



Figure 1 (Bowers). The visual patterns A and a each map onto the phonological code /ei/, which results in the coactivation of A and a each time one of the patterns is presented, via feedback. Associative learning principles within the orthographic system act to map together these coactive patterns to produce an abstract letter code.

In this view, then, the perceptual categories for individual letters and words are structured in accordance with background knowledge. Various findings such as the word superiority and phoneme restoration effects have been described as compatible with the view that the relevant perceptual systems are encapsulated, because the top-down influences may occur within the module. However, as the above evidence indicate, the perceptual representations of words that might support these effects are located outside the putative early visual system.

I do not mean to suggest that these findings are incompatible with Pylyshyn's main thesis that early vision is informationally encapsulated. But the findings do restrict the types of computations that such a system may perform. At least in the domain of reading, the perceptual categories for words (and letters) reside outside the early visual system, and it remains to be seen whether other categories, such as structural descriptions of objects are completely determined on the basis of visual information, or whether nonvisual sources of evidence constrain this knowledge as well.

## Complexities of face perception and categorisation

Vicki Bruce, Steve Langton, and Harold Hill

Department of Psychology, University of Stirling, Stirling FK9 4LA, Scotland.  
[www.stir.ac.uk/departments/humansciences/psychology/](http://www.stir.ac.uk/departments/humansciences/psychology/)  
[vicki.bruce@stir.ac.uk](mailto:vicki.bruce@stir.ac.uk)

**Abstract:** We amplify possible complications to the tidy division between early vision and later categorisation which arise when we consider the perception of human faces. Although a primitive face-detecting system, used for social attention, may indeed be integral to "early vision," the relationship between this and diverse other uses made of information from faces is far from clear.

Zenon Pylyshyn is to be congratulated for setting out so clearly his "natural constraints" theory of early vision (EV). Here we amplify possible complications to the tidy division between EV and later categorisation which arise when we consider the perception of human faces.

In Note 13 (sect. 5.1), Pylyshyn raises one possible exception to the rule that surface descriptions within EV cannot be influenced a top-down manner. A hollow face mask, when viewed from a certain distance, will appear as a real face, with its nose nearer than

its eyes to the viewer. This categorisation overrides the assumption that light tends to come from above. To be consistent with actual lighting from above, the illusory face will be seen lit from below. Actually the illusion over-rides more than just apparent direction of lighting. When the mask is seen as a face, the percept also overrides normal constraints about the way in which an object occludes itself when a viewer moves in front of it. When seeing the face, an observer moving from side to side will see the face move to follow them in a most alarming fashion.

The illusion appears first to have been noticed by the Scottish scientist, Sir David Brewster (see Wade 1983; and Bruce & Young, 1998, for discussion), but was popularised by psychologist Richard Gregory (e.g., 1973) as an example par excellence of vision as hypothesis-testing. The face is such a likely and important event in the world (compared with a hollow face) that we persist in seeing it despite the reinterpretation of other visual information that is demanded.

However, the hollow face illusion need not necessarily result from face-specific processing. Another constraint is satisfied by the face over the mask – that objects in the world are normally convex. Johnston et al. (1992) noted a similar illusion using a quite unfamiliar convex shape – the "hollow" potato." Hill and Bruce (1993; 1994) set out systematically to investigate the influence of a number of factors on the strength of the illusion. We used the simple method of requiring participants to walk towards or away from the mask, stopping at the point where their perception reversed from concave to convex (or vice versa). Their distance from the mask at this point formed the dependent variable. We showed that upright faces gave a stronger effect than upside-down ones or unfamiliar shapes, which did not differ; bottom lighting gave a stronger illusion than top lighting; and viewing with one eye gave a stronger illusion than two – and these effects appear to be independent. So, there seems to be a preference for convexity, but an additional one for the face shape over other convex but unfamiliar configurations. Our results were consistent with the idea that a set of independent modules (for stereo, for shape-from shading and so forth) each fed information to a common stage where representation of 3D surface was computed – in Marr's terms, the 2.5-D sketch. On this model, the "familiar shape" or, perhaps "face detection" module would access the same stage.

Is it a cop out to allow one kind of categorisation to sneak in to affect EV in this way? Before assuming that it is only faces that gave the advantage over all other convex shapes, we would have to check that other kinds of object do not give the same effect. The prediction must be that the illusion would be equally strong for an upright or inverted hollow dog, for example, each behaving like an inverted face – showing an influence of the convexity constraint alone. Assuming such a result were obtained (and we hope that someone will now feel inspired to dip their pet into liquid plaster to find out) how does the face-detection module get in to influence EV?

There is certainly strong evidence that a face detector is innate. Goren et al. (1975) and Johnson et al. (1991) found that minutes-old newborn babies would follow face-like patterns more with their face and eyes than non-face control patterns. Recent evidence from our lab (Langton & Bruce, in press) and Jon Driver's (Driver et al., in press) suggests that faces *automatically* redirect the attention of someone viewing them. Our experiments made use of the Posner cuing paradigm, where the task is simply to respond to the onset of a target located at one of four peripheral locations. Target detection was faster when the target's appearance was preceded by a head/eye display looking towards its location. Moreover, the effect resembles *exogenous* attentional control – effects were strongest at the shortest SOA (stimulus onset asynchrony) and were found even when the cues were entirely uninformative. So, we would argue that face-ness, perhaps associated with information about head angle and direction of gaze, is a very low-level property indeed.

What is much less clear, however, is the relationship between a primitive face-detecting system, used for social attention, and the

myriad and complex other systems involved in deciphering different kinds of meaning from the face. Faces are not just categorised *as faces*, but as male or old or pretty or intelligent-looking faces, as the face of a pop star or of the President, as a face looking worried or content. The kinds of visual descriptions needed for these different categorisations are very different, and there are neuropsychological dissociations between different kinds of use made of facial information. Thus identification or expression processing may be relatively impaired or spared following different kinds of brain injury (e.g., Young et al. 1993). The face even manages to influence speech processing, as in the McGurk effect (McGurk & Macdonald 1976), and visual facial speech processing also doubly dissociates from expression and identity processing (Campbell et al. 1986). The idea that there is a modular EV stage feeding a categorical cognitive system seems too simple, and begs a number of really interesting and difficult questions about the flexibility of representations needed for different kinds of socially important activity. Such distinctions go well beyond the division between action-oriented perception and object recognition discussed within the target article. So, while we like Pylyshyn's essay very much, we feel it still involves an over-simple distinction between "seeing" and "seeing as" (Fodor & Pylyshyn 1981) – and doesn't say enough about different varieties of "seeing for."

## Visual perception is too fast to be impenetrable to cognition

Jean Bullier

*Cerveau et Cognition, INSERM 371, 69675 Bron Cedex, France.*  
**bullier@lyon151.inserm.fr**

**Abstract:** Neuroscience studies show many examples of very early modulation of visual cortex responses. It is argued that such early routing is essential for a rapid processing of information by the visual system.

The evidence presented in the target article for the inflexibility of early vision is impressive but it is mostly based on results from experimental psychology. Recent work in the neurosciences reveals that the situation is probably more complicated than argued by Pylyshyn.

Much of the difficulty of interpretation lies with the definition of early vision. Although Pylyshyn denies the possibility of localizing it in specific structures, section 3.2 implies that it corresponds to information processing in the visual cortex. According to Pylyshyn's thesis, there should be no modulation of activity in visual cortex other than that related to spatial attention. There is clear evidence to the contrary: many reports have established that neuronal responses in occipital, parietal, frontal, and inferotemporal visual cortex of monkeys depend on the significance of the stimulus for subsequent behavior (eye or arm movement). Significance can be attached to the orientation, the direction of movement, the shape or the color of the stimulus (Chelazzi et al. 1993; Haenny & Schiller 1988; Mottler 1994a; 1994b; Schall et al. 1995; Treue & Maunsell 1996) or it may involve cueing through another sensory modality (Haenny et al. 1988).

The relationship between visual perception and single cell activity has recently been the subject of much interest. Binocular rivalry provides a powerful approach to this question. When two different stimuli are presented in each eye, monkeys, as well as humans, experience sudden switches between the two competing percepts. In the inferotemporal cortex and, to a lesser extent, in lower-order visual areas, neuronal responses change dramatically when the animal signals that its visual perception flips between the two percepts (Leopold & Logothetis 1996; Sheinberg & Logothetis 1997). Similar findings have been recently demonstrated in the human brain: metabolic activity of occipital, temporal, and parietal cortex appears to follow the changes in percepts without any changes in the stimuli (Lumer et al. 1998). Such modulations

demonstrate that the neuronal firing in inferotemporal cortex gives a closer image of the perception than of the stimulus that triggers it, although it is not clear whether the modulations are driven by bottom-up or top-down processes.

Another way to define early vision is to use the temporal dimension. Early vision corresponds to the early parts of the responses to visual stimulation. Measurements of the latencies of neurons to visual stimulation show substantial overlap in the different cortical areas of the primate. Numerous neurons in area V1 are activated later than some neurons in frontal, parietal or inferotemporal cortex (Nowak & Bullier 1997; Schmolesky et al. 1998). It is therefore difficult to reduce early vision to extrastriate visual cortex and most cortical areas contain neurons that respond early to visual stimulation. In this respect, the results of Schall and his collaborators are remarkable: they show that the visual responses of neurons in the primate frontal cortex are modulated at very short latencies (50–70 msec post stimulus) by the color of a stimulus that signals whether or not it is to be the target of an eye movement (Bichot et al. 1996). These effects are delayed by less than 10 msec with respect to visual responses in area V1, thus demonstrating the capacity of the decision system to influence *very* early vision.

There have also been several reports of early influences of categorization on visually evoked responses in human. Thorpe and his collaborators demonstrated that categorizing visual scenes as containing or not containing an animal modulates the strengths of responses as early as 150 msec after the stimulus (Thorpe et al. 1996). Recent results (Thorpe et al., personal communication) suggest that this modulation takes place in the inferotemporal cortex. Other groups have also reported early modulation of responses in inferotemporal cortex by whether or not a stimulus can be interpreted as a human face (Bentin et al. 1996; Jeffreys 1996).

In conclusion, evidence from the neurosciences indicates that early visual responses are strongly modulated by the intention of the individual to make a movement to a target or by the categorization of visual stimuli. Thus, at least in such experimental situations, cognition penetrates down to the earliest levels of the visual system to facilitate responses relevant to subsequent behavior. Such facilitation effects appear to be mediated by feedback connections (Hupé et al. 1998; Lamme 1995; Lamme et al. 1998). The probable reason for the penetrability of early vision by the cognitive system is the visual system's need to process information rapidly despite the slow nature of computations by individual neurons (Nowak & Bullier 1997). Together with massive parallelism, early routing of information (i.e., penetrability) is one of the main strategies set up by the nervous system to achieve its remarkable capacities in the temporal domain.

## The cognitive impenetrability of cognition

Patrick Cavanagh

*Department of Psychology, Harvard University, Cambridge, MA 02138.*  
**patrick@wjh.harvard.edu visionlab.harvard.edu**

**Abstract:** Cognitive impenetrability is really two assertions: (1) perception and cognition have access to different knowledge bases; and (2) perception does not use cognitive-style processes. The first leads to the unusual corollary that cognition is itself cognitively impenetrable. The second fails when it is seen to be the claim that reasoning is available only in conscious processing.

Pylyshyn argues that some parts of vision are not subject to the operation of reasoning and problem solving – they are cognitively impenetrable. He notes that perceptual decisions are often made despite the observer's knowledge that the percept must be wrong. This makes perception irrational in Pylyshyn's view, leading to the