

VISUAL OBJECT RECOGNITION

Nikos K. Logothetis and David L. Sheinberg

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

KEY WORDS: object representation, perceptual categorization, monkey, electrophysiology, inferotemporal cortex

ABSTRACT

Visual object recognition is of fundamental importance to most animals. The diversity of tasks that any biological recognition system must solve suggests that object recognition is not a single, general purpose process. In this review, we consider evidence from the fields of psychology, neuropsychology, and neurophysiology, all of which supports the idea that there are multiple systems for recognition. Data from normal adults, infants, animals, and brain-damaged patients reveal a major distinction between the classification of objects at a basic category level and the identification of individual objects from a homogeneous object class. An additional distinction between object representations used for visual perception and those used for visually guided movements provides further support for a multiplicity of visual recognition systems. Recent evidence from psychophysical and neurophysiological studies indicates that one system may represent objects by combinations of multiple views, or aspects, and another may represent objects by structural primitives and their spatial interrelationships.

INTRODUCTION

An essential behavior of animals is the visual recognition of objects that are important for their survival. Human activity, for instance, relies heavily on the classification or identification of a large variety of visual objects. We rapidly and effortlessly recognize these objects even when they are encountered in unusual orientations, under different illumination conditions, or partially occluded by other objects in a visually complicated environment.

How is this performance accomplished by the brain? What kind of information does the visual system derive from the retinal image to construct

descriptions of sets of object features that capture the invariant properties of objects? How are such descriptions stored, and how are they activated by the viewed object? Are object representations general, or are they specific to an action or to a cognitive process, such as learning, planning, or reasoning?

These questions have historically been addressed by scientists in a variety of disciplines, including cognitive psychology (Pinker 1985, Biederman 1987, Banks & Krajecek 1991), neurobiology (Gross 1973, Gross et al 1993, Miyashita 1993, Rolls 1994), neuropsychology (Humphreys & Riddoch 1987a, 1987b; Damasio et al 1990; Farah 1990; Grüsser & Landis 1991), and computation and engineering (Marr 1982, Ullman 1989, Koenderink 1990, Aloimonos 1993). In this chapter we review selected work from each of the aforementioned fields that, in combination, shed increasing light on the internal workings of this system.

Our aim is to provide evidence that a multipurpose general recognition system does not actually exist. Instead, in the process of biological recognition, multiple representations of an object are formed, each specific to the transformations required by either perception or action. The reviewed literature suggests that the recognition of prototypical members of an object category, the encoding of dynamic and plastic transformations of objects or object parts, the identification of individual members of a homogeneous object class, and the planning of movements habitually made when interacting with familiar objects rely on different representations that are formed in different neural sites or by different interconnectivity patterns.

We start with an overview of the basic capacities and limitations of the primate recognition system. After a brief description of some general principles of object categorization, we discuss the performance of human and nonhuman primates in different recognition tasks and relate this performance to relevant theoretical models. We then survey a number of human neuropsychological and animal lesion studies showing that damage in different regions of the brain often results in a selective disruption of different recognition processes. In the final section, we discuss findings from psychophysical and electrophysiological experiments in the monkey that examine the role of single neurons of cortical areas thought to be essential in the formation of object representations.

CATEGORIZATION

The world has an infinite number of stimuli that can be discriminated from one another, to an arbitrary degree of detail. Which discriminations are essential for a given recognition system, and what is the basis for organizing information into equivalence classes? In this section, we examine how humans classify and recognize objects, and then we provide evidence that the same principles most likely underlie categorization performed by other biological

recognition systems. Specifically, we show that the generalizations about the world that allow us to categorize objects are not the product of the development of language, but are instead of perceptual origin. Perceptual categorizations, in turn, reflect the redundant, correlational structure of the environment and occur most often at the level at which individual members of categories are most similar to each other and maximally different from members of other categories.

Object Classes and Taxonomies

In human cultures, object categories are usually designated by words that capture the common functional properties of the category's members. Brown (1958) considered the question of why everyday "things" (e.g. pineapples and dimes) are referred to by the same name by most members of a society. He concluded that "[t]he most common name for each of these categorizes them as they need to be categorized for the community's nonlinguistic purposes. The most common name is at the level of usual utility" (Brown 1958, p. 16). The idea that categories may actually reflect more than just linguistic constructs was examined systematically by Rosch and her colleagues (1976a). They showed that human conceptual categories have a perceptual basis and are determined by the high correlational structure of the real world, in which certain combinations of attributes are more probable than others. For instance, attributes such as "feathers" and "wings" co-occur often, while combinations such as "feathers" and "wheels" generally do not. Bundles of such co-occurring attributes form the basis of a natural classification for objects.

Rosch et al (1976a) argued that the world contains "intrinsically separate things," and that there exists a taxonomy for objects within which categories are related to each other by class inclusion. Such categories form natural groupings of stimuli with different perceivable characteristics, or "cues." Cues with high frequency within a given category and low frequency in all other categories are valid category predictors. For example, the cue "long neck" has extremely high validity for the category giraffe because it reliably predicts the presence of a giraffe. The cue "hoofed foot," on the other hand, has low validity for the same category because all ungulate mammals, in addition to giraffes, have hooves. The notion of cue validity extends to categories and is conceived of as the sum of the cue validities of each of the category's features (Reed 1972, Rosch et al 1976a).

General categories, such as mammals, are highly inclusive but have low cue validity, since few perceivable characteristics are shared among their members. Categories such as doberman, on the other hand, are very specific but also have low cue validity because many properties are shared with other categories at the same level of abstraction (e.g. setter, pointer, golden retriever). The most

inclusive category within which attributes are common to most category members is what Rosch and colleagues called the basic-level category, e.g. dog, and it is the category that has the maximum cue validity. Classifications more general than the basic level are called superordinate categories, while those that are more specific are called subordinate categories (Rosch et al 1976a).

When human subjects are asked to list as many attributes that apply to certain objects as they can, they report the greatest increase in the number of characteristic features when describing objects at the basic level (Rosch et al 1976a, Tversky & Hemenway 1984). Objects in the same basic-level category are also manipulated using common motor sequences and share considerable shape similarity with each other but not with objects of most other groups. The similarity of basic objects is such that shape-based averaging of two members of the category will often yield a new object that can also be recognized as a category member. In fact, in the case of a highly homogeneous group of objects like faces, photographic averages from two separate 20-person groups, selected according to gender and age, have been shown to yield two "average" faces remarkably similar to each other (Katz 1953).

Empirically, recognition of objects at the basic level often occurs more rapidly and more accurately than the recognition of objects at any other taxonomic level. Exceptions to this rule are atypical exemplars of basic categories that have pronounced shape differences from the prototype, or central tendency, of the class. For example, humans usually identify "penguins" or "racing cars" faster as such than as "birds" or "cars," which are the basic-level classifications for these objects (Jolicoeur et al 1984, Murphy & Brownell 1985). To describe the level at which specific objects are first accessed irrespective of inclusiveness or cue validity, Jolicoeur et al (1984) coined the term "entry point" of recognition.

Interestingly, the entry point of individual objects, which usually coincides with the basic level of classification, can shift to the subordinate level when perceivers become especially sensitive to subtle differences between objects of the same class. For example, Rosch et al (1976a) noted major differences in the descriptions of object attributes between experts in a field and unspecialized subjects. An airplane mechanic, for instance, when asked to list airplane attributes, spontaneously reported a large number of attributes of airplanes that are potentially available to the casual observer but that are usually ignored. In a systematic study of this phenomenon, Tanaka & Taylor (1991) showed that for experts in a field (such as bird watching), subordinate categories become as differentiated as basic-level categories. This entry point change is also evident in naming latencies, which become as short as those of the basic classifications as expertise increases.

In summary, humans systematically categorize objects in the world based on natural groupings of attributes. Do, however, such categorization principles

also apply in situations where the observer has no prior conceptual information about the objects to be classified? Most importantly, do they apply for the nonverbal observer? If natural categories do develop independently of pre-existing conceptual or linguistic labels, then the same principles of categorization may underlie the recognition skills of other animals, in which the neural representations of objects can be studied directly using neurophysiological techniques. At least three lines of evidence, discussed briefly below, suggest that categorization may indeed rely on principles applying to any recognition system.

Perceptual Categorization

LEARNING NOVEL STIMULI Evidence suggesting some universal principles in the formation of categories comes from recognition experiments with visually novel objects that are unrelated to any previously experienced verbal codes or abstract concepts. Such experiments show that in the process of learning basic objects, humans can detect consistent features of minimum interindividual variability, e.g. features of high cue validity, that characterize most exemplars of an object class, thereby extracting class invariances. In their seminal work, Posner & Keele (1968) probed the representations stored by humans when learning to classify patterns with individual variance around a common abstract structure. They used dot patterns (Figure 1a) as prototypes, and created individual category instances (exemplars) at specified distances, or deviations, from the original pattern by applying statistical distortion rules (Posner et al 1967). Their subjects were taught to classify distorted patterns constructed from three different prototypes, and they were subsequently tested in a recognition task in which they were exposed to the previously viewed patterns, the prototype pattern, and new, distorted patterns. Interestingly, subjects recognized the prototype pattern almost as quickly and as accurately as they recognized previously memorized patterns, even though they had never been directly exposed to it, suggesting that the prototype is a main constituent in the category's memorial representation (see also Franks & Bransford 1971, Strange et al 1970). Moreover, information about the central tendency of such sets of exemplars was found to be extracted and stored during learning and not during the process of recognition (Posner & Keele 1970, Homa et al 1973, Strange et al 1970).

Categorizations are not based upon a recognition threshold that, once exceeded, definitively endows a stimulus with class membership. In other words, class boundaries are not formed by sharp transition hypersurfaces in a multidimensional feature space. Instead, a familiarity continuum exists, according to which the probability of a correct classification depends on the structural typicality of the stimulus, determined by the closeness to the class prototype

