short-term memories do not penetrate the visual system. However, evidence suggests that the ventral stream is not purely perceptual. In delayed-match-to-sample tasks, where a monkey has to indicate whether a sample matches a previously presented cue, V4 responses are often better related to cue responses long after the cue has disappeared and a subsequent sample has appeared (Ferrera et al. 1994). This shows clearly that a cognitive component is present in V4, one not reducible to effects of attention. On the other hand, it is also clear that V4 contributes significantly to visual processing (Schiller & Lee 1991; de Weerd et al. 1996). Thus V4 appears to violate the impenetrability of visual perception. Similarly, the dorsal stream too exhibits delay responses, for example, in the lateral imtraparietal area (LIP; Gnadt & Andersen 1998), while being involved in visual processing, such as the representation of salience (Gottlieb et al. 1998). Thus it appears that in addition to attentional effects, short-term memory effects need to be added to the possible cognitive penetration of the visual system.

A further feature that cognitive impenetrability seems to require is that the task being performed only be reflected in how attention is allocated within the visual system (sect. 4.3). In other words, the prediction for PET studies would be that whereas different visual areas may get more or less activated depending on the task, the underlying perceptual network should be quite similar. A network analysis, however, has shown that depending on the task very different areas cooperate, even within the visual system (McIntosh et al. 1994). Hence it seems that the effect of a task is not limited to reallocation of attentional resources; instead, a nonattentional task-dependent component can affect visual processing.

In his target article, Pylyshyn allows for the possibility that the visual system may have to be relabeled the "spatial system," because it may well be that multimodal information converges before cognition contributes (sect. 7.1). This possibility receives support from the finding that auditory stimuli are perceived as early, even when subjects know they are simultaneous. The place at which auditory and visual signals converge need not be a "central" representation. Indeed, neurons in the inferior temporal cortex (IT) respond to auditory stimuli if they are paired with visual stimuli but not otherwise (Gibson & Maunsell 1997), as if IT neurons coded the association of auditory and visual stimuli. This association is not spatial per se, however; rather, it is based on identity. Similarly, responses in area V4 can reflect tactile signals (Maunsell et al. 1991). Thus it seem extraretinal signals can enter the visual system even if their spatial component is not the important feature.

According to the notion of cognitive impenetrability, there is a hard division between early vision and late vision where only attention is able to affect early vision (sect. 4.3). This division is reminiscent of the one between sensory and motor processing stages. Although this is easy to identify at the level of sensory transducers and of muscles, it is a lot fuzzier closer to the sensorimotor transformation. Evidence from the dorsal stream suggests that even quite early parietal areas already code the intention to make movements (Snyder et al. 1997). [See also Jeannerod: "The Representing Brain" *BBS* 17(2) 1994.] The ventral stream, on the other hand, is involved in functions that also lead to movements, but at different time scales (Goodale 1998). It appears then, that the brain may not be divided "horizontally" into different processing stages, but rather "vertically" into different parallel sensorimotor circuits, each subserving a different competence that is called upon, depending on the context. Similarly, perceptual and cognitive factors may not be divided by a hard line, but may be interwoven into different sensory-cognitive circuits that can be recalled selectively, depending on the capability required. Most circuits, however, are likely to contain visual components that are cognitively impenetrable.

Vision and cognition: Drawing the line

Andrew Hollingworth[a] and John M. Henderson[b]

[a]*Department of Psychology, Michigan State University, East Lansing, MI 48824-1117.* [b]*Department of Psychology and Cognitive Science Program, Michigan State University, East Lansing, MI 48824-1117.*

andrew@eyelab.msu.edu   eyelab.msu.edujohn@eyelab.msu.edu

Abstract: Pylyshyn defends a distinction between early visual perception and cognitive processing. But exactly where should the line between vision and cognition be drawn? Our research on object identification suggests that the construction of an object's visual description is isolated from contextually derived expectations. Moreover, the matching of constructed descriptions to stored descriptions appears to be similarly isolated.

As Pylyshyn states in his target article, few would argue that cognitive representations such as contextually derived expectations or beliefs modulate the types of information processed at the very earliest stage of vision (i.e., retinal stimulation). Thus, the critical question is: At what functional stage of perceptual processing do such representations begin to interact with visual information derived from the retinal image? Current computational theories of visual perception tend to break down the perception of meaningful stimuli into three functional stages. First, primitive visual features (e.g., surfaces and edges) are extracted from retinal information. Second, these features are used to construct a description of the structure of a stimulus. Third, the constructed description is matched against stored descriptions. Pylyshyn argues that the line between vision and cognition should be drawn between stages two and three. Specifically, cognitively derived expectations and beliefs do not interact with visual processing up to the construction of a visual description, but may influence the matching stage, perhaps by modulating the threshold amount of activation necessary to trigger a match to a particular object type.[1]

Recent research in our laboratory, however, indicates that the division between vision and cognition may occur even further upstream than Pylyshyn suggests, at least in the realm of real-world object identification. We have used a forced-choice object discrimination paradigm (similar to that developed by Reicher, 1969) to investigate the influence of scene context on object identification while avoiding the interpretative difficulties of signal detection methodology (Hollingworth & Henderson 1998; in press). In this paradigm, participants see a real-world scene for a short time (150–250 msec.). The scene contains a target object that is either semantically consistent with the scene (i.e., likely to appear) or semantically inconsistent (i.e., unlikely to appear). The scene is followed by a brief pattern mask, which is followed by two object alternatives of equivalent semantic consistency, only one of which corresponds to the target object. The participants' task is to indicate which object alternative had been presented in the scene.

To test whether expectations derived from meaningful scene context interact with the initial perceptual analysis of objects in the scene, we employed a token discrimination manipulation (Hollingworth & Henderson, in press). The forced-choice screen presented a picture of the target object and a picture of a different token of that conceptual type (e.g., a sedan and a sports car). If consistent scene context interacts with early visual processing to facilitate the visual analysis of consistent objects (see e.g., Biederman et al. 1982; Boyce et al. 1989), token discrimination should be better when the target object is semantically consistent versus inconsistent with the scene in which it appears. Contrary to this prediction, no such advantage was obtained. To test whether meaningful scene context interacts with the matching stage of object identification, we employed an object type discrimination manipulation (Hollingworth & Henderson 1998; in press). After presentation of the scene, the forced-choice screen contained a label describing the target object and a label describing another object of equivalent semantic consistency but of a different conceptual type (e.g., "chicken" and "pig" after a farm scene). If consistent scene context reduces the amount of perceptual information needed to reach threshold activation indicating that a particular object type is present (see, e.g., Friedman 1979; Palmer 1975), object type discrimination should be better when the target object is consistent versus inconsistent with the scene. Contrary to the prediction of this weaker version of the interactive hypothesis, we found no advantage for the discrimination of consistent object types. In fact, one experiment revealed a reliable advantage for inconsistent object discrimination. These results suggest that object identification occurs essentially independently of contextually derived expectations, though such information can be used post-perceptually to make decisions about which objects are likely or unlikely to have been present in a scene (Hollingworth & Henderson 1998).

How could the matching of constructed object description to stored descriptions occur independently of semantic information stored in memory about that object type? We propose that a constructed object description is matched to stored descriptions pre-semantically, with a successful match allowing access to semantic information about that object type. Thus, the hypothesized knowledge base for a visual module responsible for object identification would include (1) visual features and the routines necessary to compute a visual description from these features and (2) stored descriptions of object types. This general framework is consistent with the behavioral, neuropsychological, and neuroscientific evidence reviewed by Pylyshyn indicating that early visual processing is isolated from general world knowledge. In addition, it is consistent with current theories of object recognition that propose little or no role for cognitively derived expectations or beliefs in object recognition (Biederman 1995; Bülthoff et al. 1995; see also Marr & Nishihara 1978). Finally, it is consistent with the evidence from the implicit memory literature that there are independent memory systems for the representation of object form (i.e., a structural description system) and the representation of semantic and other associative information about objects (Schacter 1992).

In summary, our research suggests that the visual subsystems responsible for constructing a description of a visual stimulus and for comparing that description to stored description are functionally isolated from knowledge about the real-world contexts in which objects appear. This research supports Pylyshyn's proposal that much of the important work of vision takes place independently of expectations and beliefs derived from semantic knowledge about real-world contingencies.

NOTE

1. Clouding this issue somewhat, Pylyshyn proposes that higher-order visual primitives (e.g., geons under Biederman's 1987 theory) should be considered part of the semantic system that does not interact with early vision. However, it seems likely to us (and consistent with Pylyshyn's larger thesis) that higher-order visual primitives could comprise part of the non-semantic knowledge base of a visual module dedicated to the construction of a three-dimensional visual description.

We all are Rembrandt experts – or, How task dissociations in school learning effects support the discontinuity hypothesis

Régine Kolinsky[a,b] and José Morais[b]

[a]*Fonds National de la Recherche Scientifique, B-1000, Brussels, Belgium;*
[b]*Laboratoire de Psychologie Expérimentale, Université Libre de Bruxelles, CP. 191, B-1050 Brussels, Belgium.*
rkolins@ulb.ac.be jmorais@ulb.ac.be

Abstract: We argue that cognitive penetration in non-early vision extends beyond the special situations considered by Pylyshyn. Many situations which do not involve difficult stimuli or require expert skills nevertheless load on high-level cognitive processes. School learning effects illustrate this point: they provide a way to observe task dissociations which support the discontinuity hypothesis, but they show that the scope of visual cognition in our visual experience is often underestimated.

Pylyshyn's main claim is that there is a discontinuity between "early" vision and cognition. We certainly agree with this. We also agree with his acknowledgement that vision as a whole is cognitively penetrable, being modulated by attentional and decisional factors. However, to illustrate penetration of non-early vision by cognition, Pylyshyn presents rather special cases of visual processing (their being special he himself acknowledges): he refers either to tasks which obviously include problem solving, that is, search on difficult-to-perceive stimuli such as fragmented figures, or to the case of trained experts, who are clearly more able than (we) novices to authenticate a Rembrandt or to determine the sex of chicks.

We will argue that cognitive penetration in non-early vision extends far beyond these special tasks, stimuli or observers. Our claim does not concern decisional or response selection processes as examined by Signal Detection Theory or ERP studies (about which we agree with most of Pylyshyn's arguments). Rather, we claim that many situations which do *not* involve difficult stimuli or require expert skills nevertheless load on high-level, cognitive processes. School learning effects can be used to make this point clear, as they might provide a methodological tool to observe task dissociations which, we will argue, ultimately support the discontinuity hypothesis.

Earlier reports on school learning effects in vision stem mainly from studies, which, under the impact of Vygotsky's approach to cognitive development and of the transactional functionalism and New Look movements (e.g., Bruner 1957; Ittelson 1952), stressed the individuals' social and cultural differences. Yet many cross-cultural studies either examined high-level representations (like the use of functional vs. perceptual categorization criteria, e.g., Greenfield & Bruner 1966), or failed to control for correlated (genetic and environmental) variables (see Deregowski 1989 and associated commentaries).

Nevertheless, in the last twenty years, many experimental studies have been devoted to a special sort of school learning, namely alphabetization. The consequences of acquiring an alphabetic system for mental representation were stressed in developmental studies (e.g., Liberman et al. 1974) and later in adult studies (e.g., Morais et al. 1979). For our purpose, what matters is that, before that seminal work, no distinction was made between, on the one hand, *perceptual discrimination* among phonemes (e.g., distinguishing between "cat" and "rat") and, on the other hand, *phonemic awareness,* namely, the explicit representation of speech as a sequence of phonemes, as demonstrated in phoneme counting, deletion or reversal. This distinction was suggested by the observation that pre-literate children and illiterate adults are unable to perform intentional operations at the level of the phoneme while most literate children and *ex-illiterates* who learned to read and write as adults succeed in these tasks. Lack of phonemic awareness does not prevent the illiterates from being perfectly able to discriminate between pairs of stimuli that differ only in one phoneme or phonetic feature (Adrián et al. 1995; Scliar-Cabral et al. 1997). This dissociation has led to various theoretical developments (e.g., Kolinsky 1998; Morais & Kolinsky 1994; 1995).

Going back to vision, we suggest that comparing schooled to unschooled people can provide new insights into the distinction between what we call visual perception (early vision, according to Pylyshyn's terminology) and *visual cognition* (that part of vision penetrated by cognition according to Pylyshyn).

Our own studies have shown that unschooled adults have serious difficulties performing tasks like part-verification, dimensional filtering, and orientation judgment, which require that attention is directed to a specific component of the stimuli (e.g., Kolinsky et al. 1990; 1987). By contrast, no difference is observed between unschooled and schooled adults in tasks which do not require such explicit selective attention and analysis, for example, when separability of parts or dimensions as well as line orientation registration are estimated by the occurrence of illusory conjunctions (i.e., errors in which properties correctly extracted from several objects are blended into a new, mentally created object,