chapter four

Object Recognition

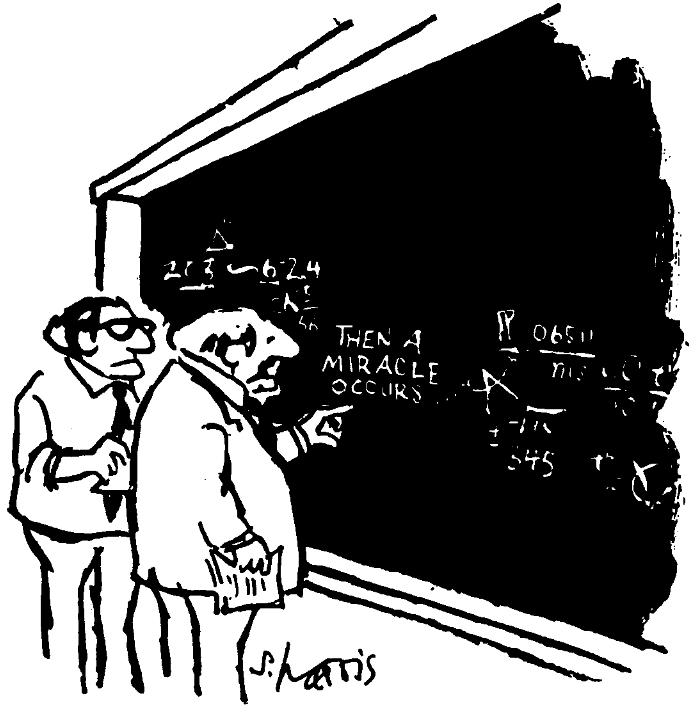
M.J. Farah: The cognitive Neuroscience of Vision. OUP, 2000

4.1 Object representation in inferior temporal cortex: a miracle occurs

The visual representations of the retina, LGN, and the occipital lobe are all retinotopic images. Retinotopy is a ubiquitous organizing principle for the representations of early and intermediate vision. But as we saw in the last chapter, the information that is explicitly available in such representations is not particularly useful for object recognition. Images bundle together the true shape of an object and the perspective from which it is viewed, whereas the identity of the object is of course related only to the former.

Accordingly, the neural substrates of visual recognition are not among the retinotopic areas just mentioned. Instead, they are located in inferior temporal areas in both monkey and man. Lesions to this area have devastating effects on animals' performance in tasks testing object perception, and on human object recognition after neurological disease or injury. The results of single unit recordings in IT are consistent with this. Compared to V4 and the visual areas preceding it, neurons in inferotemporal cortex show considerable constancy over changes in viewing conditions, and virtually no retinotopy.

How can visual representation change so radically going from V4 to IT, just one synapse away? This transformation, from image to object, is reminiscent of the famous cartoon shown in figure 4.1. In this chapter I will try to better characterize the miracle, if not fully explain it, calling upon lesion and single unit recording studies in monkeys, and lesion and neuroimaging studies in humans.



"I think you should be more explicit here in step two."

Figure 4.1 Sidney Harris's classic cartoon, which about sums up our understanding of the neural information processing performed between V4 and IT.
From Harris, "I think you should be more . . ." in Chalk Up Another One: The Best of Sidney Harris, *New Brunswick, NJ, Rutgers University Press, 1992; copyright Sidney Harris.*

4.2 The neural bases of shape recognition in monkeys

Lesion studies in monkeys

The experimental study of temporal cortex and visual object recognition dates back to the research of Kluver and Bucy (1937), on what is

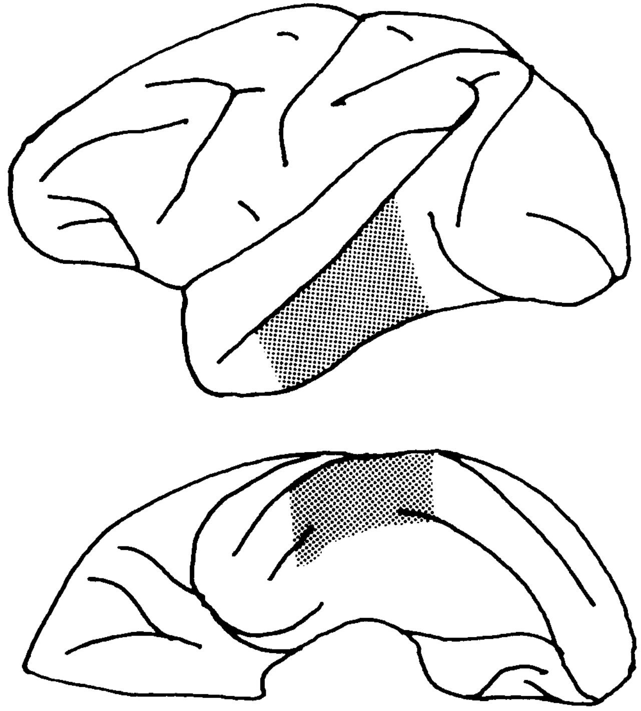


Figure 4.2 Inferotemporal cortex in the monkey brain.
From P. Dean, "Visual behaviour in monkeys with inferotemporal lesions," in D. J. Ingle et al. *(eds),* Analysis of Visual Behavior, *Cambridge, MA, MIT Press, 1982.*

now known as the Kluver–Bucy syndrome. These researchers removed the entire temporal lobes of monkeys bilaterally, and found complex changes in social, sexual, and eating behavior of the animals. Later research attempted to fractionate the syndrome and relate specifically visual impairments to specific cortical regions within the temporal lobe. The inferior temporal gyrus, also known as inferotemporal cortex or von Bonin and Bailey area TE, was eventually shown to be the critical area for producing the visual deficits (Mishkin, 1954, 1966; Mishkin and Pribram, 1954). Figure 4.2 shows the location of this area in the macaque brain.

The functional role of inferotemporal cortex in vision was initially conceptualized in terms of visual learning, rather than visual object recognition as discussed so far in this chapter. However, this difference has more to do with terminology and with the particulars of the experimental tasks used in these early laboratories than with any substantive distinction between the visual abilities impaired in the monkeys and what we would call visual object recognition.

In the typical experimental paradigm, monkeys were trained to respond differentially to one stimulus, the target stimulus, presented in advance of or alongside other "choice" stimuli. The animal would be required to press a response button under the choice stimulus that matched the target in order to obtain a reward, and performance was typically measured in terms of number of learning trials to reach a criterion. Compared to normal monkeys and operated control monkeys, monkeys with inferotemporal lesions showed severe impairments in these tasks (e.g., Blum *et al.*, 1950; Mishkin, 1966; Pribram, 1954). Assessment of the visual fields, acuity, and visual thresholds of these monkeys showed that the impairments could not be attributed to elementary visual sensory impairments. Assessment of discrimination learning in modalities other than vision confirmed the specificity of the impairments for visual discrimination learning (see Plaut and Farah, 1990, for a more detailed review).

Two other early findings suggest that the impairment of IT-lesioned monkeys is not in visual learning *per se*, but in object representation. First, IT lesions cause a loss of previously acquired visual discriminations (e.g., Pribram, 1954). This finding is more clearly analogous to an impairment of visual object recognition, in that the monkeys have lost knowledge of familiar objects. Second, IT-lesioned monkeys show qualitative as well as quantitative abnormalities in their visual discrimination learning, and these qualitative abnormalities are suggestive of an inability to represent visual objects *per se*, as opposed to position, size, brightness, or local features. For example, they may generalize their responses on the basis of just one dimension or feature of the target stimulus (e.g., Butter, Mishkin and Rosvold, 1965; Butter, 1968; Iwai, 1985), and have been noted to ignore shape altogether in favor of brightness (Iverson and Weiskrantz, 1967).

Iwai (1985) showed that even when IT-lesioned monkeys had succeeded in learning to discriminate between the target stimulus of a triangle and the wrong choice of a circle, they were not doing so on the basis of shape *per se*. Instead, they seemed to be responding to the parallelism between the base of the triangle and the edge of the board

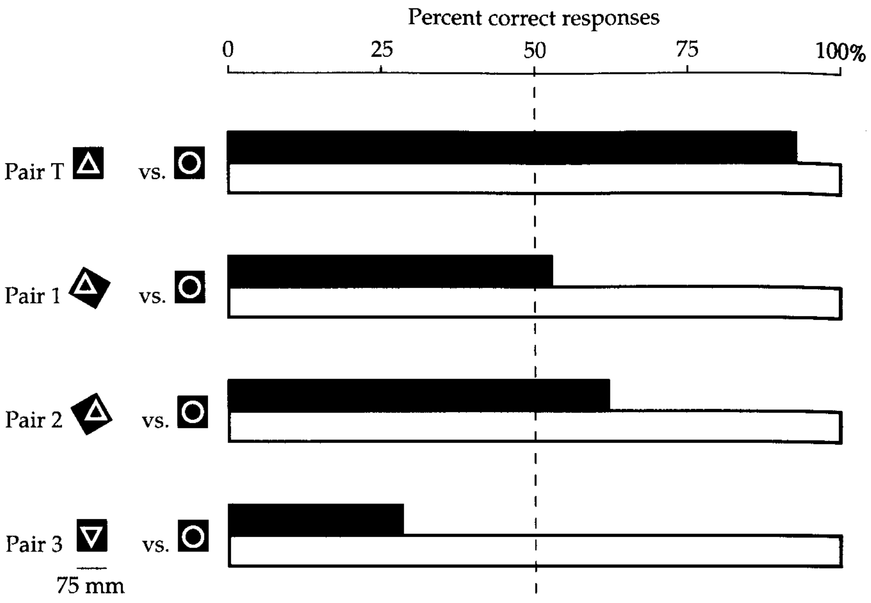


Figure 4.3 Performance of IT-lesioned monkeys in a visual discrimination task, showing their inability to transfer a learned discrimination to displays of the same shapes when the local spatial characteristics of the stimuli were changed. *From E. Iwai, "Neuropsychological basis of pattern vision in macaque monkeys,"* Vision Research, *25, 1985, with permission from Elsevier Science.*

on which the stimuli were presented. When the board was rotated, but the stimuli remained in the same orientation, the monkeys could no longer perform the discrimination. Furthermore, as shown in figure 4.3, Iwai found that these monkeys showed transfer of learning to new discriminations when the spatial location of some of the earlier patterns' features is maintained, but not when the same features were shifted in space. This suggests that, to the extent that the IT-lesioned monkeys could learn a discrimination, they were treating it as a spatial discrimination rather than a shape discrimination.

Once the general hypothesis of defective object representation after IT lesions was accepted, inquiry moved on to the next stage: What is the nature of stimulus representation in IT? Much research in the 1970s and subsequently has been aimed at answering this question. The general approach has been to infer which stimulus properties are

normally encoded (or not encoded) in IT representations by showing which stimulus properties IT-lesioned monkeys are impaired at using (or not impaired at using) as a basis for discrimination. On the basis of our current knowledge, a reasonable short answer might be: IT represents aspects of the intrinsic shape of a stimulus that are useful for recognition, and omits most aspects of stimulus appearance that depend on viewing conditions. Only a few representative studies will be reviewed here. More detail can be found in Plaut and Farah (1990).

Position is one visual property that is clearly a red herring for purposes of object recognition, and normal monkeys will easily generalize a visual discrimination learned in one hemifield to the other. Monkeys with bilateral IT lesions are impaired at this generalization, however (Gross and Mishkin, 1977; Seacord, Gross, and Mishkin, 1979). This implies that they have lost representations in which position is not represented, in other words, representations that are general across positions. The retinal image size of the stimulus is another visual property that depends on viewing conditions and not just intrinsic object geometry, and this is another property that IT-lesioned monkeys have trouble ignoring. For example, Humphrey and Weiskrantz (1969) trained monkeys to discriminate disks of two absolute sizes, varying their distance and hence their retinal image size. IT-lesioned monkeys were unable to relearn the discrimination, instead responding on the basis of retinal size or distance. This implies that IT representations normally encode the absolute size of an object, an intrinsic object property useful for recognition, rather than its distance *per se* or its retinal image size.

IT-lesioned monkeys are also impaired at generalizing across views of the same stimulus in a different orientation (Weiskrantz and Saunders, 1984), implying that orientation is yet another of the incidental image properties that has been laundered out of IT representations. A discrepant finding was reported by Holmes and Gross (1984), who found no impairment in discriminating size- and orientation-transformed stimuli, but this may have to do with the relatively simple nature of the discriminations (J vs. π, or P vs. T, pairs which can be distinguished on the basis of a local feature such as a hook or a closed loop) and the fact that only a single positive stimulus (hence feature) had to be learned by the monkeys. Finally, variations in illumination prevent IT-lesioned monkeys from seeing the equivalence of objects (Weiskrantz and Saunders, 1984), implying that IT representations are unaffected by patterns of shadow and light falling on object surfaces. IT-lesioned monkeys show little or no impairment in tasks that require discriminating between (as opposed to generalizing between) patterns and their mirror images (Gross, 1978), suggesting that handedness is yet another dimension over which the normal observer tends to generalize on the basis of object representations in IT.

4.3 Single unit studies in monkeys

The technique of single cell recording was applied to inferotemporal cortex by Charles Gross and collaborators beginning in the late 1960s. Early recordings from anaesthetized animals showed large bilateral receptive fields responsive to visual stimuli (e.g., Gross, Schiller, Wells, and Gerstein, 1967). Although some cells responded well to virtually any visual stimulus, others responded with some degree of selectivity to shape, color, or texture (see Desimone, Schein, Moran, and Ungerleider, 1985, for a review). Unlike cells in V4, the main source of input to IT, cells in inferotemporal cortex are not retinotopically organized (Desimone and Gross, 1979), but tend to cluster in groups with similar response properties (Fuster and Jervey, 1982). Recordings from awake animals have shown the ways in which neuronal activity is dependent on task demands. The responses of IT cells are enhanced during visual discrimination, compared to when the monkey need not perform any actions contingent on the stimulus (Richmond and Sato, 1987), and become larger and more selective as the difficulty of the discrimination increases (Spitzer, Desimone, and Moran, 1988). However, IT cells do not carry motivational information *per se*; they are not sensitive to the association of a stimulus with reward (Rolls, Judge, and Sanghera, 1977). Most striking was the observation that some cells in IT are tuned to highly specific aspects of stimulus shape. For example, Gross, Rocha-Miranda and Bender (1972) recorded from a cell that responded vigorously to a monkey hand, with diminished responses to increasingly different-shaped stimuli, as shown in figure 4.4.

Although the finding of a hand-selective cell was met with surprise and outright skepticism at the time, many different laboratories have subsequently observed IT cells with highly selective responses for particular patterns and objects (e.g., Baylis, Rolls, and Leonard, 1985; Desimone, 1991; Miyashita, Date, and Okuno, 1993; Perrett, Mistlin, and Chitty, 1987; Tanaka, Saito, Fukada, and Moriya, 1991; Yamane, Kaji, and Kawano, 1988). Figure 4.5 shows examples of the shapes for which neurons in IT show selectivity.

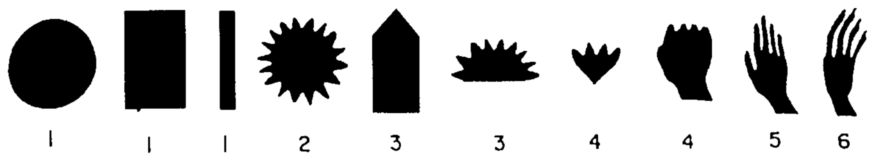


Figure 4.4 The range of stimuli used to test the selectivity of a "hand cell" in monkey IT cortex. The more different the stimulus shape from a monkey hand, the smaller the cell's response.
From C. G. Gross et al., "Visual properties of neurons in inferotemporal cortex of the macaque," Journal of Neurophysiology, *35, 1972.*

In many ways, these neurons appear to be representing objects. One manifestation of this is their general preference for real objects: they respond more vigorously to three-dimensional objects or models of objects than to their outline silhouettes (Desimone, Albright, Gross, and Bruce, 1984). Indeed, they are selective for objects and may be relatively nonselective for Adelson and Bergen's (1991) "stuff": Sary, Vogel, and Orban (1993) identified neurons that were selectively responsive to a particular shape defined by luminosity differences (e.g., a white star on a black background) and found that they were also responsive to the same shape defined by texture cues and motion cues (e.g. a star-shaped region of speckles with the same average luminosity as its background, defined by larger, sparser speckles or speckles moving in a different direction).

Many IT cells are selective for faces, some even showing selectivity for one face over another (Baylis, Rolls, and Leonard, 1985). "Face cells" cease to respond if the features of the face are present but scrambled (Desimone *et al.*, 1984), suggesting that the overall structure of the face is important, and not simply the presence of local features. This conclusion was strengthened by a quantitative study in which neuronal responses were best predicted by combinations of various inter-feature distances within the face (Yamane, Kaji, and

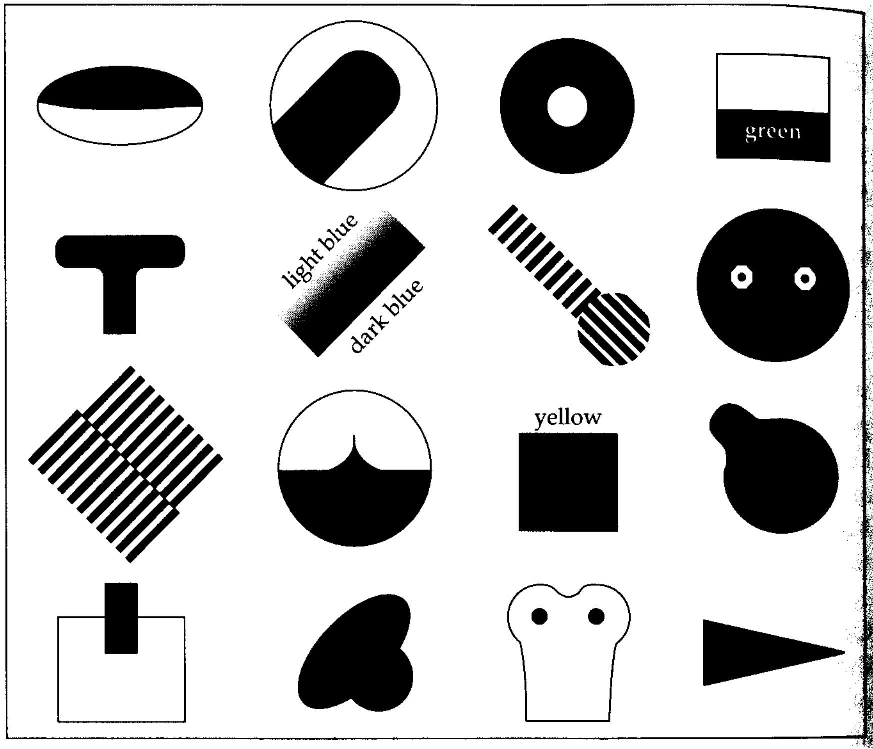


Figure 4.5 Examples of stimulus patterns for which cells in IT cortex show selectivity.

From K. Tanaka, "Inferotemporal cortex and object vision," Annual Review of Neuroscience, *19, copyright 1996 by Annual Reviews.*

Kawano, 1988). Further discussion of face cells will be deferred until the next chapter.

A fuller characterization of the information represented by cells in IT comes from experiments in which specific properties of a stimulus are varied while recording from a cell responsive to that stimulus (see Tanaka, 1996, for a comprehensive review). The results of these experiments are generally consonant with the conclusions of the lesion studies reviewed earlier, and with the general view that IT represents objects *per se* as opposed to incidental image features. For example, the position (e.g., Desimone, Albright, Gross, and Bruce, 1984), retinal image size (e.g., Sato, Kawamura, and Iwai, 1980), and picture plane orientation (e.g., Desimone *et al.*, 1984) have relatively small effects on cells' responses to an optimal shape, as illustrated by the data in figure 4.6.

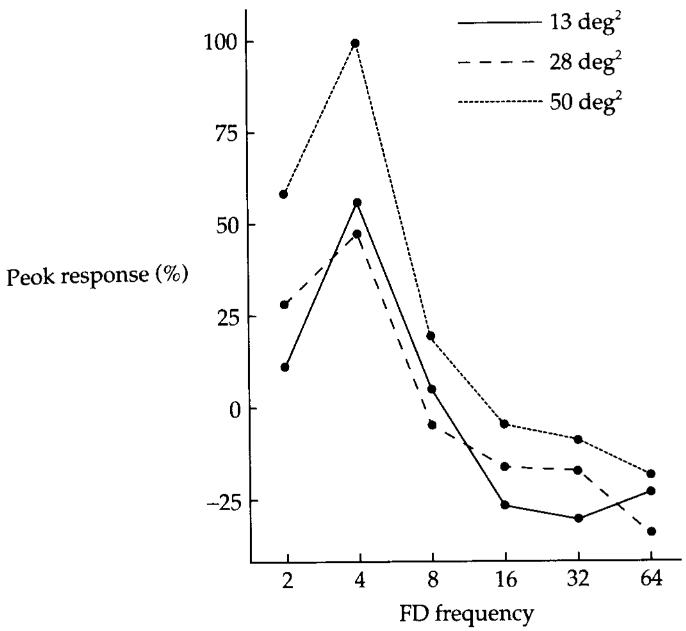


Figure 4.6 The response strength of a shape selective cell as a function of shape similarity (represented on the x-axis as Fourier Descriptor Frequency) and as a function of stimulus size (dotted, dashed, and solid lines). Note that there is shape selectivity, in that the functions are peaked, but the selectivity is not absolute; there is a generalization gradient to other similar shapes. Similarly, the selectivity shows size invariance, in that all functions are peaked for the same FD frequency, but the size invariance is not absolute; the cell responds more vigorously to one size than to the others.

From R. Desimone et al., *"Contour, color and shape analysis beyond the striate cortex,"* Vision Research, *25, 1985, with permission of Elsevier Science.*

Changes in depth orientation create more complex changes in the retinal image than changes in picture plane orientation, and the effect on IT cells, responses are less consistent. Perrett, Smith, Potter, Mistlin, Head, Milner, and Jeeves (1985) report face cells that respond preferentially to profile or frontal views of faces, as well as cells that generalize to some degree over depth rotations. Hasselmo, Rolls, Baylis and Nalwa (1989) report similar findings, and note that some orientation-independent cells maintain a preference for one face over another across rotations in depth. The effects of picture plane and depth rotations on cells' responses to nonface objects have been systematically investigated by Logothetis, Pauls and Poggio (1995) using complex wire frame and amoeba stimuli. They report some generalization, better for picture plane than depth rotations, but in no case was orientation-invariance complete.

4.4 Disorders of shape recognition in humans

The earliest clues about the neural bases of object recognition came not from the laboratory but from the neurology clinic, specifically from study of patients with visual agnosia. Visual agnosia is a blanket term for a wide array of visual disorders affecting object recognition, in which elementary visual functions such as acuity and visual fields are grossly intact, or at least adequate to allow for recognition (see my 1990 book on agnosia for a taxonomy and detailed review). Agnosias are commonly divided into the "apperceptive" and "associative" varieties, a distinction introduced by Lissauer (1890). According to Lissauer, object recognition could be impaired because the object is not adequately perceived, or because the percept fails to be associated with relevant knowledge in memory. Agnosic patients whose perception is obviously impaired, despite intact or at least adequate visual sensory function, were classified as apperceptive agnosics on the assumption that their impairment lay in the stage of "apperception." Agnosic patients whose perception seemed grossly intact were classified as associative agnosics on the assumption that their impairment lay in the stage of "association" of percept and memory. In the words of Teuber (1968), these patients experience "a normal percept, stripped of its meaning."

The apperceptive/associative distinction is valid in the sense that there are agnosic patients with and without blatant perceptual impairments, and their underlying problems do appear to be different. In other words, there is reason to draw a line between two general types of patients, on purely empirical grounds. However, the interpretation suggested by Lissauer's terms "apperceptive" and "associative" is probably wrong. The underlying problem in associative agnosia is likely to be perceptual too, and not one of "association." In fact, of the two types of visual agnosia most relevant to the issue of shape perception, one of them is associative visual agnosia; the other is a disorder that is usually grouped with the apperceptive agnosias, termed "perceptual categorization deficit."

4.5 Associative visual agnosia

Although the term "associative visual agnosia" has itself been used to cover a range of disorders (see Farah, 1990), in its narrow sense it refers to an impairment in visual object recognition that is not attributable to defective semantic knowledge of the objects nor to clinically apparent perceptual difficulty. To be considered an associative agnosic, a patient must demonstrate the following features: First, he or she must have difficulty recognizing visually presented objects. This must be evident in ways other than just naming, such as sorting objects by category (e.g., putting kitchen utensils together, separate from sports equipment) or pantomiming the objects' functions. If the trouble is confined to naming objects, and is not manifest in nonverbal tests of recognition, then the problem is either anomia or, if confined to the naming of visual stimuli, optic aphasia (see chapter 9). Second, the patient must demonstrate that knowledge of the objects is available through modalities other than vision, for example by tactile or auditory recognition, or by verbal questioning (e.g., what is an egg beater?). Some dementias may result in a loss of knowledge about objects regardless of the modality of access, and this is distinct from a visual agnosia (e.g., Hodges, Patterson, Oxbury, and Funnell, 1992; Martin and Fedio, 1983; Warrington, 1995). Third, the patient must be able to see the object clearly enough to describe its appearance, draw it, or answer whether it is the same or different in appearance compared with a second stimulus.

An interesting illustrative case of associative visual agnosia was reported by Rubens and Benson (1971). Their subject was a middle-aged physician who became agnosic following an acute hypotensive episode. His mental status and language abilities were normal, his visual acuity was 20/30, and although he had a right homonymous hemianopia (blindness in his right visual field) his visual fields were certainly larger than those of many non-agnosic individuals. They report that:

> For the first three weeks in the hospital, the patient could not identify common objects presented visually, and did not know what was on his plate until he tasted it. He identified objects immediately on touching them. When shown a stethoscope, he described it as "a long cord with a round thing at the end," and asked if it could be a watch . . . He was never able to describe or demonstrate the use of an object if he could not name it. . . . He could match identical objects, but not group objects by category (clothing, food) . . . He was unable to recognize members of his family, the hospital staff, or even his own face in the mirror . . . Remarkably, he could make excellent copies of line drawings and still fail to name the subject . . . He easily matched drawings of objects he could not identify, and had no difficulty discriminating between complex nonrepresentational patterns differing from each other only subtly. He occasionally failed because he included imperfections in the paper or printer's ink." (pp. 308–9)

In this classic case we see all the elements of associative agnosia: Impaired recognition of visually presented objects, demonstrated verbally and nonverbally, in a patient with normal intellect and apparently adequate visual perception. Recognition of objects through other modalities is intact, and copying and matching ability appear remarkably preserved. Figure 4.7 shows four drawings that this patient was unable to recognize, along with his excellent copies. Figure 4.8 shows the copies of four other agnosic patients, demonstrating the generality of the striking dissociation between perception (as measured by copying ability, at least) and recognition.

Evidence for a shape perception impairment

How can someone be of sound mind, see pictures clearly enough to produce the copies shown in figures 4.7 and 4.8, and yet not recognize the pictures? This constellation of abilities and impairments seems almost paradoxical, and perhaps for this reason the very existence of visual agnosia has been doubted (e.g., Bay, 1953; Bender and Feldman, 1972). The good drawings and preserved matching ability of such patients also invites the conclusion that perception is intact, and that the fault lies in the process of associating a normal percept with memory knowledge.

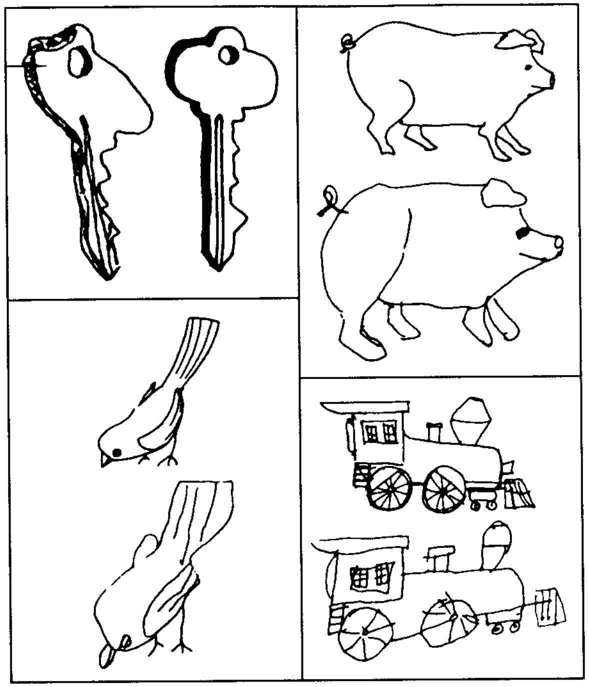


Figure 4.7 Copies of pictures made by an associative visual agnosic who could not recognize the pictures, either before or after copying them.
From A. B. Rubens and D. F. Benson, "Associative visual agnosia," Archives of Neurology, *24, 1971, with permission of the American Medical Association.*

Although a failure of association is one possible explanation, it is also possible that perception itself is at fault despite appearances to the contrary. Several considerations suggest that a perceptual impairment underlies associative agnosia. First, although the final products of these patients' copying efforts are often normal, the process by which they produce the copies is generally reported to be abnormal. The words "slavish" and "line-by-line" are often used in describing the manner of copying in these cases (e.g., Ratcliff and Newcombe, 1982; Wapner, Judd, and Gardner, 1978), including the patient of

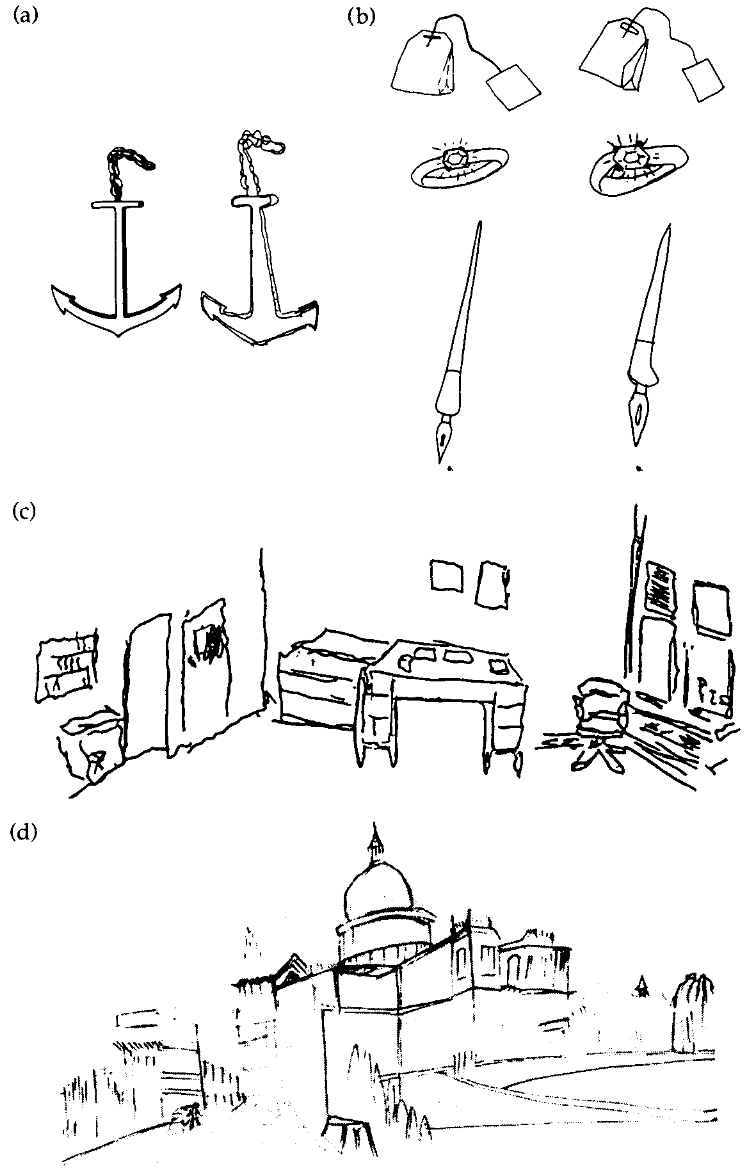


Figure 4.8 More examples of the good-quality copies made by associative visual agnosics who do not recognize their subject matter. (a) an anchor; (b) a teabag, ring, and pen; (c) the office in which the patient was sitting; (d) St. Paul's Cathedral.

From G. W. Humphreys and M. J. Riddoch, To See But Not to See: A Case Study of Visual Agnosia, *Hillsdale, NJ, Lawrence Erlbaum Associates, 1987, reprinted by permission of Psychology Press Limited; M. J. Farah,* Visual Agnosia: Disorders of Object Recognition and What They Tell Us About Normal Vision, *Cambridge, MA, MIT Press, 1990; W. Wapner* et al., *"Visual agnosia in an artist,"* Cortex, *14, 1978.*

Rubens and Benson, who was observed copying by Brown (1972). My own observations of L.H., an agnosic to be described in more detail in the following chapter, is that his drawings are executed abnormally slowly, with many pauses to check the correspondence of each line of the copy and the original. The impressive rendition of St. Paul's Cathedral by Humphreys and Riddoch's (1987) case H.J.A. impresses us in a different way when we learn that he spent 6 hours on it!

In evaluating the copying techniques of associative visual agnosics as evidence for a visual perceptual impairment, we should consider the alternative possibility that decreased availability of semantic knowledge might interfere with copying. A normal person's semantic grasp of what an object is might be expected to help a person keep the object's elements in working memory while it is being copied. However, it does seem unlikely that an absence of top-down semantic support for perception or perceptual working memory would be responsible for a 6-hour copying session! Nor does it seem able to explain the slavish line-by-line approach reported in so many cases, as normal subjects do not copy meaningless patterns in this way.

Several other observations are consistent with an impairment in visual perception, although these vary in their decisiveness. Associative visual agnosic patients are also abnormally sensitive to the visual quality of stimuli, performing best with real objects, next best with photographs, and worst with line drawings, an ordering reflecting increasing impoverishment of the stimulus (e.g., Levine and Calvanio, 1989; Ratcliff and Newcombe, 1982; Riddoch and Humphreys, 1987; Rubens and Benson, 1971). Tachistoscopic presentation, which also reduces visual stimulus quality, also impairs associative agnosic performance dramatically. Although this would seem to be *prima facie* evidence for a visual impairment, an absence of top-down semantic support can also account for an increase in sensitivity to visual factors (Tippett and Farah, 1994).

Potentially more decisive evidence comes from the nature of the recognition errors made by associative agnosics. The vast majority of errors are visual in nature, that is, they correspond to an object of similar shape (e.g., Levine, 1978; Ratcliff and Newcombe, 1982). For example, on four different occasions when I asked case L. H. to name a picture of a baseball bat, he made four different errors, all reflecting shape similarity: paddle, knife, baster, thermometer. The subject of

Davidoff and Wilson (1985) made some semantic as well as visual errors, but she was able to correct her semantic errors later when offered a forced choice between her initial answer and the correct one, whereas her visual errors were less tractable. Although visual errors can be accounted for by impaired access to semantic knowledge (Hinton and Shallice, 1991), such accounts predict accompanying semantic errors. Therefore, for those cases in which visual shape errors are found in the absence of semantic errors, it is likely that visual shape perception is at fault.

The matching of unfamiliar faces and complex meaningless designs, in which semantics would not play a role, also provides decisive evidence for a visual perceptual impairment. Changing the angle or lighting in the photograph of a face impairs agnosics' ability to match unfamiliar faces (Shuttleworth, Syring, and Allen, 1982). The matching of abstract geometric forms is even less likely to depend on semantic knowledge than the matching of unfamiliar faces. Recall that Rubens and Benson's patient occasionally mistook flaws in the paper or printer's ink for a part of the design, reminiscent of IT-lesioned monkey's use of local, idiosyncratic features in visual discrimination learning. Levine (1978) administered a visual discrimination learning task to an associative agnosic, and found her unable to learn a subtle discrimination between two patterns after 30 trials.

In sum, associative visual agnosics appear to be the human analog of the IT-lesioned monkeys described earlier. A variety of evidence suggests that they fail to recognize objects because they fail to represent their shape in a normal way. The extremely slow and slavish copying technique, the sometimes isolated occurrence of visual shape errors, and abnormalities in performance at matching abstract designs, all point fairly directly to a shape perception impairment. The analogy holds anatomically as well. Although the human lesions tend to be somewhat more posterior than in the monkey brain, they are inferior and generally include temporal as well as occipital cortex.

4.6 Perceptual categorization deficit

Warrington and her colleagues have described another type of visual recognition impairment, which they term "apperceptive agnosia,"and which they characterize as an impairment of perceptual categorization. Because the term "apperceptive agnosia" has been used in a variety of different ways by different authors, and because it has been used most consistently to label the disorder of grouping discussed in chapter 3, I have referred to the present disorder as "perceptual categorization deficit" (see Farah, 1990, for a detailed review of the literature on this form of agnosia). The cardinal feature of perceptual categorization deficit, first documented by Warrington and Taylor (1973), is an inability to recognize objects viewed from unusual perspectives, or to match pairs of objects depicted in one usual and one unusual perspective. Warrington (1985) has cited unpublished data showing that the same type of patient also has difficulty recognizing objects photographed under conditions of uneven or unusual illumination. Figure 4.9 shows examples of the kinds of stimuli used in this research.

On the face of things, perceptual categorization deficit appears to be the loss of just those "miraculous" representations discussed at the outset of this chapter. Indeed, Warrington's research on perceptual categorization deficit was the only neuropsychological evidence cited by David Marr in his landmark (1982) book on vision, and he presented it as bearing on the representations underlying object recognition. In this context, he interpreted the disorder as an inability to transform the image representation to an object-centered representation of shape, from which perspective and other aspects of the viewing conditions had been eliminated.

Although perceptual categorization deficit has attracted the attention of many leading researchers since Marr as a source of clues to the mechanisms of orientation invariance, there are reasons to doubt its direct relevance. First, these patients are not impaired in everyday life. Their deficit is manifest only on specially designed tests. This is in sharp contrast to associative visual agnosics just described, who are significantly handicapped by their visual disorder. Perhaps more to the point, it is also in contrast to the predicted effects of derailing vision at a retinotopic or image-based stage of representation.

A second and related point is that these patients have not been demonstrated to have an impairment in matching objects across different views. What, you say? Although readers may remember learning that perceptual categorization deficit involves a problem in matching different views of objects, all that has been demonstrated is

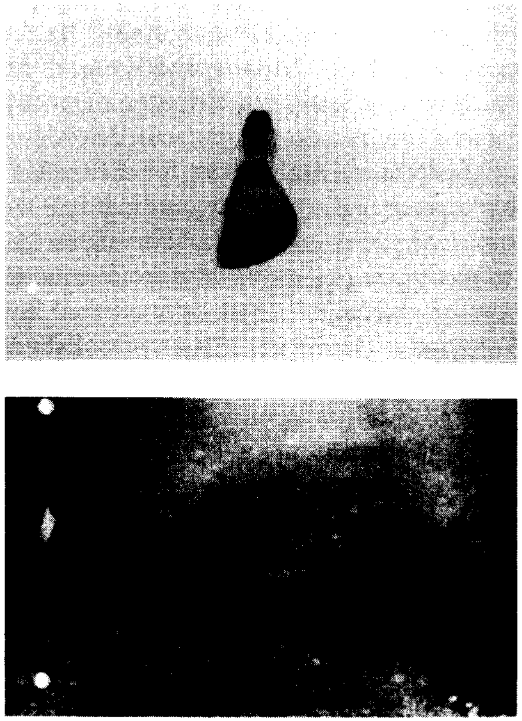


Figure 4.9 Examples of photographs used to test for a perceptual categorization deficit. (a) unusual view (b) unusual lighting.

From M. J. Farah, Visual Agnosia: Disorders of Object Recognition and What They Tell Us About Normal Vision, *Cambridge, MA, MIT Press, 1990.*

a problem matching a usual to an *unusual* view. Although one could construct a test in which different usual views of objects must be matched, the tests used so far have always included an unusual view. In my experience normal subjects often require a few seconds to identify these unusual views, and published data show that their performance is not without error (e.g., Warrington and Taylor, 1973). This raises the possibility that the recognition or matching of unusual views requires a kind of effortful processing above and beyond object perception proper. Such processing might more aptly be called visual problem solving than visual recognition.

A third reason for questioning whether perceptual categorization deficit results from a loss of the visual shape representations normally

used in object recognition comes from its associated neuropathology. Everything we know about the localization of visual shape representation in nonhuman primates implicates the ventral visual system bilaterally. Perceptual categorization deficit in humans generally follows unilateral right hemisphere lesions, and is particularly associated with parietal damage (Warrington and Taylor, 1973).

In sum, despite the initial impression that perceptual categorization deficit represents a selective impairment of viewpoint-invariant object recognition, a closer look at both behavior and anatomy casts doubt on this idea. Indeed, although Warrington (e.g., 1985) once viewed perceptual categorization as the first of two main stages of object recognition (the second being the access of semantic knowledge), in more recent writings she has stated that "we would now wish to argue that perceptual categorization systems may be an optional resource rather than an obligatory stage of visual analysis" (Warrington and James, 1988).

4.7 Neuroimaging studies of object recognition in humans

The recently developed techniques of PET and fMRI have the potential to localize object recognition processes in the human brain with greater precision than is possible with naturally occurring lesions. Functional neuroimaging can also be used to answer certain questions about functional characteristics of object recognition, through inferences based on localization information. Geoffrey Aguirre and I recently surveyed the neuroimaging literature on object recognition (Farah and Aguirre, 1999). From a large set of published studies that involved viewing or making judgments about visually presented stimuli, we found 17 whose design made it possible to at least roughly isolate visual recognition *per se*.

The 17 relevant studies are listed in table 4.1. Beyond a shared affinity for Snodgrass and Vanderwart pictures, they are a heterogeneous collection of designs. Some simply contrasted passive viewing of visual stimuli (e.g., line drawings of objects, printed words or pseudowords, photographs of faces) with passive viewing of control stimuli (e.g., fixation points, scrambled pictures, textures). Others

Table 4.1 Studies which roughly isolate visual recognition *per se*. (From M. J. Farah and G. K. Aguirre, 1999.)

Study	*Task*
Words	
Petersen *et al.*, 1988	*Passive* viewing of words vs. fixation
Petersen *et al.*, 1990	*Passive* viewing of words and pseudo-words vs. passive viewing of false fonts
Howard *et al.*, 1992	Read aloud visually presented words vs. view false fonts and say "crime"
Price *et al.*, 1994, exp. 1	Read aloud visually presented words vs. perform feature decision on false fonts
Price *et al.*, 1994, exp. 2	*Passive* viewing of words vs. passive viewing of false fonts
Menard *et al.*, 1996	*Passive* viewing of words vs. fixation
Puce *et al.*, 1996	*Passive* viewing of letter strings (nonwords) vs. passive viewing of textures
Polk *et al.*, 1998	*Passive* viewing of AltErNAtIng case words vs. passive viewing of consonant strings
Objects	
Sergent *et al.*, 1992a	Living/nonliving judgment regarding Snodgrass and Vanderwart (S&V) pictures vs. fixation
Sergent *et al.*, 1992b	Living/nonliving judgment regarding S&V pictures vs. judge gratings as vertical or horizontal
Kosslyn *et al.*, 1994	Matching S&V pictures with their names vs. viewing random patterns of lines
Kosslyn *et al.*, 1995	Picture verification performed upon S&V-style line drawings of objects and auditorily presented "entry level" words vs. scrambled lines and words
Malach *et al.*, 1995	*Passive* viewing of objects vs. passive viewing of phase randomized pictures
Menard *et al.*, 1996	*Passive* viewing of S&V pictures vs. fixation
Kanwisher *et al.*, 1997	*Passive* viewing of S&V pictures (and novel S&V-style objects) vs. passive viewing of scrambled lines
Faces	
Sergent *et al.*, 1992b	Gender categorization of faces vs. judge gratings as vertical or horizontal
Haxby *et al.*, 1994	Matching faces across shifts of gaze vs. alternating button presses to scrambled faces
Haxby *et al.*, 1996	Encoding (viewing) faces vs. alternating button presses to scrambled faces
Puce *et al.*, 1996	*Passive* viewing of faces vs. passive viewing of textures
McCarthy *et al.*, 1997	*Passive* viewing of faces amongst phase randomized objects vs. viewing of phase randomized objects

Table 4.1 (cont'd)

contrasted active experimental tasks with control tasks intended to match at least some of the processing demands of the experimental task other than the need for object recognition. The experimental tasks included judgments such as living versus nonliving, name verification (e.g., is this a *tree*?), and for faces, verification of male versus female. The control tasks in these studies used stimuli such as scrambled pictures or gratings that were either passively viewed or the object of different judgments, such as horizontal versus vertical.

An optimist might view the heterogeneity in the designs of these studies as an opportunity to identify the cortical areas that participate in visual recognition independent of the particulars of task and stimulus. A pessimist might expect the variability in designs, especially the imperfect ways in which control tasks are matched to experimental tasks, to obscure the true neural locus of visual recognition. Figure 4.10, showing the 84 activation maxima from the 17 studies, suggests that the pessimist's prediction may be closer to the truth. The only generalization that one can make, on the basis of these data, is that visual recognition is a function of the posterior half of the brain!

Before giving the pessimist the last word, let us explore this data set a bit further to see if there are clusters of maxima, within the overall scatter, corresponding to particular aspects of task design or

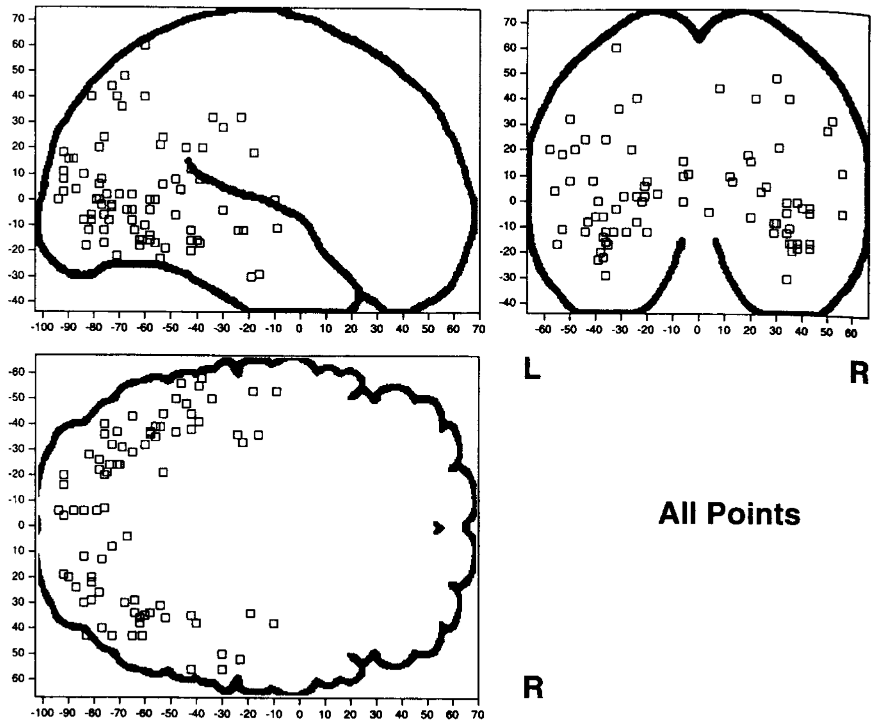


Figure 4.10 Activation maxima from 17 neuroimaging studies of visual recognition.

From M. J. Farah and G. K. Aguirre, "Imaging visual recognition: PET and fMRI studies of the functional anatomy of human visual recognition," Trends in Cognitive Sciences, *3, 1999.*

stimuli. The first distinction to look at, if task variability and imperfect control conditions are a concern, is the active versus passive nature of the experimental task. Active tasks, because they involve more processing beyond simply seeing and recognizing the stimulus, are prone to spurious maxima if the control condition fails to match perfectly the nonrecognition processing. Figure 4.11 shows the maxima associated with the contrasts between experimental and control conditions for active and passive tasks separately. The active tasks cover a slightly broader range of brain than the passive, but the difference hardly accounts for the overall scatter. Both active and passive tasks produce widely distributed maxima.

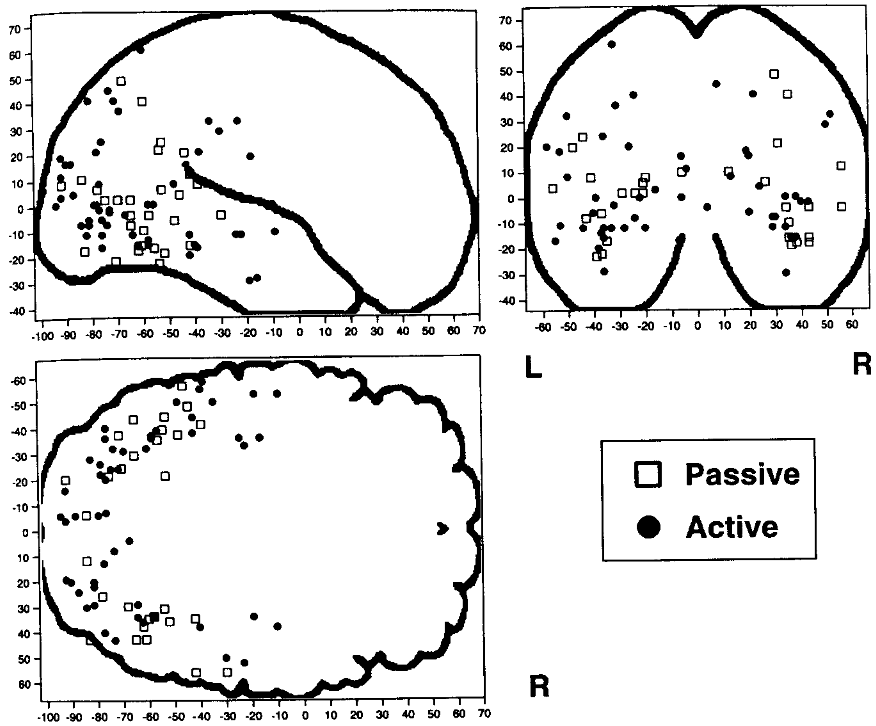


Figure 4.11 Maxima subdivided into those derived from subtractions between passive object viewing and passive baseline tasks, and those derived from subtractions between active object recognition tasks (e.g., living/nonliving classification) and corresponding active baseline tasks.

From M. J. Farah and G. K. Aguirre, "Imaging visual recognition: PET and fMRI studies of the functional anatomy of human visual recognition," Trends in Cognitive Sciences, *3, 1999.*

The possibility that different categories of stimuli may be recognized using different neural systems is a question that will be taken up in more detail in the following two chapters. It is an example of an issue concerning the functional organization of visual recognition, rather than its anatomical localization *per se*, that can be addressed using neuroimaging data. If the regions activated by object, face, and word recognition are segregated into different parts of visual cortex, this would support a category-specific organization. For present purposes, the possibility of category-specific recognition systems is of

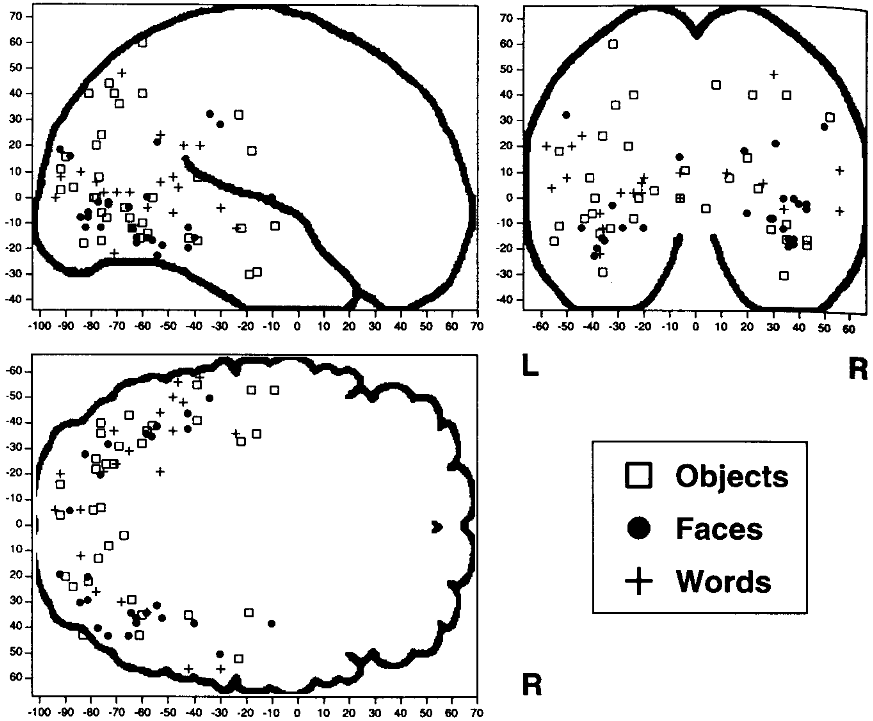


Figure 4.12 Maxima subdivided into those derived from visual recognition of objects, faces, and printed words.

From M. J. Farah and G. K. Aguirre, "Imaging visual recognition: PET and fMRI studies of the functional anatomy of human visual recognition," Trends in Cognitive Sciences, *3, 1999.*

interest as a way of explaining the seemingly nonfocal nature of activation maxima associated with visual recognition. Perhaps the scatter apparent in figures 4.10 and 4.11 can actually be subdivided into some number of more compact non-overlapping clusters. Figure 4.12 shows that subdividing the studies by category of stimulus does not greatly reduce the scatter.

On the basis of the findings summarized so far, it would be fair to say that functional neuroimaging has not taught us much regarding the neural bases of object recognition in humans. Different studies produce different results, and the source of the variability is unclear. It does not appear to result from the different categories of stimuli used in these studies, nor does it appear to result from the variability in the tasks used to study recognition. What could be wrong?

What neuroimaging studies localize is the psychological process, or processes, by which an experimental task and a control task differ. With this in mind, look again at the designs summarized in table 4.1. The experimental tasks generally do require more visual recognition than the control tasks. The problem is that this is not the only difference between the experimental and control tasks. In the passive viewing tasks, the experimental and control stimuli often differ dramatically, by such gross measures as luminance flux, size, and complexity. In the active tasks, both stimuli and task instructions differ dramatically. There is little wonder that the results of these studies do not superimpose.

There is no reason why functional neuroimaging studies cannot be designed to better isolate the processes of interest, and indeed a few good examples already exist. Most of these specifically address the issue of specialization for different categories of stimuli, and were designed using experimental and control tasks that differ minimally. Because they do not isolate object recognition *per se*, but instead isolate specific subtypes of object recognition relative to one another, they will be described in the two chapters that follow, on the question of specialized recognition systems for faces and printed words.

4.8 Neural representations underlying object recognition: a computational interpretation

Having reviewed a broad array of empirical findings on visual object processing in the brains of humans and their primate cousins, we are now in a position to try to deduce some constraints on the nature of the underlying representations.

Coordinate system: empirical evidence

The evidence from single unit recordings and lesions in monkeys rules out the simplest versions of a viewer-centered or an environment-centered coordinate system. The relatively invariant responses of at least some IT cells to a given shape over changes in position, size, and picture plane orientation relative to viewer and environment are not consistent with a coordinate system anchored to either. The impairment of IT-lesioned monkeys in generalizing learned visual object discriminations to new views of the objects, and their normality at learning to discriminate different views of a single object, also suggest that IT neurons possess some degree of viewpoint-invariance. Finally, the ability of IT-lesioned monkeys to generalize a learned discrimination to new patterns when some of the features of the earlier patterns stay in the same position relative to the monkey and/or the environment, but not when the same features are shifted to a new position (see figure 4.3), is further evidence for an abnormal reliance on viewer-centered or environment-centered representation and hence a loss of more abstract representations of shape.

Although these data clearly rule out the use of a plain viewer-centered or environment-centered coordinate system, they do not definitively implicate an object-centered coordinate system in IT. Recall that when a viewer-centered system is augmented with associative learning and normalization processes, it too will enable viewpoint-invariant object recognition. In the terms used to describe the problem at the outset of this chapter, there are two ways to deal with the bundling together of shape and perspective in retinotopic representations. One is to undo the bundle, and this is equivalent to computing an object-centered representation. The other, less aesthetic but easier to accomplish, is simply to sort the bundles according to the objects that gave rise to them. They can be sorted according to their intertransformability (e.g., this viewpoint-dependent representation can be rotated and enlarged to match that one) or through explicit learning (e.g., this viewpoint-dependent representation is my grandmother and so is that one) or a combination (as proposed by Tarr and Pinker, 1989).

In short, it is possible that IT does not house object-centered representations *per se*, but rather the ability to associate multiple viewer-centered representations and/or transform one viewer-centered representation to another. Two empirical observations lend some degree of support to the latter alternative, although the issue is far from resolved. First, the invariances for position, size, and orientation that we see in the responses of IT neurons are always imperfect (see figure 4.6). Indeed some studies, with wire and amoeba-like stimuli, find rather limited orientation invariance (Logothetis, Pauls, and Poggio, 1995). This is not what would be expected if objects' shapes were being represented in an object-centered coordinate system, which does not contain perspective information. In contrast, it is easier to see how perspective could have residual effects on the processing of a system that never eliminated perspective information in the first place. Unusual views might be less well-learned or require additional normalization with consequent additional likelihood of error. A second observation that lends credence to the viewer-centered alternative is the demonstrated ability of IT neurons to learn associations between patterns (Miyashita, Date, and Okuno, 1993). These cells have been shown to acquire selectivity for arbitrary pairs of stimuli that have been repeatedly associated, a necessary ability for deriving invariances from viewer-centered representations through learning.

Foldiak (1991) has proposed a simple computational mechanism by which viewpoint-independent representations could emerge from seeing a given object from different perspectives. He combined the idea that different views of an object are often clustered in time, with the idea that cells' activity takes some time to decay. The consequence of these two ideas is the following: An active cell in a higher visual area such as IT might remain active throughout the time that a moving object activates first one set of cells then another in earlier retinotopic areas, and by correlation-driven learning this will associate both of the retinotopic representations with the same higher-level representation. Wallis and Rolls (1997) have developed similar ideas in the context of the physiology of the different visual areas, going from V1 to IT.

Primitives: empirical evidence

Surprisingly, no research has directly addressed the nature of the geometric primitives used in primate, including human, object recognition. Nevertheless, there are clues available from a number of sources that show a reassuringly high degree of agreement in pointing to either surface-based or volumetric primitives for the shape representations underlying object recognition in IT. Discriminating between surface-based and volumetric primitives is not possible at present, but at least contour-based primitives can be tentatively ruled out.

Two indications of noncontour-based representation are available in the literature on IT lesions in monkeys. First, these animals are impaired at perceiving shape equivalence over changes in the pattern of shadow and light falling on the object. Such changes do not affect the depth information needed to derive surface and volumetric representations, but they do affect the pattern of spurious contours. This suggests that the object perception of IT-lesioned monkeys is abnormally reliant on contour information, and hence that they differ from normal monkeys by an inability to derive noncontour-based representations. The finding that IT lesions also impair monkeys' ability to perceive shapes in random dot stereograms, which have no contours, provides additional evidence that the function of IT includes noncontour-based representation. Recordings from IT neurons are also consistent with this interpretation. The preference of these neurons for three-dimensional objects, or models of objects, over flat outline shapes suggests the importance of surface properties such as texture, shadow, and disparity, which provide cues to the surface or volumetric shape. Perhaps most compelling is the finding that IT neurons respond selectively to shape whether defined by luminosity differences, which form the basis for static contour, or by texture or motion differences, which do not give rise to contours in the sense of elongated zones of transition from light to dark. Research on the face perception of agnosic patients also suggests that they may be more dependent on contours than a normal human, in that they have difficulty seeing the equivalence of faces photographed from the same angle but with a different play of light and shadow. Their heightened sensitivity to the differences between drawings, photographs, and real objects may also reflect an impaired ability to extract or infer surface and/or volume information.

The evidence that IT represents shape in terms of surface-based or volumetric primitives contrasts with at least one common interpretation of the response properties of cells in earlier occipital areas, reviewed in chapter 1, according to which they represent edges and contours. Thus, one way in which the representation of the stimulus appears to be transformed in going from early occipital to inferotemporal representations is that the building blocks of shape representation go from contours to some higher-order geometric primitive, either surfaces or volumes.

Organization: empirical evidence

Studies of object vision in monkeys have relatively little to tell us concerning the degree and type of organization imposed on object shape by the primate visual system. The one source of direct evidence is the finding that face cells show greatly diminished responses to scrambled faces. If face parts were explicitly represented as units of shape in their own right in a hierarchy of shape representation, then the representation of the scrambled face would still be partially equivalent to the representation of the intact face at the part level of the hierarchy. The lack of response to scrambled faces suggests that face cells do not embody a hierarchically organized representation of shape. However, as will be argued in the next chapter, this particular aspect of face cell function may well be unrepresentative of the cells involved in representing nonface objects.

Turning to the human evidence, the behavior of some agnosic patients seems very relevant to the issue of hierarchical shape representation. When shown an object or picture that they cannot recognize, agnosics frequently guess its identity on the basis of its local parts or features. For example, an animal with a long tapered tail might engender "rat" or "mouse" as a guess. A baby carriage whose wheels have metal spokes might be called a "bicycle." This behavior invites interpretation in terms of a hierarchical system of shape representation, whose lower level part representations are relatively intact but whose higher level integration of the parts is damaged or unavailable. Riddoch and Humphreys (1987) have explicitly suggested that such an impairment in the integration of local parts into higher and more global levels of a shape hierarchy may underlie certain cases of agnosia. They introduced the term "integrative agnosia" for such cases.

In addition to the use of local parts for guessing the identity of objects, they point to several other aspects of agnosic performance that seem consistent with this interpretation, specifically: Impaired recognition of briefly presented stimuli (because, they argue, if parts are serially encoded more time will be required), impaired recognition of overlapping drawings (because impaired part integration will be further taxed by the possibility of misconjoining the parts of different objects), impaired discrimination of real objects from pseudo-objects composed of mismatched parts of real objects, and greater impairment relative to normal subjects at recognizing more complex depictions (because these contain more parts).

Because the idea of integrative agnosia has important implications for the issue of the organization of visual object representations, let us scrutinize it further. Although there is no doubt that an impairment in integrating shape parts into global wholes is consistent with the findings just listed, such an impairment is not the only way to account for these findings.

First, consider the basic finding that agnosics may guess the identity of objects based on a single correctly perceived part. While consistent with an impairment in integration of parts, it is also consistent with almost any type of impairment in shape processing capacity, as the shape of a part will always be simpler than the shape of a whole object. Above and beyond this, in any system for which there is a fixed probability of recognizing a given shape (part or whole), there will be more successes with just parts than with just wholes, simply because parts are more numerous.

The other features of integrative agnosia are similarly ambiguous with respect to the underlying impairment in shape representation. The slower speed of agnosic object recognition is hardly a unique prediction of impaired part integration, nor is the detrimental effect of overlapping pictures, as almost any impairment of shape perception one can think of would be expected to slow the process and make it less robust to interfering contours. Similarly, object decision would be expected to be impaired whenever shape perception is defective in any way. The difference in performance between silhouettes and detailed drawings after unspecified perceptual impairment could take the form of better performance the more information is available (hence drawings better than silhouettes) or better performance the simpler that shape to be perceived (hence silhouttes better than drawings), but certainly the latter prediction is not unique to a specific impairment of part integration.

In sum, we know little at this point about the organization of object shape representations. There is no evidence from monkeys or humans that specifically implicates a hierarchical organization for the object representations of IT.

Implementation: empirical evidence

With respect to the type of search process that underlies visual object recognition, the question can be posed thus: Are there two tokens of a high-level object representation, one derived from the stimulus and one waiting in memory against which the stimulus representation is matched? Or does the stimulus get encoded and recoded, starting in early visual areas in which the representation is determined largely by the innate structure of the visual system, and ending with still just one token representation in higher-level areas, in which the representation is determined by a structure that results from learning? In the former case, one can point to distinct perceptual and mnemonic representations of the object, within high-level visual areas. In the latter case, there is no distinction between perception and memory; if one's memory is changed or disrupted, so is one's high-level perception. High-level visual representations are perceptual, in the sense that they are derived from stimulus input, and they are mnemonic in the sense that the pattern of weights responsible for their derivation is determined by experience (in contrast to the smaller role of experience in setting the weights at earlier stages of visual processing).

If object search is implemented in the first way, in common with search in symbol-manipulating computers, then it should in principle be possible to destroy the memory representation but retain the high-level perceptual representation of the object. If object search is implemented in the second way, in common with neural network computation, then impaired performance on tests of object recognition (memory) will always be accompanied by impaired performance on tests of object perception. Although no direct tests of this prediction have been made in either the monkey or the human literature, the apparent universality of impaired object perception in associative agnosia, discussed earlier, is more consistent with a neural network implementation of search.

The degree of distributedness of object representation has been addressed most directly in the single unit recording literature. The striking specificity of IT neurons for particular shapes, even for one face over another, might seem to suggest the kind of one stimulus–one neuron system of representation that is equivalent to a localist implementation. However, even these highly selective neurons show some degree of generalization, responding in varying degrees to different faces. Thus, for a given object or face, a number of neurons will be active to varying degrees, equivalent to a distributed representation (Young and Yamane, 1992).

The spatial scale of functional neuroimaging, and the necessity of combining data from multiple trials, makes comparable evidence impossible to obtain from humans. However, one of the neuroimaging studies cited earlier is nevertheless relevant to the issue of distributed representation. Kanwisher, Woods, Iacoboni, and Mazziotta (1997) compared patterns of brain activity while their subjects viewed line drawings of real objects, line drawings of made-up objects, and scrambled line drawings that had no three-dimensional interpretation as an object. As expected, they found inferotemporal activation associated with viewing the objects, relative to the scrambled displays. They also found equivalent activation associated with viewing the made-up objects. This is consistent with a distributed system of representation, in which a made-up object can be represented by a novel ensemble of the same parts used to represent familiar objects.

The graded way in which object recognition breaks down after IT lesions is also indicative of a system of distributed representation. IT-lesioned monkeys and human agnosics do not lose the ability to recognize arbitrary subsets of all objects, such as tall things with corners. Agnosias may be more or less severe, consistent with more or less of a distributed representation having been damaged, but by and large they affect all objects equally. There are two well-established exceptions to this generalization, to which we now turn. Both face recognition and printed word recognition may make use of cortical representations that are to some extent segregated from each other and from object representation.

chapter five

Face Recognition

5.1 Are faces "special"?

Everything that was said in chapter 3 about the problem of object recognition would seem to apply equally well to the problem of face recognition. Aside from finding certain exemplars of this category particularly endearing, it is hard to see the difference between faces and other objects. As illustrated in figure 5.1, faces present us with highly variable images depending upon the angle from which we view them and the positions of their moveable parts. Whether the content of the image is a common object or a face, our visual system must create a representation that is invariant over at least a range of such viewing conditions, yet discriminates among exemplars.

This very reasonable sounding argument for common mechanisms underlying face and object recognition is contradicted by an array of empirical findings in developmental psychology, psychophysics, and neuropsychology. The neuropsychological evidence, from brain-damaged humans and from neuroimaging studies of normal humans, is arguably the strongest evidence and will be the focus of this chapter. Just two examples of evidence from outside of neuropsychology will be described here.

Developmental psychologists have shown that we come into the world predisposed to treat faces differently from other objects. For example, human infants only 30 minutes of age will track a moving