**Introductory Remarks to Symposium 21**

# Visual representations for memory and recognition

*Nikolaus F. Troje (1) and Martin Giurfa (2,3)*

**(1): Fakultät für Psychologie, Ruhr-Universität-Bochum, D-44780 Bochum, Germany**
**(2): AG Neurobiologie, Institut für Biologie, Freie Universität Berlin, Königin-Luise-Str. 28/30, D-14195 Berlin, Germany**
**(3): Laboratoire de Cognition Animale, Université Paul-Sabatier – Toulouse III, 118 Route de Narbonne, 31062 Toulouse Cedex 4, France**

One of the main tasks of the brain is to construct a perceptual world that corresponds to its environment. To this end, the information gathered by an organism is organised in the form of internal representations. Representational ability is based on the ability to encode a stimulus in memory and to activate such a trace even if the original stimulus is not available in the current environment. In that way, appropriate reactions to a changing world are possible.

Visual recognition is a fundamental task for humans and for most non-human animals. Animals are confronted with the necessity of recognising objects and movements on the basis of specific features that underlie visual representations. Such representations may vary in their degree of plasticity, i.e. in the possibility of generalising the stimulus trace to novel stimuli in the environment. Some representations may be extremely rigid due to a pixel-based organisation in the form of templates that require perfect matching to ensure recognition. Other visual representations may be extremely flexible as they can involve important levels of generalisation. For instance, generalisation of visual objects can be achieved by focussing the animal's selective attention on a specific feature common to a great variety of visual objects. Also, generalisation may be based on a set of combined features that create a specific configuration that has to be identified in novel stimuli to trigger recognition. Template matching, feature detection or configural learning are some of the possibilities available to the visual system.

The challenging task that the visual system (as any sensory system) confronts is that of recognition in an unpredictable world (see Fig. 1). It is very unlikely that two sensory events are identical. Consider face recognition as an example: recognising a face involves generalisation from views that have been experienced in the past to a novel previously unseen view. Even if the particular face is familiar to the observer, his/her experience with that face is limited to only a certain number of views. The current image cast onto the observer's retina is most likely one that was never seen before. The orientation of the face and the illumination may be slightly different. Also the particular shape of the face may be altered by facial expression and speech. Despite these variations, our visual system usually recognises the face. Effortlessly, it identifies the novel view with previously experienced instances of the same face. To this end it accesses a specific representation of the face that results from the observer's experience with that face. How does this process work? How does the visual system extract just those invariants that contain diagnostic information that helps to solve the recognition task? How does it transform the raw sensory information such as to establish a representation that supports generalization and extraction of diagnostic invariants? Of course, these questions are not only relevant to the human visual system. Any visual system must cope with roughly the same set of requirements and constraints and other sensory modalities also have to face the problem of a changing world.

**Figure 1:** Mark Tansey's "*The Innocent Eye Test*". The painting illustrates the fundamental questions on visual recognition and the underlying visual representations.

Although the demands are similar across species, the solutions implemented by an organism may depend on the resources and the past and present species-specific circumstances that define its visual environment. They may further depend on the particular visual task and on particular object classes of interest. The mechanisms for human face recognition might be different from the ones used by bees to recognize flowers. However, there may be still some common principles that are derived directly from the requirement to extract diagnostic invariants.

There are several ways to approach the problem of the nature of visual representations. One possible approach is to analyse the stimuli that a visual system must deal with in order to postulate mechanisms for recognition and classification that best fit the structure and statistics of the stimulus class. Another possible approach is to determine what a visual system does by analysing behaviour in experiments designed to deduce the mechanisms by which recognition is achieved. In such experiments, human subjects or animals are required to identify, classify or discriminate between visual stimuli. Both approaches are closely interrelated and feed into each other. Principles derived theoretically have to be tested in behavioural or psychophysical experiments and results from such experiments provide a basis for new theoretical principles. Together both approaches enable the formulation of hypotheses about the neuronal implementations of the observed behaviour. This, leads to a third important approach, namely the verification of such hypotheses on stimulus representation with physiological methods.

**Computational approaches**

The central question raised by these studies is: what kind of information is contained in the sensory message and how is it encoded? This question is crucial for both the scientist trying to understand the information flow from the ambient environment into the animals' perceptual system as well as for the animal itself that has to evolve an appropriate neurophysiological hardware to perceive and understand biologically relevant events in its world. A central tool is thus an algorithm for mapping the raw sensory data into a representational space such that their analysis and the search for diagnostic invariants is supported. In general, the mapping can be specific to a particular objects class or sensory category thus reflecting a-priori knowledge about the structure and statistics of the objects class. Human faces will certainly be represented differently from blossom patterns.

Many representational spaces that have been suggested in the past can be characterized by their ability to linearize sensory data. Generally highly non-linear input data are transformed such that they can be mapped into a linear space in a way that linear operations become meaningful. In an appropriate linear space, extracting diagnostic invariants from a set of instances of an object or an object class can now be achieved by averaging out noisy or irrelevant information. Generalization involves interpolation between already experienced instances, an operation that is best supported in a proper linear space.

**Behavioural studies**

Experiments testing visual discrimination are essential to delimit generalisation capabilities and thus the kind of underlying recognition. Such experiments usually rely on the capacity to associatively learn visual cues. Certainly, organisms may exhibit spontaneous preferences independent of conditioning, but the essential property common to humans and non-human animals is the fact that they can be *trained* to stimuli designed to answer precise questions on visual recognition. This is clearly illustrated by the model systems that are presented in this symposium: fruit flies, honeybees, electric fishes and pigeons, all learn to associate visual cues with an unconditioned stimulus. What changes from one model system to the other is the specific recognition strategy that, in turn, may depend on the stimuli used, on the training task or on the species-specific history. Under certain experimental circumstances, template-based recognition seems to occur in fruit flies, honeybees and electric fishes. However, more flexible performances based on elective attention to a single feature or to a combination of features can also be found in many animals such as fruit flies, honeybees and pigeons. Flexibility can even allow performing visual categorisations. Typically, categorisation is demonstrated if animals 1) solve a discrimination in which reward is not signalled by a single stimulus, but, rather, by a variety of stimuli that share some common characteristics, and 2) transfer the information about the reinforced features to novel instances. These prerequisites have been fulfilled in visual object recognition of animals as diverse as honeybees and pigeons. In both cases, categorisation experiments allow uncovering the features employed for recognition and thus making precise hypotheses about the underlying encoding mechanisms of visual information.

**Physiological studies**

A fundamental task of the cognitive neuroscientist is to access the nature of the internal representations. To this end, physiological experiments are useful to find out in which specific way the neural system implements a representation. Studying the response properties of single neurones or of ensembles of neurones when the experimental subject is visually stimulated with given visual cues is a traditional method that allowed understanding the basic nature of visual encoding in the mammalian cortex. The use of similar methods for recording from single units in the insect brain has allowed identifying neuronal elements dedicated to the processing of specific visual cues such as orientation detectors, movement detectors, colour-opponent neurones, etc. These findings again emphasize that the mammalian and the insect brain certainly differ in their degree of complexity but have implemented similar principles to cope with the problem of visual recognition. More recently, another traditional approach, the use of brain-damaged subjects with naturally occurring lesions, has been complemented with functional neuroimaging studies. This now allows studying visual recognition in the normal human brain. PET and fMRI studies have attempted to isolate the substrates of human visual recognition but their outcome is still unsatisfactory due to the lack of agreement between studies and the scatter in the identified neural substrates. Nevertheless improvements in these techniques are expected such that they can be extremely valuable in detecting how, where and when a visual representation is built in the brain

**Conclusions**

The contributions presented at this symposium illustrate some of the various attempts that have been made to identify the processes involved in visual object representation and recognition. In choosing a comparative approach encompassing studies on invertebrates and vertebrates, we wanted to stress the variety of visual cues and strategies employed to achieve visual recognition, but also the generality of common principles implemented to this end. The study of diversity but also of generality may allow understanding how different brains construct visual representations that correspond to their respective environments. To achieve such an understanding, the interactions between the three kinds of studies mentioned above, the theoretical, the behavioural and psychophysical, and the physiological ones, must be intensified. The study of visual recognition and the nature of visual representation is a task that can be successful only if an integrative approach is chosen.