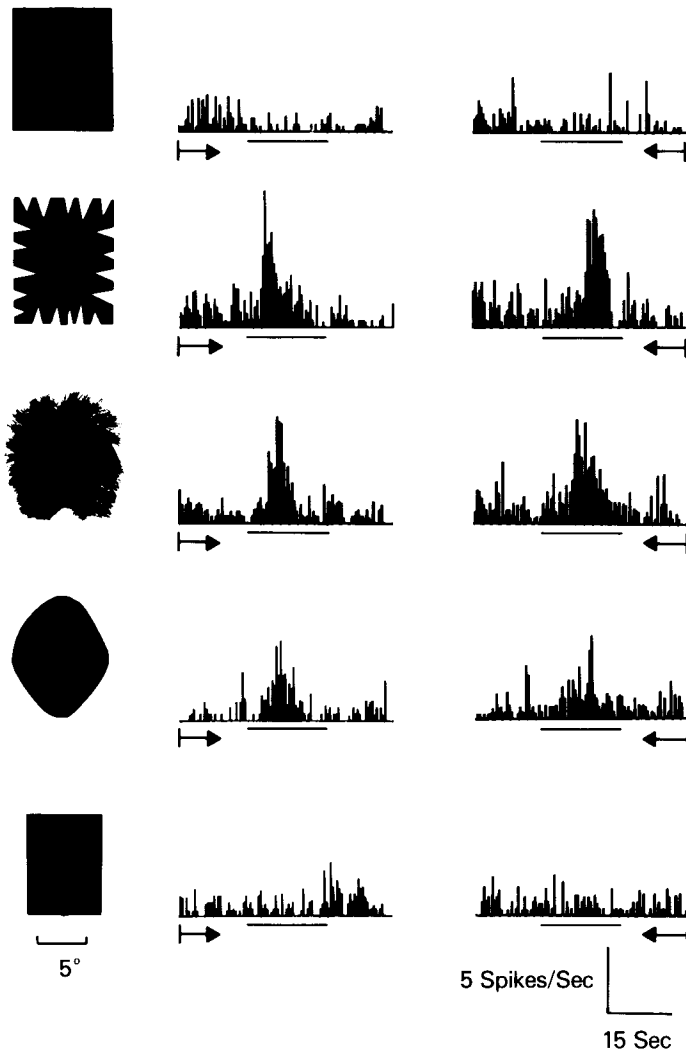

Figure 3. Responses of a unit selective for patterns with irregular edges. Stimuli were presented as dark shadows on a tangent screen, moving at 1.2°/sec. Stimuli with irregular edges elicited the best responses, while stimuli with straight edges were ineffective. See also the legend to Figure 2.

and to a cutout with irregular edges but not at all to rectangular shapes with straight edges. Since responses to the shapes with irregular edges could be elicited anywhere within a large receptive field, the selectivity of the unit was apparently due to a global property of the receptive field and not to any local receptive field structure.

Whereas the unit described in Figure 3 responded to any shape with irregular edges, a few units responded only to shapes with particular complex edges. All of the units maintained their selectivity throughout their large receptive field, but the center of gaze was the most sensitive portion of the field. Four of the shape-selective cells were also sensitive to either color or texture, in addition to shape.

Color. Ten (15%) of the selective units were selective for stimulus color. We classified a unit as color-selective if it responded more strongly to a particular broad-band colored stimulus than to a white stimulus of any luminance. Six of the color-selective units could be driven with a colored bar. The remaining four units that did not respond to a colored bar required either a specific colored complex shape, colored texture, or combination of colors (see Fig. 4).

Texture. Four units (6%) did not respond to any bar tested but did respond to random dot patterns and patches of artificial hair or fur, which we interpreted as selectivity for some type of texture. In each case, the unit responded to a small textured pattern (2° to 3° wide) throughout the central 15° of the visual field, although the best response, as with most IT units, was to the stimulus positioned at the center of gaze. Defocusing the stimuli, by placing a two-diopter lens in front of the animal's eye, completely eliminated the response; however, this apparent selectivity for fine texture was not simply a preference for high spatial frequencies *per se*, as none of these units responded to square-wave gratings tested over a range of 0.25 to 16.0 cycles/degree at several orientations. Three of the units were also sensitive to either shape or color, in addition to texture.

Unclassified selective units. Twenty-eight (42%) of the selective units responded selectively to subsets of the complex objects and weakly or not at all to edges and bars, but we were unable to determine which specific features of the complex stimuli were related to the responses. The objects which elicited a response from these units shared no obvious stimulus feature or "meaning" (e.g., a brush and a hand) which distinguished them from the objects that were ineffective. Furthermore, we were unable to elicit a response from any of the simple shapes, colors, or textures we tried. Although we had the impression that these cells were selective for stimulus features, the dimensions along which they were selective eluded us.

Units selective for specific objects. Previous studies have reported a few units in IT cortex selective for faces (Perret et al., 1982) or more rarely for hands (Gross et al., 1972). Because such highly selective units were not searched for systematically throughout IT cortex, there was a possibility that a much larger proportion of IT units might be selective for specific objects. In the present study, we sampled units throughout a wide region of IT cortex, testing a variety of complex objects, but found only 5 of 151 units that appeared to respond selectively to specific objects. Two of these units responded best to hands and three to faces. The results from units that responded best to hands will be described next, and the results from units that responded best to faces will be described in the following section.

An example of an experiment on one of the two units that responded best to hands is illustrated in Figure 5. In the initial manual testing, the unit responded to the model of the hand but not to any of the other complex objects, nor to bars, edges, or gratings tested at several orientations. The unit responded to the hand throughout a receptive field at least 20° wide, with the best response occurring when the hand was positioned over
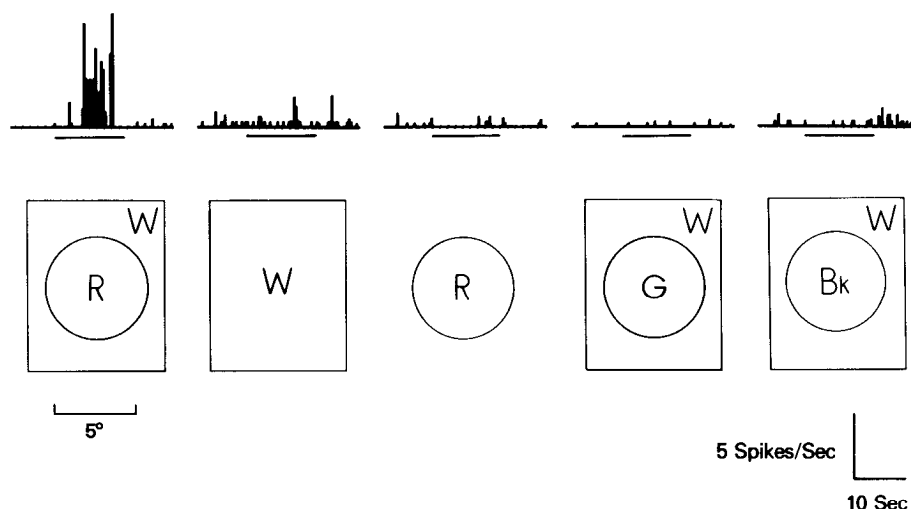
Figure 4. Responses of a unit selective for red and white color combination. Stimuli were constructed of colored paper. *W*, white (50 ft-L); *R*, red (35 ft-L), *G*, green (37 ft-L); *Bk*, black (2 ft-L). The background had a luminance of 5 ft-L. Stimuli were moved at 2.0°/sec from the contralateral into the ipsilateral visual field and were visible within a 15° window centered on the fovea. The unit responded only to the combination red and white moving stimulus. There was no response to the moving red stimulus alone, even if the luminance of the white static background was increased to the luminance of the white stimulus.

the fovea. Furthermore, the unit responded to photographs of human and monkey hands of any orientation, and of any size from 2.5° to 10° wide. The unit was then tested with a variety of complex objects, photographs, and cutouts, using computer-controlled presentation. The unit responded strongly to the model of the hand, both front and back, but not at all to other models, such as faces. White cutouts with the shape of the hand also elicited strong responses, but not quite as strong as to the colored, three-dimensional model. Thus, it is likely that the response to the model of the hand was based primarily on its shape but the color or internal detail of the model probably contributed to the unit's response. Consistent with this interpretation, there was a small response to a scrambled photograph of the hand, which preserved the overall color and many of the internal details of the hand but not its shape. The unit's response to the shape of the hand appeared to be very specific since "grating-like" cutouts elicited no response, even though they mimicked the periodicity of the fingers of the hand cutout as well as the general pattern of "spokes" emanating from one end of a central "core." By contrast, a cutout of a monkey hand, which was much smaller than the cutout of the human hand and had less prominent fingers, elicited a good response. Fingers proved to be a critical feature of the hand cutout, since filling in the spaces between them completely eliminated the response. Finally, the unit was tested with the cutout of the human hand swept through the receptive field in different orientations and directions. All orientations elicited good responses, as did all directions of motion except for one of the directions in which the base of the hand cutout was the leading stimulus in the receptive field. The other unit that responded selectively to hands gave similar results.

In summary, the two units that responded best to hands were very sensitive to the shape of the hand, not responding at all to similar (but not hand-like) shapes. Furthermore, the selectivity of the units for the shape of the hand was largely invariant over changes in position, size, and orientation. Finally, adding the color and internal details of a hand to the shape of the hand seemed to improve the units' response.

Part 2: Neurons selectively responsive to faces

In part 1, we surveyed the responses of IT units to a large variety of complex objects and found three units that appeared to respond best to faces. These units were each found in separate, widely spaced penetrations. We did not find other such units in their immediate vicinity, making detailed analysis of the phenomenon difficult. Because cells selective for faces had been found previously in the superior temporal sulcus (Bruce et al., 1981; Perret et al., 1982), we decided to search in this region and subsequently located a small zone within the sulcus where we consistently found face-selective cells on adjacent penetrations. This zone was located between 5 and 7 mm anterior to the interaural plane near the junction of the lower bank of the sulcus with the fundus (see Fig. 1). The zone was within IT cortex, ventral to the superior temporal polysensory area (Bruce et al, 1981).

Fifty units were sampled on 10 penetrations. All units were tested initially with the same set of complex objects and assortment of edges and bars used in the survey. Of these units, 17 (34%) responded strongly to the model of the face in the set of complex objects and weakly or not at all to any other stimulus tested. The remaining units did not respond well to any stimulus tested. No unit responded to auditory or somesthetic stimuli.

All of the cells that initially responded strongly to the model of the face in the set of complex objects were subsequently tested with a variety of different monkey and human faces, some of which had components removed or rearranged, and were also tested informally with a real human face. In addition, eight of the cells were tested with sine-wave gratings presented on a CRT screen. Because the properties of the three face-selective cells recorded in the survey (part 1) did not appear to differ significantly from those of the 17 additional cells of part 2, they are included in the following analyses.

All of the face-selective cells responded to all faces tested, including the real human face. None of the responses to faces were critically dependent on the exact position of the stimulus within the central 20° of the visual field nor on the stimulus size. Furthermore, when components such as the eyes or snout were removed, the responses to a face were always reduced, but no single component ever elicited the same response as the complete face. None of the cells tested responded to sine-wave gratings.

The results from the cell illustrated in Figure 6 are representative of many of the face-selective cells. In the initial
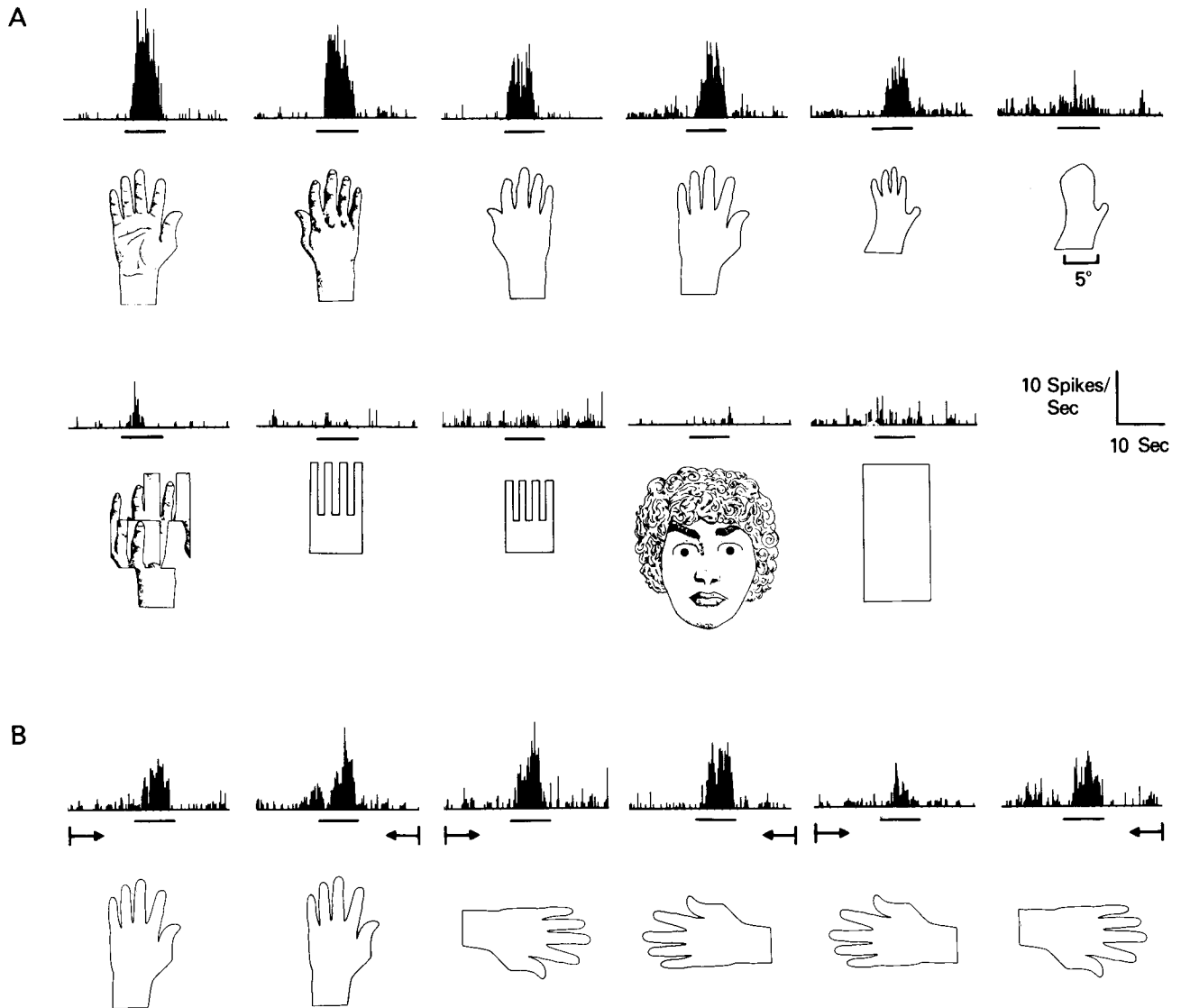
A

B

Figure 5. Responses of a unit that responded more strongly to hands than to any other stimulus tested. *Drawings* under each *histogram* were traced from the stimuli. *A*, Comparison of responses to hands versus other patterns. Stimuli were randomly interleaved and included the front and back of a model of a human hand, white cutouts with the same shape as the human hand model, a cutout with the shape of a monkey hand, a cutout of a monkey's hand with the space between the fingers eliminated, a scrambled photograph of the model of the human hand (10 rearranged pieces), two "grating-like" hands, a model of a human face, and a plain rectangle. Stimuli were moved at 1.2°/sec from the contralateral into the ipsilateral visual field and were visible within a 15° window centered on the fovea. *B*, Responses to a stimulus with the shape of a hand, in different orientations. The contralateral visual field is represented on the *left* of each *histogram* and the ipsilateral field on the *right*. The *arrows* indicate the direction of stimulus motion and the direction of time in the *histograms*. Other conditions were as in *A*.

Figure 6. Responses of a unit that responded more strongly to faces than to any other stimulus tested. *A*, Comparison of responses to faces, to faces with components removed, and to a hand. Stimuli were photographic slides, presented for 2.5 sec, indicated by the *bar* under each *histogram*. All stimuli were centered on the fovea. *Drawings* under each *histogram* were traced from stimuli. *1*, Monkey face in natural color; *2*, same monkey face with components rearranged (four pieces); *3*, second monkey face in color; *4*, same monkey face with snout removed; *5*, eyes removed; *6*, color removed; *7*, human face; *8*, hand. The *bar graph* at *top left* indicates summed responses to each stimulus. Responses were computed from the firing rate of the unit during the stimulus presentation minus the average firing rate before the stimulus presentation. *0* represents the base line firing rate. Removing any component of the face reduced the response, while scrambling the components eliminated the response. *B*, Responses to a monkey face in different degrees of rotation. All stimuli were colored slides; other conditions were the same as in *A*. Responses decreased as the face was rotated from frontal to profile view. As in *A*, removing the eyes from the frontal view reduced the response. *C*, Responses to faces in different locations within receptive field. The stimulus was the same as in *B* (frontal view of face). The stimulus was centered on the horizontal meridian of the visual field. *I*, Ipsilateral visual field; *C*, contralateral visual field; *FOV*, fovea. The best response was to the stimulus positioned over the fovea. *D*, Responses to sine-wave gratings and bars. Gratings ranged from 0.25 to 8 cycles/degree, and bars were 0.1° wide. All stimuli had a vertical orientation. Gratings were drifted at 1 cycle/sec for 5 sec, and bars were moved at 1°/sec. Stimuli were generated on a 10° diameter CRT display. *LB*, Light bar; *DB*, dark bar. There was no response to either bars or gratings of any frequency. See also the legend to Figure 5.
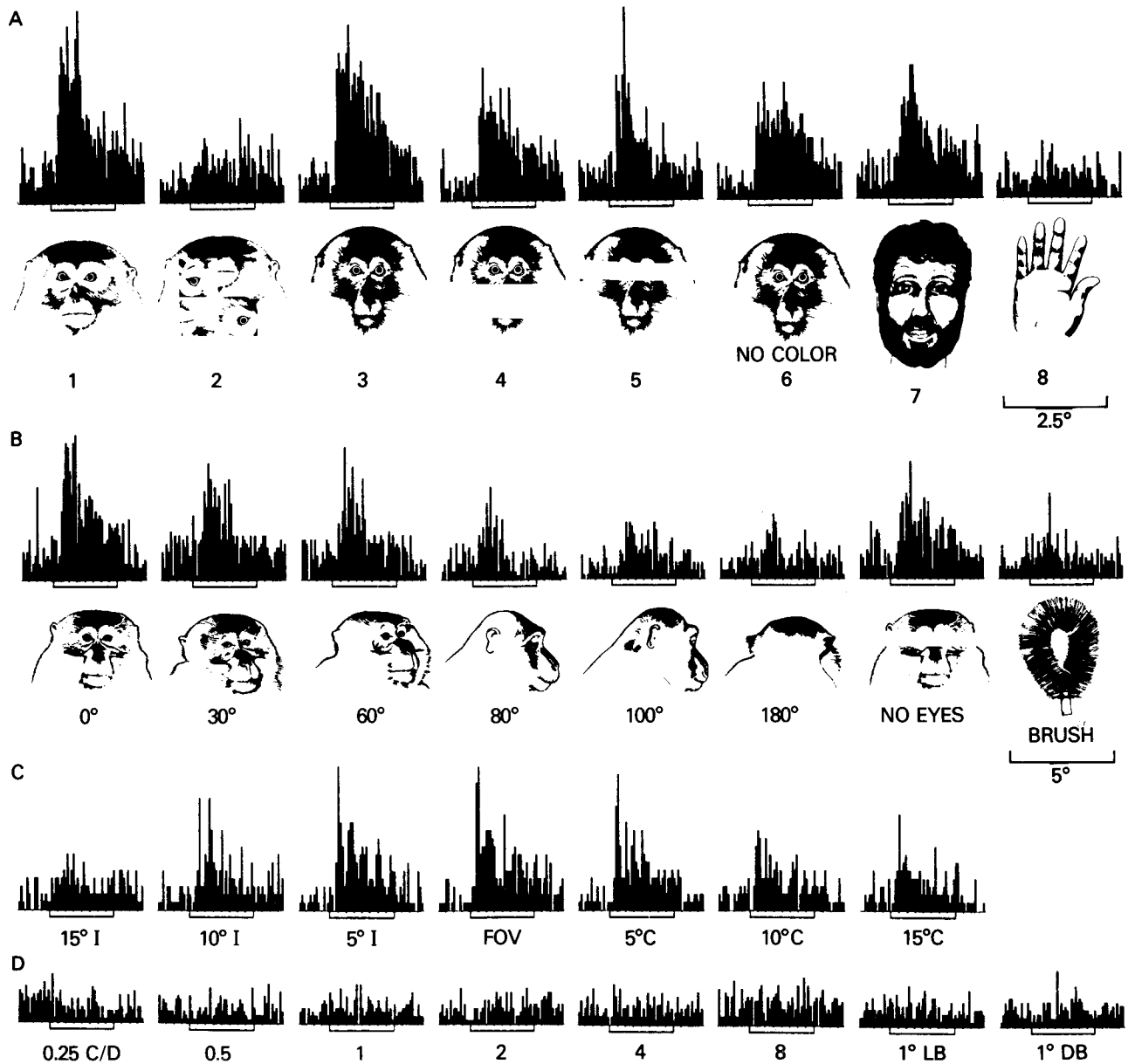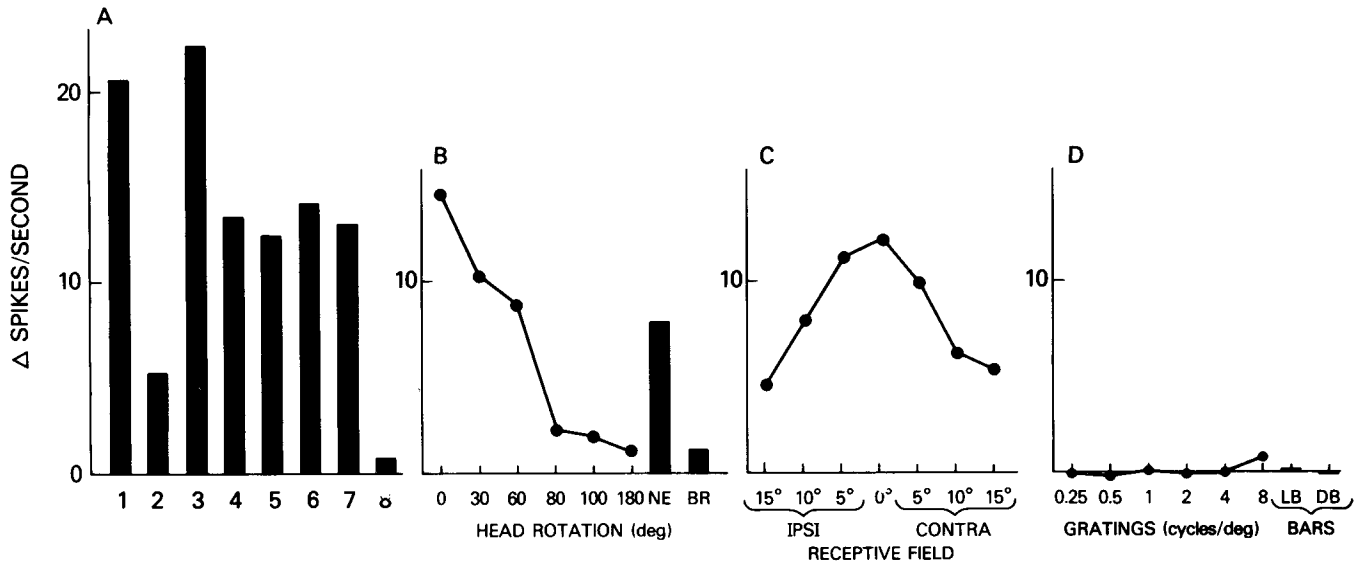
Figure 6

manual testing, the unit responded to faces anywhere within a receptive field 40° to 60° wide that was centered on the fovea. It responded very weakly to other complex objects and not at all to bars or edges. To investigate the apparent selectivity of the cell for faces, we tested the cell with computer-controlled presentations of several series of stimuli.

One series of stimuli consisted of projected color slides of a human face, a human hand, and monkey faces, some with components removed or rearranged (Fig. 6A). The two different monkey faces tested elicited strong responses; these responses were reduced about 40%, but were not eliminated, by removing either the eyes, the snout, or the color. Thus, the eyes, snout, and color each contributed to the response, but no one component alone was sufficient to explain the response to the original face. In addition, scrambling the internal features of the face greatly reduced the response, indicating that a particular configuration of the internal features was essential. The response to the human face was less than the response to the monkey face; however, the human face was larger than the monkey face, and this may have contributed to the difference. The hand elicited almost no response. Likewise, sine-wave gratings (Fig. 6D), presented as a separate series of stimuli, elicited essentially no responses. We also presented a set of shape-descriptor stimuli that, in a separate study (Desimone et al., 1982; Schwartz et al., 1983), had proved effective in driving many IT units, but these stimuli were completely ineffective in driving this particular neuron.

The cell was also tested with monkey faces shown in different orientations (Fig. 6B). The cell responded well (but not quite as strongly) to the front view of the monkey face in this series, even though the face was twice the size of those presented in the first series. The cell was maximally sensitive to the front view of a monkey face, and its response fell off as the head was rotated into the profile view. A common property of all but 2 of the 17 units recorded in part 2 (but none of the three face-selective units recorded in part 1) was sensitivity to the three-dimensional orientation of the face, some units responding best to the front view of the face and some to the profile.

The cell responded to a face throughout its large receptive field, but the best response was to a face centered on the fovea (Fig. 6C). A projected slide of a monkey face centered on the fovea and at 5°, 10°, and 15° in both the contralateral and ipsilateral visual fields elicited the best response when it was on the fovea and progressively weaker responses as it was moved into the periphery. Even at the more peripheral visual field locations, however, the cell responded only to faces and not to other stimuli. All of the cells selective for faces had large receptive fields but responded best to faces presented at the fovea.

Figure 7 shows the responses of another face-selective neuron, in this case one that responded selectively to profiles of faces. The unit responded to two profile views that differed in size but not at all to the back of the head or to the front of the face with or without eyes (Fig. 7A). There was no response to one-half of a frontal view, to an outline of a profile filled with uniform grey (which preserved the outer boundary shape of the head but no internal details), to an outline filled with fur, or to an outline filled with the features of a frontal view (Fig. 7B). Likewise, the neuron did not respond to either the anterior or posterior portions of the face profile when either portion was presented separately. Thus, the response of this neuron, like the neuron decribed above (Fig. 6), appeared to depend on the overall configuration of many facial features.

Whereas some of the cells that responded selectively to faces were very sensitive to the orientation of heads in depth, e.g., a frontal versus a profile view, none of the cells, of 10 tested, were very sensitive to the orientation of faces in the frontal plane, e.g., upright versus sideways. An example is shown in Figure 8. The cell responded best to human and monkey faces

in profile, and, like other such profile-sensitive cells, this was true whether the profile was facing into the contralateral or ipsilateral visual field. Moreover, the cell maintained its selectivity for profiles even when the head was rotated in the frontal plane, so that the profile was facing either upward or downward. Thus, the cell appeared to be sensitive to the orientation of the head in depth, independent of its orientation in the frontal plane.

In summary, we studied 20 cells (parts 1 and 2) that appeared to be selective for faces. Each cell responded to human and monkey faces throughout its large receptive field. Removing any component or feature of the face, such as the eyes, snout, color, or anterior or posterior portion of the head, always reduced the response. Rearranging the internal components virtually eliminated the response. Many of the units were sensitive to the orientation of the head in depth (frontal view versus profile), but none of the units tested was very sensitive to the orientation of the face in the frontal plane.

Discussion

Using both simple and complex stimuli, we attempted to identify some of the critical features underlying the stimulus selectivity of IT neurons. To our knowledge, this is the first systematic survey of stimulus selectivity in IT cortex. We found that even when tested with a large variety of both simple and very complex stimuli, many IT neurons either did not respond well to any or reponded similarly to all. Like neurons in other visual areas, most of the remaining, stimulus-selective IT cells responded to many visual stimuli and, thus, were not narrowly tuned for a specific stimulus. Many of the selective cells were selective for shape, color, texture, or combinations of the three, and they maintained their selectivity throughout their large receptive fields. Finally, we found a small proportion of IT units that appeared to respond selectively to face or hands. In the remainder of the "Discussion" we will consider each class of IT unit in detail.

Unresponsive or not selective. The proportion of units unresponsive to any visual stimulus we tested, 13%, was very similar to the proportion of unresponsive units reported in a previous study (Gross et al., 1972). In addition to unresponsive units, we found another 14% of the units to be so poorly driven that it was not possible to study their stimulus selectivity. A possible explanation for the large proportion of unresponsive or poorly responsive units might be the anesthesia or conditions of immobilation; however, in a recent study of IT units in awake, behaving monkeys, a similar proportion of poorly responsive or unresponsive units was found (Richmond et al., 1983). Another possibility is that some poorly reponsive cells were actually selective for stimuli that we never tested. Consistent with this possibility is the fact that a few IT neurons were found that were so highly selective for a particular stimulus that they would have been classified as unresponsive if that stimulus had not been tested.

Of the responsive units, a large proportion (25%) was not clearly stimulus-selective. As is the case with the unresponsive units, such apparently nonselective units might have responded better to a stimulus that we never tested. Alternatively, these cells might have a behavioral function, such as directing attention, that is unrelated to the specific stimulus within the receptive field.

Cells selective for shape, color, or texture. Half of the selective units were selective for shape, color, or texture or combinations of the three. Some of these units were sensitive to the length, width, or color of a bar and, thus, resembled units that have been found in striate and prestriate cortex but with much larger receptive fields. Other selective IT units did not respond well to simple rectangular stimuli, such as bars. A few such units only responded to textured patterns, and others only reponded to complex shapes. Since this study was completed, we have
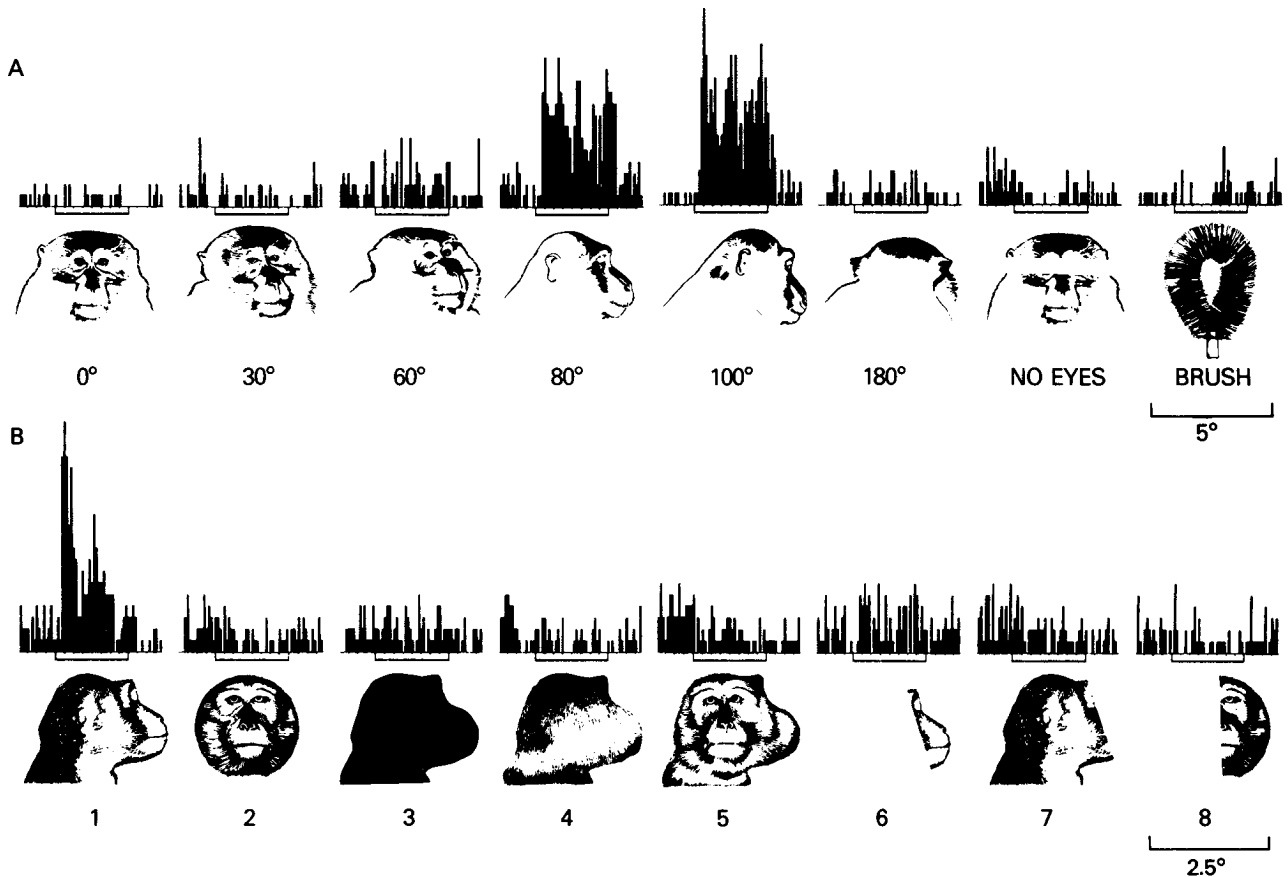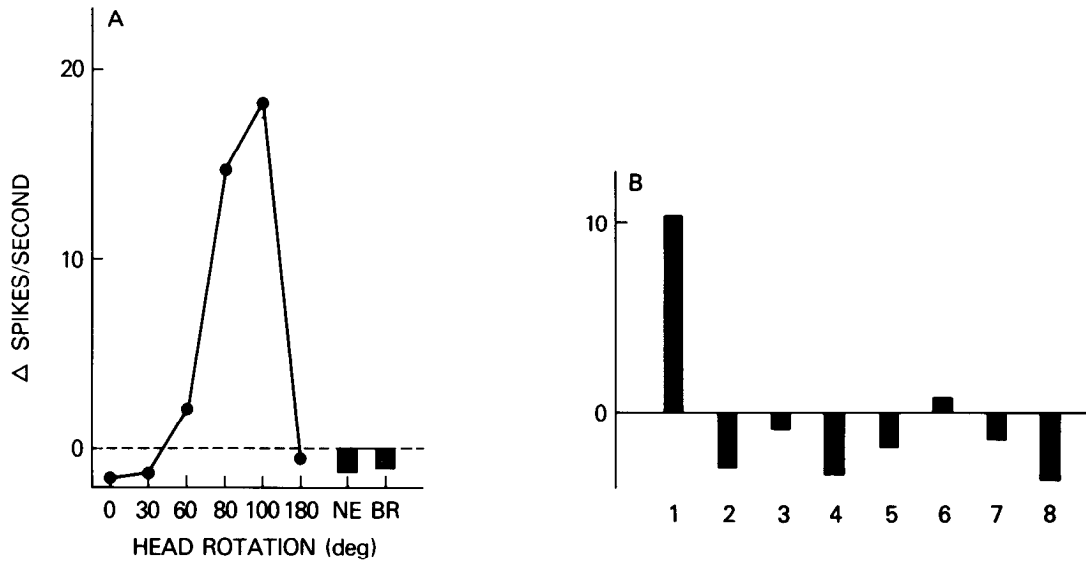
Figure 7. Responses of unit that responded better to profiles of faces than to any other stimulus tested. *A*, Responses to a monkey face in different degrees of rotation. Best responses were to the profiles of the face. *B*, Responses to profile of face and to profile with components removed or altered. *1*, Profile of face; *2*, frontal view of face inscribed in circle; *3*, outline of profile filled with uniform grey; *4*, outline of profile filled with "fur"; *5*, outline of profile filled with features from the frontal view; *6*, anterior portion of face profile; *7*, posterior portion of face profile; *8*, one-half frontal view of face. Removing or altering any of the components of the profile eliminated the response. In both *A* and *B*, stimuli were color slides. For conditions of presentation and figure conventions, see Figure 6.

investigated shape selectivity in IT cortex with a set of parametric shape descriptors (Schwartz et al., 1983). We confirmed that many IT units are sensitive to stimulus shape and found that their sensitivity to shape can be described in terms of a quantifiable shape dimension. Thus, shape appears to be a particularly salient stimulus feature in IT cortex.

Units selective for faces or hands. In 1969 we first reported an IT unit that appeared selective for the shape of a monkey hand (Gross et al., 1969), and 3 years later we reported photographs of faces to be the best stimuli we could find for a few IT cells (Gross et al., 1972). Although these reports encouraged speculation that object recognition was mediated by IT cells
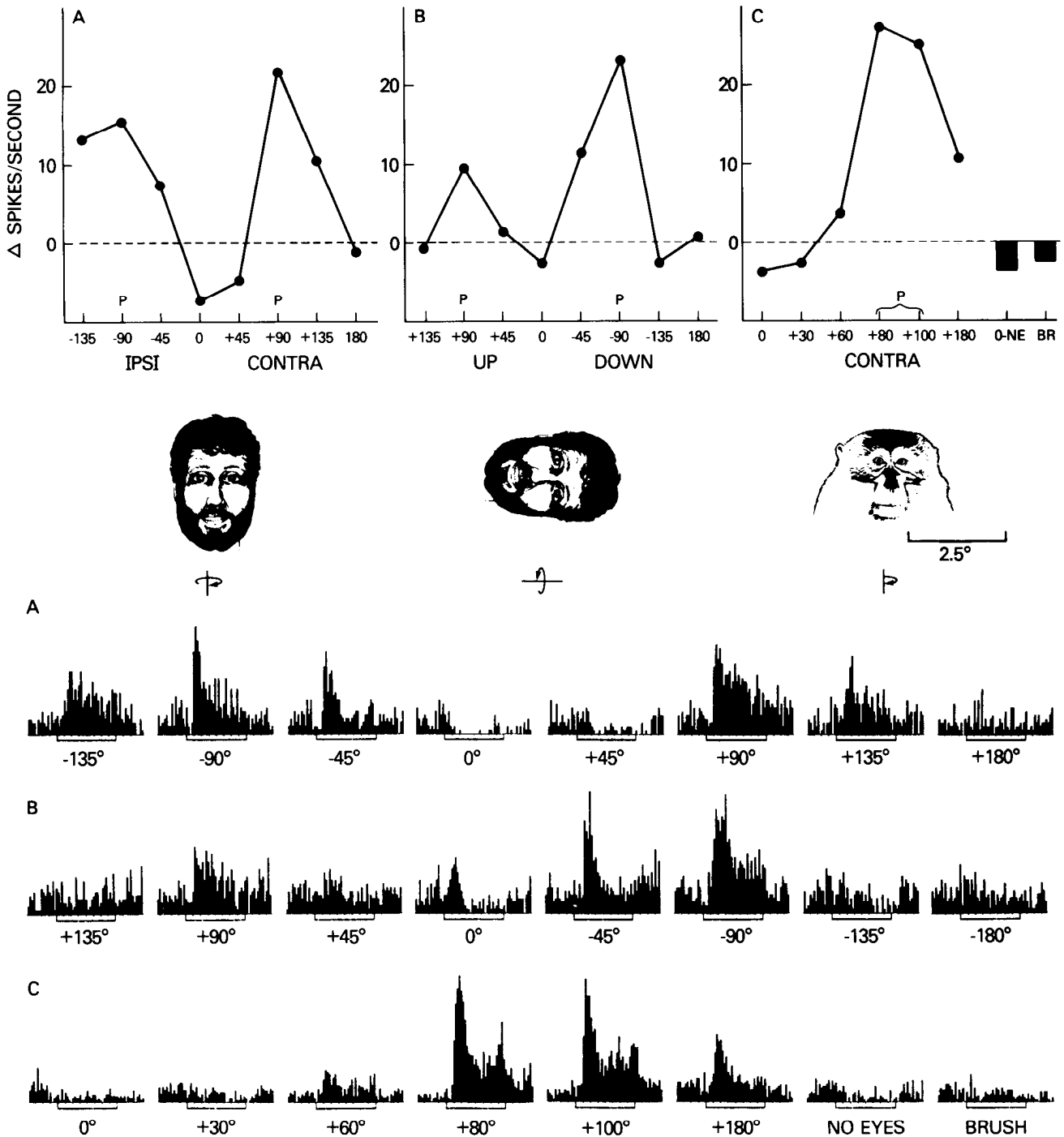
Figure 8. Responses of unit that responded better to profiles of faces than to any other stimulus tested. *A,* Responses to an upright human face presented with different rotations about the center of the head. *0°,* frontal view; *+90°,* profile facing into contralateral field; *+180°,* back of head; *−90°,* profile facing into the ipsilateral field. The unit responded to the face in profile view, indicated as *P* on the *bar graph. B,* Unit continued to respond to profile view, even with head on its side. *C,* Responses to monkey face in profile view, for comparison. A frontal view of a monkey's face without eyes and a brush were also presented, as controls. See also the legend to Figure 6.

selective for specific objects, there were no subsequent attempts to search for such cells systematically and the phenomenon remained unanalyzed. The results of the present study suggest that cells selective for specific objects are rare throughout IT cortex. Nonetheless, we found two hand-selective cells which we studied in much greater detail than previously and with computer-controlled, randomized presentation of stimuli. Furthermore, we found an additional 20 cells which responded selectively to faces. A study by Perret et al. (1982), initiated concurrently with this one, has also reported face-selective cells, in a portion of the superior temporal sulcus partially

overlapping with IT cortex (see the discussion of location below). They recorded in awake, behaving animals and had no control over the location of the stimulus on the retina. However, the central findings in the two studies are the same, and each serves as a replication of the other under different conditions. (An exception was our observation of consistency of the response over changes in location, which they were unable to test.)

For several reasons the responses of the hand- and face-selective units in the present study were unlikely to have been due to general arousal. First, none of the units responded to

intense somesthetic or auditory stimuli. Second, although the models of the snake and spider elicited strong emotional reactions from monkeys tested in their home cage, none of the units in our sample selective for faces or hands responded to these other stimuli. Third, general arousal, or even arousal confined to visual stimuli, could not explain why some units were selective for hands and other units for faces. Indeed, among the face-selective units, some units responded preferentially to the frontal view of the face and other units to the profile.

It is also unlikely that the face- and hand-selective units were actually selective for some simple, local stimulus feature such as a line of a particular length, or a specific patch of color, or a certain texture in a particular retinal location. Each of these units responded to a wide variety of hands or faces, including real ones, plastic models, and photographs or slides. Furthermore, each responded selectively to both monkey and human faces or hands over a wide range of sizes and orientations throughout a large receptive field. It is hard to imagine a specific "simple" stimulus feature that would remain invariant over such transformations. Conversely, none of the units responded well to pictures of hands or faces that had been scrambled, even though many component lines and edges remained intact.

To what extent might it be more appropriate to describe the two hand-selective cells as selective for a specific shape rather than for a hand? The two hand-selective units that were studied in detail responded best to the outline or shadow of a variety of monkey and human hands but not to many other shapes. As the shape of the hand was altered so that it looked to us less like a hand, the responses of the units declined or disappeared. For example, a grating-like hand, which mimicked the periodicity of the hand's fingers and the general property of spokes radiating from one side of a central core, elicited a much smaller response than more hand-like shapes. Still, one might argue that it was the curvature of the fingers in combination with their periodicity that was the crucial feature for the neurons' responses and not "handness" *per se*. Similarly, although the responses of these units were enhanced when the hand had a skin color and appropriate internal details, this might have reflected a color or texture specificity of the cells. At this point, however, the list of the stimulus conditions that elicit the maximum response begins to approach a description of the essential features of a hand, i.e., "four long and one short curved spokes, all emanating from one end of a central core, flesh colored, and covered with fine texture." While a simpler sensory feature may eventually be found that would explain the properties of such cells, at present the most appropriate description of their properties appears to be that they respond best to hands.

Analogous arguments apply to the cells apparently selective for faces. Faces contain numerous shapes, contours, colors, and textures, and all of these components must be arranged in a limited set of configurations if they are to look like a face. Every face-selective unit we studied was sensitive to many features of a face. Removing the eyes, snout, outline of the head, or color always reduced the response. Furthermore, the overall configuration was crucial. Scrambling the components of the face always eliminated the unit's response. Thus, many of the components of the face appeared to contribute to the unit's selectivity, and the list of stimulus requirements approached a general description of a face.

The face-selective cells reported in part 2 were clustered in the portion of IT cortex that lies deep in the lower bank and fundus of the superior temporal sulcus. We have previously found similar cells within the superior temporal sulcus in the polysensory area located immediately dorsal to IT cortex in the fundus and upper bank of the sulcus (Bruce et al., 1981). Unlike cells in IT cortex, over half of the cells in the superior temporal sulcus can be activated through more than one sensory modal-

ity. In the present study, we distinguished the face-selective cells in IT cortex from those in the superior temporal polysensory area on the basis of both their location in the sulcus and their exclusively visual responses (and those of cells in their vicinity). Thus, face-selective cells may be distributed across two adjacent areas within the sulcus. Others have also found face-selective cells concentrated within the superior temporal sulcus, within what we would consider to be both the superior temporal polysensory area and IT cortex (Perret et al., 1982). As we have never found such a high local concentration of face-selective cells in any other portion of the cortex, part of the superior temporal sulcus may be relatively specialized for the analysis of faces. Interestingly, the superior temporal sulcus is heavily interconnected with the amygdala (Aggleton et al., 1980), where face-selective cells have also been found (Sanghera et al., 1979). Rolls (1981) has suggested that these sites may be parts of a distributed system involved in the analysis of faces.

IT cortex and neural theories of pattern recognition. Numerous ablation experiments have demonstrated that IT cortex is crucial for normal pattern recognition. Therefore, it is appropriate to ask how the properties of IT neurons bear on neural theories of pattern recognition. More than 20 years ago Hubel and Wiesel (1962) suggested that a hierarchical sequence of receptive field types in striate cortex forms the first stages of the neural basis of pattern recognition. According to their original scheme, as one progresses from simple to complex to hypercomplex cells, receptive field size becomes larger, the exact location of the sitmulus within the receptive field becomes less crucial, and there is greater stimulus specificity. Although the organization of striate cortex is now known to be rather more complicated than Hubel and Wiesel's original hierarchy, at least the first two of the general trends they proposed do appear to continue in IT cortex, namely, an increase in receptive field size and less specificity for stimulus location. What is the function of such large receptive fields? Because IT neurons respond to a stimulus thoughout a receptive field much larger than the stimulus, it has been suggested that IT neurons mediate the perceptual equivalence of visual objects across retinal translation (Gross and Mishkin, 1977; Seacord et al., 1979). In addition, in the present study we found neurons that responded to stimuli regardless of changes in stimulus size within the receptive field; this has been confirmed for a much larger sample of IT neurons in a separate study (Schwartz et al., 1983). As with changes in position, changes in stimulus size normally cause changes in the absolute firing level of the neuron; thus, the absolute response of an IT neuron rarely exhibits size or position constancy. Nonetheless, to the extent that the relative preference for a particular stimulus is maintained over changes in size and position within the receptive field, IT neutrons may participate in the mechanism of size and position constancy.

What about the third trend within Hubel and Wiesel's hierarchy, namely, increased specificity? Many IT cells responded equally to nearly every stimulus tested, and most of the stimulus-selective cells gave at least a small response to virtually every stimulus tested, especially visually complex stimuli. Thus, as in other visual areas (see review by Desimone et al., 1984), most IT neurons are not narrowly tuned for a specific stimulus; the image of a visually complex stimulus on the central retina will probably activate most of the cells in IT cortex. Presumably, therefore, the neural representation of objects in IT cortex is reflected in the pattern of activity across a population of cells and not in the activity of individual cells that respond only to specific objects. The cells selective for faces, and more rarely for hands, may be an exception to this scheme.

Special place for the analysis of faces. Why should faces be treated differently from other objects by IT cortex? Primates have evolved an elaborate facial musculature that permits a wide range of facial expression. Given the importance of faces

in primate social communication, we can speculate that there may also have been selective pressure to evolve neural mechanisms for the analysis and recognition of faces and facial expression. Consistent with this hypothesis is the evidence from humans that faces are perceived as special stimuli at birth (Field et al., 1982), that there are special perceptual mechanisms for the analysis of faces in adults (Yin, 1969), and that specific perceptual deficits in the recognition of faces (prosopagnosia) may follow from occipitotemporal lesions (Meadows, 1974; Whiteley and Warrington, 1977; Nardelli et al., 1982).

There are at least two examples of specialized neural mechanisms that have evolved to facilitate social communication in other species. One is the specialized structures that mediate the perception and generation of song in birds (Leppelsack and Vogt, 1966; McCasland and Konishi, 1981), and another is the cortex specialized for language in man. In each case, specialized perceptual mechanisms have evolved that are separate from those of audition in general. In man, the supramodal language cortex of the temporal lobe is located within the second temporal convolution, adjacent to the auditory association cortex (Rasmussen and Milner, 1975). Interestingly, this appears to be the location of the superior temporal sulcus in the monkey, which contains both face-selective and polysensory cells. Thus, this portion of the primate brain may be a fertile zone for the development of supramodal mechanisms for communication.

References

Aggleton, J. P., M. J. Burton, and R. E. Passingham (1980) Cortical and subcortical afferents to the amygdala in the rhesus monkey (*Macaca mulatta*). Brain Res. *190:* 347–368.

Bruce, C. J., R. Desimone, and C. G. Gross (1981) Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque. J. Neurophysiol. *46:* 369–384.

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

Desimone, R., and C. G. Gross (1979) Visual areas in the temporal cortex of the macaque. Brain Res. *178:* 363–380.

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

Desimone, R., E. L. Schwartz, T. D. Albright, and C. G. Gross (1982) Inferior temporal neurons selective for stimulus shape. Assoc. Res. Vis. Ophthalmol. Abstr. *22:* 238.

Desimone, R., S. J. Schein, and T. D. Albright (1984) Form, color, and motion analysis in prestriate cortex of macaque monkey. In *Study Group on Pattern Recognition Mechanisms*, C. Ghagas, ed., Pontifical Academy of Sciences, Vatican City, in press.

Field, T. M., R. Woodson, R. Greenberg, and D. Cohen (1982) Discrimination and imitation of facial expressions by neonates. Science *218:* 179–181.

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.

Gross, C. G. (1972) Visual functions of inferotemporal cortex. In *Handbook of Sensory Physiology*, R. Jung, ed., Vol. VIII/3B, pp. 451–482, Springer-Verlag, Berlin.

Gross, C. G., and M. Mishkin (1977) The neural basis of stimulus equivalence across retinal translation. In *Lateralization in the Nervous System*, S Harned, R. Doty, J. Jaynes, L. Goldberg and G. Krauthamer, eds., pp. 109–122, Academic Press, New York.

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

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.

Gross, C. G., D. B. Bender, and G. L. Gerstein (1979) Activity of inferior temporal neurons in behaving monkeys. Neuropsychologia *17:* 215–229.

Hubel, D. H., and T. N. Wiesel (1962) Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. J. Physiol. (Lond.) *160:* 106–154.

Jarvis, C. D., and M. Mishkin (1977) Responses of cells in the inferior temporal cortex of monkeys during visual discrimination reversal. Soc. Neurosci. Abstr. *3:* 564.

Leppelsack, H. J., and M. Vogt (1966) Responses of auditory neurons in the forebrain of the songbird to stimulation with species-specific sounds. J. Comp. Physiol. *107:* 263–274.

McCasland, J., and M. Konishi (1981) Interactions between auditory and motor activities in an avian song control nucleus. Proc. Natl. Acad. Sci. U. S. A. *178:* 7815–7819.

Meadows, J. C. (1974) The anatomical basis of prospagnosia. J. Neurol. Neurosurg. Psychiatry *37:* 489–501

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

Mishkin, M. (1966) Visual mechanisms beyond the striate cortex. In *Frontiers of Physiological Psychology*, R. Russel, ed., pp. 77–123, Academic Press, New York.

Nardelli, E., F. Buonanno, G. Coccia, A. Fiaschi, H. Terzian, and N. Rizzuto (1982) Prosopagnosia: Report of four cases. Eur. Neurol. *21:* 289–297.

Perret, D. I., E. T. Rolls, and W. Caan (1982) Visual neurones responsive to faces in the monkey temporal cortex. Exp. Brain Res. *47:* 329–342.

Rasmussen, T., and B. Milner (1975) Clinical and surgical studies of the cerebral speech areas in man. In *Cerebral Localization*, K. J. Zulch, O. Creutzfeldt, and G. C. Galbraith, eds., pp. 238–257, Springer-Verlag, Berlin.

Richmond, B. J., and L. M. Optican (1983) Encoding of 2-D patterns by single units in inferior temporal cortex. Assoc. Res. Vis. Ophthalmol. Abstr. *24:* 229.

Richmond, B. J., R. H. Wurtz, and T. Sato (1983) Visual responses of inferior temporal neurons in the awake rhesus monkey. J. Neurophysiol. *50:* 1415–1432.

Ridley, R. M., and G. Ettlinger (1973) Visual discrimination performance in the monkey: The activity of single cells in infero-temporal cortex. Brain Res. *55:* 179–182.

Ridley, R. M., N. S. Hester, and G. Ettlinger (1977) Stimulus- and response-dependent units from the occipital and temporal lobe of the unanesthetized monkey performing learnt visual tasks. Exp. Brain Res. *27:* 539–552.

Rolls, E. T. (1981) Processing beyond the inferior temporal visual cortex related to feeding, memory, and striatal function. In *Brain Mechanisms of Sensation*, Y. Katsuki, R. Norgren, and M. Sato, eds., pp. 241–269, John Wiley & Sons, New York.

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

Sanghera, M. K., E. T. Rolls, and A. Roper-Hall (1979) Visual responses of neurons in the dorsolateral amygdala of the alert monkey. Exp. Neurol. *63:* 610–626.

Sato, T., T. Kawamura, and E. Iwai (1980) Responsiveness of inferotemporal single units to visual pattern stimuli in monkeys performing discriminations. Exp. Brain Res. *38:* 313–319.

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

Seacord, L., C. G. Gross, and M. Mishkin (1979) Role of inferior temporal cortex in interhemispheric transfer. Brain Res. *167:* 259–272.

Ungerleider, L. G. (1984) Contrasts between the corticocortical pathways for pattern and spatial vision. In *Study Group on Pattern Recognition Mechanisms*, C. Chagas, ed., Pontifical Academy of Sciences, Vatican City, in press.

Whiteley, A. M., and E. K. Warrington (1977) Prosopagnosia: A clinical, psychological and anatomical study of three patients. J. Neurol. Neurosurg. Psychiatry *40:* 395–403.

Yin, R. K. (1970) Face-recognition by brain-injured patients: A dissociable ability? Neuropsychologia *8:* 395–402.