BEHAVIORAL AND BRAIN SCIENCES (1999) **22,** 341–423
Printed in the United States of America

Is vision continuous with cognition? The case for cognitive impenetrability of visual perception

Zenon Pylyshyn
Rutgers Center for Cognitive Science,
Rutgers University, New Brunswick, NJ 08903
zenon@ruccs.rutgers.edu **ruccs.rutgers.edu/faculty/pylyshyn.html**

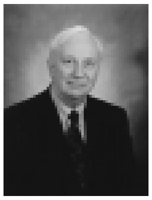
Abstract: Although the study of visual perception has made more progress in the past 40 years than any other area of cognitive science, there remain major disagreements as to how closely vision is tied to cognition. This target article sets out some of the arguments for both sides (arguments from computer vision, neuroscience, psychophysics, perceptual learning, and other areas of vision science) and defends the position that an important part of visual perception, corresponding to what some people have called early vision, is prohibited from accessing relevant expectations, knowledge, and utilities in determining the function it computes – in other words, it is cognitively impenetrable. That part of vision is complex and involves top-down interactions that are internal to the early vision system. Its function is to provide a structured representation of the 3-D surfaces of objects sufficient to serve as an index into memory, with somewhat different outputs being made available to other systems such as those dealing with motor control. The paper also addresses certain conceptual and methodological issues raised by this claim, such as whether signal detection theory and event-related potentials can be used to assess cognitive penetration of vision.

A distinction is made among several stages in visual processing, including, in addition to the inflexible early-vision stage, a pre-perceptual attention-allocation stage and a post-perceptual evaluation, selection, and inference stage, which accesses long-term memory. These two stages provide the primary ways in which cognition can affect the outcome of visual perception. The paper discusses arguments from computer vision and psychology showing that vision is "intelligent" and involves elements of "problem solving." The cases of apparently intelligent interpretation sometimes cited in support of this claim do not show cognitive penetration; rather, they show that certain natural constraints on interpretation, concerned primarily with optical and geometrical properties of the world, have been compiled into the visual system. The paper also examines a number of examples where instructions and "hints" are alleged to affect what is seen. In each case it is concluded that the evidence is more readily assimilated to the view that when cognitive effects are found, they have a locus outside early vision, in such processes as the allocation of focal attention and the identification of the stimulus.

Keywords: categorical perception; cognitive penetration; context effects; early vision; expert perception; knowledge-based vision; modularity of vision; natural constraints; "new look" in vision; perceptual learning; signal detection theory; stages of vision; top-down processes; visual agnosia; visual attention; visual processing

1. Introduction

The study of visual *perception* is one of the areas of cognitive science that has made the most dramatic progress in recent years. We know more about how the visual system works, both functionally and biologically, than we know about any other part of the mind-brain. Yet the question of why we see things the way we do in large measure still eludes us: Is it only because of the particular stimulation we receive at our eyes, together with our hard-wired visual system? Or is it also because those are the things we expect to see or are prepared to assimilate in our mind? There have been, and continue to be, major disagreements as to how closely perception is linked to cognition – disagreements that go back to the nineteenth century. At one extreme some see perception essentially as building larger and larger structures from elementary retinal or sensory features. Others accept this hierarchical picture but allow centripetal or top-down influences within a circumscribed part of vision. Then there is "unconscious inference," first proposed by von Helmholtz and rehabilitated in modern times in Bruner's (1957) New Look movement in American psychology. According to this view, the perceptual process is like science itself; it consists in finding partial clues (either from the world or from one's knowledge and expectations),



Zenon W. Pylyshyn is Board of Governors Professor of Cognitive Science at the Center for Cognitive Science, Rutgers University. He has published over 60 papers in journals of philosophy, psychology, and engineering, and written or edited five books. He has been president of the Society for Philosophy and Psychology and the Cognitive Science Society. He is a Fellow of the Canadian Psychological Association and the American Association for Artificial Intelligence. He received the Donald O. Hebb Award for distinguished contributions to psychology as a science, and was recently elected Fellow of the Royal Society of Canada.

0140-525X/99 $12.50

formulating a hypothesis about what the stimulus is, checking the data for verification, and then either accepting the hypothesis or reformulating it and trying again in a continual cycle of hypothesize-and-test.

Perhaps not too surprisingly, the swings in popularity of these different views of visual perception occurred not only in the dominant schools of psychology, but were also echoed, with different intensity and at somewhat different times, in neuroscience, artificial intelligence, and philosophy of science. In section 3, we will sketch some of the history of these changes in perspective, as a prelude to arguing for an independence or discontinuity view, according to which a significant part of vision is cognitively impenetrable to beliefs and utilities.[1] We will present a range of arguments and empirical evidence, including evidence from neuroscience, clinical neurology, and psychophysics, and will address a number of methodological and conceptual issues surrounding recent discussions of this topic. We will conclude that although what is commonly referred to as "visual perception" is potentially determined by the entire cognitive system, there is an important part of this process – which, following roughly the terminology introduced by Marr (1982), we will call *early vision*[2] – that is impervious to cognitive influences. First, however, we will need to make some salubrious distinctions: we will need to distinguish between perception and the determination of perceptual *beliefs,* between the semantically coherent or rational[3] influence of beliefs and utilities on the content of visual perception,[4] and a cognitively mediated directing of the visual system (through focal attention) toward certain physical properties, such as certain objects or locations. Finally, since everyone agrees that some part of vision must be cognitively impenetrable, in section 7 we will examine the nature of the output from what we identify as the impenetrable visual system. We will show that it is much more complex than the output of sensors and that it is probably not unitary but may feed into different post-perceptual functions in different ways.

First, however, we present some of the evidence that moved many scientists to the view that vision is continuous with and indistinguishable from cognition, except that part of its input comes from the senses. We do this in order to illustrate the reasons for the received wisdom, and also to set the stage for some critical distinctions and some methodological considerations related to the interpretation generally placed on the evidence.

In 1947 Jerome Bruner published an extremely influential paper, called "Value and need as organizing factors in perception" (cited in Bruner 1957). This paper presented evidence for what was then a fairly radical view; that values and needs determine how we perceive the world, down to the lowest levels of the visual system. As Bruner himself relates in a later review paper (Bruner 1957), the "Value and need" essay caught on beyond expectations, inspiring about 300 experiments in the following decade, all of which showed that perception was infected through and through by the perceiver's beliefs about the world being perceived: hungry people were more likely to see food and to read food-related words, poor children systematically overestimate the size of coins relative to richer children, and anomalous or unexpected stimuli tend to be assimilated to their regular or expected counterparts.

Bruner's influential theory (Bruner 1957) is the basis of what became known as the "New Look in Perception." According to this view, we perceive in cognitive categories. There is no such thing as a "raw" appearance or an "innocent eye": we see something as a chair or a table or a face or a particular person, and so on. As Bruner put it, "all perceptual experience is necessarily the end product of a categorization process" and therefore "perception is a process of categorization in which organisms move inferentially from cues to category identity and . . . in many cases, as Helmholtz long ago suggested, the process is a silent one." According to Bruner, perception is characterized by two essential properties: it is categorical and it is inferential. Thus it can be thought of as a form of problem solving in which part of the input happens to come in through the senses and part through needs, expectations, and beliefs, and in which the output is the category of the object being perceived. Because of this there is no distinction between perception and thought.[5] [See also Schyns et al.: "The Development of Features in Object Concepts," *BBS* 21(1), 1998.]

Thousands of experiments performed from the 1950s through the 1970s showed that almost anything, from the perception of sentences in noise to the detection of patterns at short exposures, could be influenced by subjects' knowledge and expectations. Bruner cites evidence as far-ranging as findings from basic psychophysics to psycholinguistics and high-level perception – including social perception. For example, he cites evidence that magnitude estimation is sensitive to the response categories with which observers are provided, as well as the anchor points and adaptation levels induced by the set of stimuli, from which he concludes that cognitive context affects such simple psychophysical tasks as magnitude judgments. In the case of more complex patterns there is even more evidence for the effects of what Bruner calls "readiness" on perception. The recognition threshold for words decreases as the words become more familiar (Solomon & Postman 1952). The exposure required to report a string of letters shown in a tachistoscope varies with the predictability of the string (Miller et al. 1954): random strings (such as YRULPZOC) require a longer exposure for recognition than strings whose sequential statistics approximate those of English text (such as VERNALIT, which is a nonword string constructed by sampling 4-letter strings from a corpus of English text), and the higher the order of approximation, the shorter the required exposure. The signal-to-noise ratio at which listeners can recognize a word is lower if that word is part of a sentence (where it could be predicted more easily) or even if it occurs in a list of words whose order statistically approximates English (Miller 1962).

Similar results are found in the case of nonlinguistic stimuli. For example, the exposure duration required to correctly recognize an anomalous playing card (e.g., a black ace of hearts) is much longer than that required to recognize a regular card (Bruner & Postman 1949). Also, as in the letter and word recognition cases, the perceptual thresholds reflect the relative probabilities of occurrence of the stimuli, and even their relative significance to the observer (the latter being illustrated by studies of so-called "perceptual defense," wherein taboo words, or pictures previously associated with shock, show elevated recognition thresholds).

The results of these experiments were explained in terms of the accessibility of perceptual categories and the hypothesize-and-test nature of perception (where "hypotheses" can come from any source, including immediate context, memory, and general knowledge). There were also experiments that investigated the hypothesize-and-test view more directly. One way this was done was by manipulating the "availability" of perceptual hypotheses. For example, Bruner and Minturn (1955) manipulated the readiness of the hypothesis that stimuli were numbers versus letters (by varying the context in which the experiment was run), and found that ambiguous number-letter patterns (e.g., a "B" with gaps so that it could equally be a "13") were reported more often as congruous with the preset hypothesis. Also, if a subject settles on a false perceptual hypothesis in suboptimal conditions (e.g., with an unfocused picture), then the perception of the same stimuli is impaired (Bruner & Potter 1964).

Because of this and other evidence showing contextual effects in perception, the belief that perception is thoroughly contaminated by cognition became received wisdom in much of psychology, with virtually all contemporary elementary texts in human information processing and vision taking that assumption for granted (e.g., Lindsay & Norman 1977; Rumelhart 1977; Sekuler & Blake 1994). The continuity view also became widespread in philosophy of science. Philosophers of science such as Hanson (1958), Feyerabend (1962), and Kuhn (1972) argued that there was no such thing as objective data because every observation was contaminated by theory. These scholars frequently cited the New Look experiments showing cognitive influences on perception to support their views. Mid-twentieth-century philosophy of science was ripe for the new holistic all-encompassing view of perception that integrated it into the general framework of induction and reasoning.

The view that perception and cognition are continuous is all the more credible because it comports well with everyday experience. The average person takes it for granted that how we see the world is radically influenced by our expectations (not to mention our moods, our culture, etc.). Perhaps the most dramatic illustration of this is magic, where the magician often manipulates what we see by setting up certain false expectations. But there are also plenty of everyday observations that appear to lead to the same conclusion: when we are hungry we seem to mistake things for food and when we are afraid we frequently mistake the mundane for signs of danger. The popularity of the Sapir-Whorf hypothesis of linguistic relativity among the literate public also supports this general view, as does the widespread belief in the cultural effect on our way of seeing (e.g., the books by Carlos Castaneda). The remarkable placebo effect of drugs and of authoritative suggestions (even posthypnotic suggestions) also bears witness to the startling malleability of perception.

1.1. Where do we stand? – The thesis of this target article

Both the experimental and the informal psychological evidence in favor of the idea that vision involves the entire cognitive system appears to be so ubiquitous that you might wonder how anyone could possibly believe that a significant part of the visual process is separate and distinct from cognition. The reason we maintain that much of vision is distinct is not that we deny the evidence pointing to the importance of knowledge for visual apprehension (although in some cases we will need to reconsider the evidence itself), but that when we make certain distinctions the evidence no longer supports the knowledge-based view of vision. It is clear that what we believe about the world we are looking at *does* depend on what we know and expect. It is for that reason that we can easily be deceived – as we are in the case of magic tricks. But seeing is not the same as believing, the old adage notwithstanding, and this distinction needs to be respected. Another distinction that we need to make is between top-down influences in early vision and genuine cases of what I have called *cognitive penetration.* This distinction is fundamental to the present thesis. A survey of the literature on contextual or top-down effects in vision reveals that virtually all the cases cited are ones where the top-down effect is a within-vision effect – that is, visual interpretations computed by early vision affect other visual interpretations, separated either by space or time. The sort of influence that concerns us here originates outside the visual system and affects the content of visual perception (what is seen) in a certain meaning-dependent way that we call cognitive penetration. A technical discussion of the notion of cognitive penetrability and its implications for cognitive science is beyond the scope of this paper (but see Pylyshyn 1984, and "Computation and Cognition," *BBS* 3, 1980). For present purposes it is enough to say that if a system is cognitively penetrable then the function it computes is sensitive, in a semantically coherent way, to the organism's goals and beliefs, that is, it can be altered in a way that bears some logical relation to what the person knows (see also note 3). Note that changes produced by shaping basic sensors, say by attenuating or enhancing the output of certain feature detectors (perhaps through focal attention), do not count as cognitive penetration because they do not alter the contents of perceptions in a way that is logically connected to the contents of beliefs, expectations, values, and so on, regardless of how the latter are arrived at. Cognitive penetration is the rule in cognitive skills. For example, solving crossword puzzles, assigning the referent to pronouns and other anaphors in discourse, understanding today's newspaper, or attributing a cause to the noises outside your window are all cognitively penetrable functions. All you need to do is change what people believe (by telling them or showing them things) and you change what they do in these tasks in a way that makes sense in light of the content of the new information. Most psychological processes are cognitively penetrable, which is why behavior is so plastic and why it appears to be so highly stimulus-independent. That is why the claim that a significant portion of visual perception is cognitively impenetrable is a strong empirical claim.

The claims that we make in this paper may be summarized as follows.

1. Visual perception leads to changes in an organism's representations of the world being observed (or to changes in beliefs about what is perceived). Part of the process involves a uniquely visual system that we refer to as *early vision* (but see note 2). Many processes other than those of early vision, however, enter into the construction of visual representations of the perceived world.
2. The early vision system is a significant part of vision proper, in the sense to be discussed later (i.e., it involves the computation of most specifically visual properties, including 3-D shape descriptions).
3. The early vision system carries out complex computations, some of which have been studied in considerable detail. Many of these computations involve what is called top-down processing (e.g., some cases of perceptual "filling in" appear to be in this category – see Pessoa et al. 1998). What

this means is that the interpretation of parts of a stimulus may depend on the joint (or even prior) interpretation of other parts of the stimulus, resulting in global-to-local influences[6] such as those studied by Gestalt psychologists. Because of this, some local vision-specific memory may also be embodied in early vision.[7]

4. The early vision system is encapsulated from cognition, or to use the terms we prefer, it is cognitively impenetrable. Since vision as a whole *is* cognitively penetrable, this leaves open the question of where the cognitive penetration occurs.

5. Our hypothesis is that cognition intervenes in determining the nature of perception at only two loci. In other words, the influence of cognition upon vision is constrained in how and where it can operate. These two loci are:

(a) in the allocation of attention to certain locations or certain properties *prior to* the operation of early vision (the issue of allocation of attention will be discussed in sects. 4.3 and 6.4).

(b) in the decisions involved in recognizing and identifying patterns *after* the operation of early vision. Such a stage may (or in some cases must) access background knowledge as it pertains to the interpretation of a particular stimulus. (For example, in order to recognize someone as Ms Jones, you must not only compute a visual representation of that person, but you must also judge her to be the very person known as Ms Jones. The latter judgment may depend on anything you know about Ms Jones and her habits as well as her whereabouts and a lot of other things.)

Note that early vision is defined functionally. The neuroanatomical locus of early vision, as we understand the term in this paper, is not known with any precision. However, its functional (psychophysical) properties have been articulated with some degree of detail over the years, including a mapping of various substages involved in computing stereo, motion, size, and lightness constancies, as well as the role of attention and learning. As various people have pointed out (e.g., Blake 1995), such analysis is often a prerequisite to subsequent neuroanatomical mapping.

2. Some reasons for questioning the continuity thesis

In this section we briefly sketch some of the reasons why one might doubt the continuity between visual perception and cognition, despite the sort of evidence summarized above. Later we will return to some of the more difficult issues and more problematic evidence for what is sometimes called "knowledge-based" visual processing.

1. As Bruner himself noted (see note 5), perception appears to be rather resistant to rational cognitive influence. It is a remarkable fact about the perceptual illusions that knowing about them does not make them disappear: even after you have had a good look at the Ames room – perhaps even built it yourself – it still looks as though the person on one side is much bigger than the one the other side (Ittelson & Ames 1968). Knowing that you measured two lines to be exactly equal does not make them look equal when arrowheads are added to them to form the Müller-Lyer illusion, or when a background of converging perspective lines are added to form the Ponzo illusion, and so on. It is not only that the illusions are stubborn, in the way some people appear unwilling to change their minds in the face of contrary evidence, it is simply impossible to make some things look to you the way you know they really are. What is noteworthy is not that there are perceptual illusions, it is that in these cases there is a very clear separation between what you see and what you know is actually there – what you believe. What you believe depends on how knowledgeable you are, what other sources of information you have, what your utilities are (what is important to you at the moment), how motivated you are to figure out how you might have been misled, and so on. Yet how things look to you appears to be impervious to any such factors, even when what you know is both relevant to what you are looking at and at variance with how you see it.

2. There are many regularities within visual perception – some of them highly complex and subtle – that are automatic, depend only on the visual input, and often follow principles that appear to be orthogonal to the principles of rational reasoning. These principles of perception differ from the principles of inference in two ways.

First, perceptual principles, unlike the principles of inference, are responsive only to visually presented information. Although, like reasoning, the principles apply to representations, these representations are over a vocabulary different from that of beliefs and do not interact with them. The regularities are over a proprietary set of perceptual concepts that apply to basic perceptual labels rather than physical properties. That is why in computer vision a major part of early vision is concerned with what is called scene labeling or label propagation (Rosenfeld et al. 1976; Chakravarty 1979), wherein principles of label consistency are applied to represented features in a scene. The reason this is important is that the way you perceive some aspect of a display determines the way you perceive another aspect of it. When a percept of an ambiguous figure (like a line drawing of a polyhedron) reverses, a variety of properties (such as the perceived relative size and luminance of the faces) appear to automatically change together to maintain a coherent percept, even if it means a percept of an impossible 3-D object, as in Escher drawings. Such intra-visual regularities have been referred to by Rock (1997) and Epstein (1982) as perceptual coupling. Gogel (1973) has attempted to capture some of these regularities in what he calls perceptual equations. Such equations, though applied to cognitive representations, provide no role for what the perceiver knows or expects (though the form that these particular equations or couplings take may be understood in relation to the organism's needs and the nature of world it typically inhabits – see sect. 5.1).

Second, the principles of visual perception are different from those of inference in that in general they do not appear to conform to what might be thought of as tenets of "rationality" (see note 3 on the use of this term). Particularly revealing examples of the difference between the organizing principles of vision and the principles of inference are to be found in the phenomenon of "amodal completion." This phenomenon refers to the fact that partially occluded figures are not perceived as the fragments of figures that are actually in view, but as whole figures that are partially hidden from view behind the occluder (a distinction that is quite striking phenomenally). It is as though the visual system "completes" the missing part of the figure, and the completed portion, though it is constructed by the mind, has real perceptual consequences. Yet the form taken by an amodal completion (the shape that is "completed" or

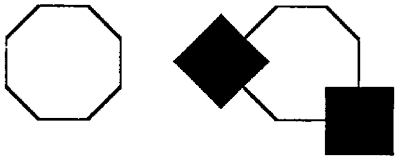


Figure 1. Kanizsa amodal completion figure. The completion preferred by the visual system is not the simplest figure despite the flanking examples. After Kanizsa (1985).

amodally perceived to be behind the occluder) follows complex principles of its own – which are generally not rational principles, such as semantic coherence or even something like maximum likelihood. As Kanizsa (1985) and Kanizsa and Gerbino (1982) have persuasively argued, these principles do not appear to reflect a tendency for the simplest description of the world and they are insensitive to knowledge, expectations, and even to the effects of learning (Kanizsa 1969). For example, Figure 1 shows a case of amodal completion in which the visual system constructs a complex and asymmetrical completed shape rather than the simple octagon, despite the presence of the adjacent examples of the latter.

3. There is a great deal of evidence from neuroscience that points to the partial independence of vision and other cortical functions. This evidence includes both functional-anatomical studies of visual pathways as well as observations of cases of visual pathologies that dissociate vision and cognition. For example, there are deficits in reasoning unaccompanied by deficits of visual function and there are cortical deficits in visual perception unaccompanied by deficits in cognitive capacity. These are discussed in sect. 3.3.

4. Finally, there are certain methodological questions that can be raised in connection with the interpretation of empirical evidence favoring the continuity thesis. In section 4 we will discuss some methodological arguments favoring the view that the observed effects of expectations, beliefs, and so on, although real enough, operate primarily on a stage of processing lying outside of what we have called early vision. The effect of knowledge can often be traced to a locus subsequent to the operation of vision proper – a stage where decisions are made as to the category of the stimulus or its function or its relation to past perceptual experiences. We will also suggest that, in a number of cases, such as in perceptual learning and in the effect of "hints" on the perception of certain ambiguous stimuli, the cognitive effect may be traced to a pre-perceptual stage where attention is allocated to different features or objects or places in a stimulus.

The rest of this article proceeds as follows. In order to place the discontinuity thesis in a historical context, section 3 will sketch some of the arguments for and against this thesis that have been presented by scholars in various areas of research. Section 4 will then discuss a number of methodological issues in the course of which we will attempt to draw some distinctions concerning stages of information processing involved in vision and relate these to empirical measures derived from signal detection theory and recordings of event-related potentials. Section 5 will examine other sorts of evidence that have been cited in support of the view that vision involves a knowledge-dependent intelligent process, and will present a discussion of some intrinsic constraints on the early visual system that make it appear as though what is seen depends on inferences from both general knowledge and from knowledge of the particular circumstances under which a scene is viewed. In section 6 we will examine how the visual system might be modulated by such things as "hints" as well as by experience and will outline the important role played by focused attention in shaping visual perception. Finally, the issue of the nature of the output of the visual system will be raised in section 7 where we will see that there is evidence that different outputs may be directed at different post-visual systems.

3. The view from computer vision, neuroscience, and clinical neurology

Within the fields of computer vision and neuroscience, both of which have a special interest in visual perception, there have also been swings in the popularity of the idea that vision is mediated by cognitive processes. Both fields entertained supporters and detractors of this view. In the following section we will sketch some of the positions taken on this issue. Showing how the positions developed and were defended in these sciences will help to set the stage for the arguments we shall make here for the impenetrability of early vision.

3.1. The perspective from artificial intelligence

In this section we offer a brief sketch of how the problem of vision has been studied within the field of artificial intelligence, or computer vision, where the goal has been to design systems that can "see" or exhibit visual capacities of some specified type. The approach of trying to design systems that can see well enough to identify objects or to navigate through an unknown environment using visual information has the virtue of setting a clear problem to be solved. In computer vision the goal is to design a system that is sufficient to the task of exhibiting properties we associate with visual perception. The sufficiency condition on a theory is an extremely useful constraint because it forces one to consider possible mechanisms that could accomplish certain parts of the task. Thus it behooves the vision researcher to consider the problems that computer vision designers have run into, as well as some of the proposed solutions that have been explored. And indeed, modern vision researchers have paid close attention to work on computer vision and vice versa. Consequently it is not too surprising that the history of computer vision closely parallels the history of ideas concerning human vision.

Apart from some reasonably successful early "model-based" vision systems capable of recognizing simple polyhedral objects when the scene was restricted to only such objects (Roberts 1965), most early approaches to computer vision were of the data-driven or so-called "bottom-up" variety. They took elementary optical features as their starting point and attempted to build more complex aggregates, leading eventually to the categorization of the pattern. Many of these hierarchical models were statistical pattern-recognition systems inspired by ideas from biology, including Rosenblatt's (1959) perceptron, Uttley's (1959) Conditional Probability Computer, and Selfridge's (1959) Pandemonium.

In the 1960s and 1970s a great deal of the research effort in computer vision went into the development of various "edge-finding" schemes in order to extract reliable features to use as a starting point for object recognition and scene analysis (Clowes 1971). Despite this effort, the edge finders were not nearly as successful as they needed to be if they were to serve as the primary inputs to subsequent analysis and identification stages. The problem was that if a uniform intensity-gradient threshold was used as a criterion for the existence of edges in the image, this would result in one of two undesirable situations. If the threshold were set low it would lead to the extraction of a large number of features that corresponded to shadows, lighting and reflectance variations, noise, or other differences unrelated to the existence of real edges in the scene. On the other hand, if the threshold were set higher, then many real scene edges that were clearly perceptible by human vision would be missed. This dilemma led to attempts to guide the edge finders into more promising image locations or to vary the edge threshold depending on whether an edge was more likely at those locations than at other places in the image.

The idea of guiding local edge-finding operators using knowledge of the scene domain may have marked the beginning of attempts to design what are known as knowledge-based vision systems. At MIT the slogan "heterarchy, not hierarchy" (Winston 1974) was coined to highlight the view that there had to be context-dependent influences from domain knowledge, in addition to local image features such as intensity discontinuities. Guided line finders were designed by Shirai (1975) and Kelly (1971) based on this approach. The idea that knowledge is needed at every level in order to recognize objects was strongly endorsed by Freuder (1986) in his proposal for a system that would use a great deal of specialized knowledge about certain objects (e.g., a hammer) in order to recognize these objects in a scene. Riseman and Hanson (1987) also take a strong position on this issue, claiming "it appears that human vision is fundamentally organized to exploit the use of contextual knowledge and expectations in the organization of visual primitives . . . thus the inclusion of knowledge-driven processes at some level in the image interpretation task, where there is still a great degree of ambiguity in the organization of the visual primitives, appears inevitable" (p. 286).

The knowledge-based approach is generally conceded to be essential for developing high performance computer vision systems using current technology. Indeed, virtually all currently successful automatic vision systems for robotics or such applications as analyzing medical images or automated manufacturing are model-based (e.g., Grimson 1990) – that is, their analysis of images is guided by some stored model. Although model-based systems may not use general knowledge and draw inferences, they fall into the knowledge-based category because they quite explicitly use knowledge about particular objects in deciding whether a scene contains instances of these objects. Even though in some cases they may use some form of "general purpose" model of objects (Lowe 1987; Zucker et al. 1975) – or even of parts of such objects (Biederman 1987) – the operation of the systems depends on prior knowledge of particulars. In addition, it is widely held that the larger the domain over which the vision system must operate, the less likely it is that a single type of stored information will allow reliable recognition. This is because in the general case, the incoming data are too voluminous, noisy, incomplete, and intrinsically ambiguous to allow univocal analysis. Consequently, so the argument goes, a computer vision system must make use of many different domain "experts," or sources of knowledge concerning various levels of organization and different aspects of the input domain, from knowledge of optics to knowledge of the most likely properties to be found in the particular domain being visually examined.

The knowledge-based approach has also been exploited in a variety of speech-recognition systems. For example, the SPEECHLIS and HWIM speech recognition systems developed at Bolt, Berenski, and Newman (BBN) (Woods 1978) are strongly knowledge based. Woods has argued for the generality of this approach and has suggested that it is equally appropriate in the case of vision. Two other speech recognition systems developed at Carnegie-Mellon University (HEARSAY described by Reddy 1975, and HARPY, described by Newell 1980) also use multiple sources of knowledge and introduced a general scheme for bringing knowledge to bear in the recognition process. Both speech recognition systems use a so-called "blackboard architecture" in which a common working memory is shared by a number of "expert" processes, each of which contributes a certain kind of knowledge to the perceptual analysis. Each knowledge source contributes "hypotheses" as to the correct identification of the speech signal, based on its area of expertise. Thus, for example, the acoustical expert, the phonetic expert, the syntactic expert, the semantic expert (which knows about the subject matter of the speech), and the pragmatic expert (which knows about discourse conventions) each propose the most likely interpretation of a certain fragment of the input signal. The final analysis is a matter of negotiation among these experts. What is important here is the assumption that the architecture permits any relevant source of knowledge to contribute to the recognition process at every stage. Many writers (Lindsey & Norman 1977; Rumelhart 1977) have adopted such a blackboard architecture in dealing with vision.

We shall argue later that one needs to distinguish between systems that access and use knowledge, such as those just mentioned, and systems that have constraints on interpretation built into them that reflect certain properties of the world. The latter embody an important form of visual intelligence that is perfectly compatible with the impenetrability thesis and will be discussed in section 5.1.

3.2. The perspective from neuroscience

The discovery of single-cell receptive fields and the hierarchy of simple, complex, and hypercomplex cells (Hubel & Weisel 1962) gave rise to the idea that perception involves a hierarchical process in which larger and more complex aggregates are constructed from more elementary features. In fact, the hierarchical organization of the early visual pathways sometimes encouraged an extreme hierarchical view of visual processing, in which the recognition of familiar objects by master cells was assumed to follow from a succession of categorizations by cells lower in the hierarchy. This idea seems to have been implicit in some neuroscience theorizing, even when it was not explicitly endorsed. Of course such an assumption is not warranted because any number of processes, including inference, could in fact intervene between the sensors and the high-level pattern-neurons.

There were some early attempts to show that some centripetal influences also occurred in the nervous system. For example, Hernandez-Péon et al. (1956) showed that the auditory response in a cat's cochlear nucleus was attenuated when the cat was attending to a visual stimulus. More recently, the notion of focal visual attention has begun to play a more important role in behavioral neuroscience theorizing and some evidence has been obtained showing that the activity of early parts of the visual system can indeed be influenced by selective attention (e.g., Haenny & Schiller 1988; Moran & Desimone 1985; Mountcastle et al. 1987; Sillito et al. 1994; Van Essen & Anderson 1990). Indeed, there is recent evidence that attention can have long term effects (Desimone 1996) as well as transitory ones. Some writers (e.g., Churchland 1988) have argued that the presence of centripetal nerve fibers running from higher cortical centers to the visual cortex constitutes prima facie evidence that vision must be susceptible to cognitive influences. However, the role of the centripetal fibers remains unclear except where it has been shown that they are concerned with the allocation of attention. What the evidence shows is that attention can selectively sensitize or gate certain regions of the visual field as well as certain stimulus properties. Even if such effects ultimately originate from "higher" centers, they constitute one of the forms of influence that we have admitted as being prior to the operation of early vision – that is, they constitute an early attentional selection of relevant properties (typically location, but see sect. 4.3 regarding other possible properties).

Where both neurophysiological and psychophysical data show top-down effects, they do so most clearly in cases where the modulating signal originates *within* the visual system itself (roughly identified with the visual cortex, as mapped out, say, by Felleman & Van Essen 1991; Felleman et al. 1997). There are two major forms of modulation, however, that appear to originate from outside the visual system. The first is one to which we have already alluded – modulation associated with focal attention, which can originate either from events in the world (exogenous control) or from cognitive sources (endogenous control). The second form of extra-visual effect is the modulation of certain cortical cells by signals originating in both visual and motor systems. A large proportion of the cells in posterior parietal cortex (and in what Ungerleider & Mishkin, 1982, identified as the dorsal stream of the visual or visuomotor system) are activated jointly by specific visual patterns, together with specific behaviors carried out (or anticipated) that are related to these visual patterns (see the extensive discussion in Milner & Goodale 1995, as well as the review in Lynch 1980). There is now a great deal of evidence suggesting that the dorsal system is tuned for what Milner and Goodale (1995) call "vision for action." What has not been reported, however, is comparable evidence to suggest that cells in any part of the visual system (and particularly the ventral stream that appears to be specialized for recognition) can be modulated in a similar way by higher level cognitive influences. Although there are cells that respond to such highly complex patterns as a face and some of these may even be viewpoint-independent (i.e., object-centered; Perrett et al. 1987), there is no evidence that such cells are modulated by nonvisual information about the identity of the face (e.g., whether it was the face expected in a certain situation). More general activation of the visual system by voluntary cognitive activity has been demonstrated by PET and fMRI studies (Kosslyn 1994), but no content-specific modulations of patterns of activity by cognition have been shown (i.e., there is no evidence for patterns of activity particular to certain interpretations of visual inputs), as they have been in the case of motor-system modulation.

It is not the visual complexity of the class to which the cell responds, nor whether the cell is modulated in a top-down manner that is at issue, but whether or not the cell responds to how a visual pattern is *interpreted,* where the latter depends on what the organism knows or expects. If vision were cognitively penetrable one might expect there to be cells that respond to certain interpretation-specific perceptions. In that case whether or not the cell responds to a certain visual pattern would appear to be governed by the cognitive system in a way that reflects how the pattern is conceptualized or understood. Studies of macaque monkeys by Perrett and his colleagues suggest that cells in the temporal cortex respond only to the *visual* character of the stimulus and not to its cognitively-determined (or conceptual) interpretation. For example, Perrett et al. (1990) describe cells that fire to the visual event of an experimenter "leaving the room" – and not to comparable movements that are not directed toward the door. Such cells clearly encode a complex class of events (perhaps involving the relational property "toward the door") which the authors refer to as a "goal centered" encoding. However they found no cells whose firing was modulated by what they call the "significance" of the event. The cells appear to fire equally no matter what the event means to the monkey. As Perrett et al. put it (p. 195):

> The particular significance of long-term disappearance of an experimenter . . . varies with the circumstances. Usually leaving is of no consequence, but sometimes leaving may provoke disappointment and isolation calls, other times it provokes threats. It would . . . appear that it is the visual event of leaving the laboratory that is important, rather than any emotional or behavioral response. In general, cells in the temporal cortex appear to code visual objects and events independent of emotional consequences and the resulting behavior.

Put in our terms, we would say that although what such cells encode may be complex, it is not sensitive to the cognitive context.

3.3. The perspective from clinical neurology: Evidence from visual agnosia

One intriguing source of evidence that vision can be separated from cognition comes from the study of pathologies of brain function that demonstrate dissociations among various functions involving vision and cognition. Even when, as frequently happens, no clear lesion can be identified, the pattern of deficits can provide evidence of certain dissociations and co-occurrence patterns of skill. They thus constitute at least initial evidence for the taxonomy of cognitive skills. The discovery that particular skill components can be dissociated from other skill components (particularly if there is evidence of double dissociation) provides a prima facie reason to believe that these subskills might constitute independent systems. Although evidence of dissociation of vision and cognition does not in itself provide direct support for the thesis that early vision is cognitively impenetrable, the fact that a certain aspect of the recognition and recall system *can* function when another aspect related to visual input fails, tends to suggest that the early computation of a visual percept may proceed independently of the process of inference and recognition under normal conditions. Of course in the absence of a detailed theory of the function of various brain areas, clinical evidence of dissociations of functions is *correlational* evidence, and like any correlational evidence it must await convergent confirmation from other independent sources of data.

Consider the example of visual agnosia, a rather rare family of visual dysfunctions in which a patient is often unable to recognize formerly familiar objects or patterns. In these cases (many of which are reviewed in Farah 1990) there is typically no impairment in sensory, intellectual, or naming abilities. A remarkable case of classical visual agnosia is described by Humphreys and Riddoch (1987). After suffering a stroke that resulted in bilateral damage to his occipital lobe, the patient was unable to recognize familiar objects, including faces of people well known to him (e.g., his wife), and found it difficult to discriminate among simple shapes, despite the fact that he did not exhibit any intellectual deficit. As is typical in visual agnosias, this patient showed no purely sensory deficits, showed normal eye movement patterns, and appeared to have close to normal stereoscopic depth and motion perception. Despite the severity of his visual impairment, the patient could do many other visual and object-recognition tasks. For example, even though he could not recognize an object in its entirety, he could recognize its features and could describe and even draw the object quite well – either when it was in view or from memory. Because he recognized the component features, he often could figure out what the object was by a process of deliberate problem solving, much as the continuity theory claims occurs in normal perception, except that for this patient it was a painstakingly slow process. From the fact that he could describe and copy objects from memory and could recognize objects quite well by touch, it appears that there was no deficit in his memory for shape. These deficits seem to point to a dissociation between the ability to recognize an object (from different sources of information) and the ability to compute an integrated pattern from visual inputs that can serve as the basis for recognition. As Humphreys and Riddoch (1987, p. 104) put it, this patient's pattern of deficits "supports the view that 'perceptual' and 'recognition' processes are separable, because his stored knowledge required for recognition is intact" and that inasmuch as recognition involves a process of somehow matching perceptual information against stored memories, then his case also "supports the view that the perceptual representation used in this matching process can be 'driven' solely by stimulus information, so that it is unaffected by contextual knowledge."

It appears that in this patient the earliest stages in perception – those involving computing contours and simple shape features – are spared. So also is the ability to look up shape information in memory in order to recognize objects. What then is damaged? It appears that an intermediate stage of "integration" of visual features fails to function as it should. While this pattern of dissociation does not provide evidence as to whether or not the missing integration process is cognitively penetrable, it does show that without this uniquely visual stage, the capacity to extract features together with the capacity to recognize objects from shape information is incapable of filling in enough to allow recognition. But "integration" according to the New Look (or Helmholtzian) view of perception, comes down to no more than making inferences from the basic shape features – a capacity that appears to be spared.

3.4. Summary

In the preceding we have reviewed a variety of evidence both for and against the thesis that vision is cognitively impenetrable. The bulk of this evidence suggests that the impenetrability thesis may well be correct. However we have left a great deal of the contrary evidence unexplained and have not raised some of the more subtle arguments for penetrability. After all, it is demonstrably the case that it is easier to "see" something that you are expecting than something that is totally unexpected and decontextualized. Moreover, it is also clear from many Gestalt demonstrations that how some part of a stimulus appears to an observer depends on a more global context – both spatially and temporally – and even illusions are not all one-sided in their support for the independence or impenetrability of vision: Some illusions show a remarkable degree of intelligence in how they resolve conflicting cues. Moreover, there is such a thing as perceptual learning and there are claims of perceptual enhancement by hints and instructions.

To address these issues we need to examine some additional arguments for distinguishing a cognitively impenetrable stage of vision from other stages. The first of these arguments is based on certain conceptual and methodological considerations. In the following section we will examine some of the information-processing stage proposals and some of the measures that have been used to attempt to operationalize them. We do so in order to provide a background for the conceptual distinctions that correspond to these stages as well as a critique of measures based on the signal-detection theory and event-related potential methodologies that have been widely used to test the penetrability of vision. Although important to the general dispute, the following section is necessarily more technical and can be omitted on first reading without loss of continuity.

4. Determining the locus of context effects: Some methodological issues

We have already suggested some problems in interpreting experimental evidence concerning the effects of cognitive context on perception. The problems arise because we need to distinguish among various components or stages in the process by which we come to know the world through visual perception. Experiments showing that with impoverished displays, sentences or printed words are more readily recognized than are random strings, do not in themselves tell us which stage of the process between stimulus and response is responsible for this effect. They do not, for example, tell us whether the effect occurs because the more meaningful materials are easier to see, because the cognitive system is able to supply the missing or corrupt fragments of the stimulus, because it is easier to figure out from fragmentary perceptual information what the stimulus must have been, or because it is easier to recall and report the contents of a display when it consists of more familiar and predictable patterns.

The existence of these alternative interpretations was recognized quite early (e.g., see Wallach 1949 and the historical review in Haber 1966) but the arguments had little influence at the time. Despite an interest in the notion of "preparatory set" (which refers to the observation that when subjects are prepared for certain properties of a stimulus they report those properties more reliably than other properties), there remained a question about when we ought to count an effect of set as occurring in the perceptual stage and when we should count it as occurring in a post-perceptual decision stage. This question was addressed in a comprehensive review by Haber (1966), who examined the literature from Kulpe's work at the turn of the century up to his own research on encoding strategies. Haber concluded that although a perceptual locus for set cannot be ruled out entirely, the data were more consistent with the hypothesis that set affects the strategies for mnemonic encoding, which then results in different memory organizations. This conclusion is consistent with recent studies (to be described later) by Hollingworth and Henderson (in press) who found no evidence that contextually induced set facilitates object perception once the sources of bias were removed from the experimental design.

Interest in this issue was rekindled in the past 30 or so years as the information processing view became the dominant approach in psychology. This led to the development of various techniques for distinguishing stages in information processing – techniques that we will mention briefly below.

4.1. Distinguishing perceptual and decision stages: Some methodological issues

Quite early in the study of sensory processes it was known that some aspects of perceptual activity involve decisions whereas others do not. Bruner himself even cites research using signal detection theory (SDT) (Swets 1998; Tanner & Swets 1954) in support of the conclusion that psychophysical functions involve decisions. What Bruner glossed over, however, is that the work on signal detection analysis not only shows that decisions are involved in threshold studies, it also shows that psychophysical tasks typically involve at least two stages, one of which, sometimes called "detection" or "stimulus evaluation," is immune from cognitive influences, while the other, sometimes called "response selection," is not. In principle, the theory provides a way to separate the two and to assign independent performance measures to them. To a first approximation, detection or stimulus evaluation is characterized by a sensitivity measure d' while response selection is characterized by a response bias or criterion measure β. Only the second of these measures was thought to capture the decision aspect of certain psychophysical tasks, and therefore it is the only part of the process that ought to be sensitive to knowledge and utilities (but see the discussion of this claim in sect. 4.2).

The idea of factoring information processing into a detection or stimulus evaluation stage and a response selection stage inspired a large number of experiments directed at "stage analysis" using a variety of methodologies in addition to signal detection theory, including the "additive factors method" (Sternberg 1969; 1998), the use of event-related potentials (ERPs), and other methods devised for specific situations. Numerous experiments have shown that certain kinds of cognitive malleability in visual recognition experiments are due primarily to the second of these stages, although other studies have implicated the stimulus evaluation stage as well. The problem, to which we will return below, is that the distinction between these stages is too coarse for our purposes, and its relation to visual perception continues to be elusive and in need of further clarification.

We begin our discussion of the separation of the perceptual process into distinct stages by considering the earliest psychophysical phenomenon to which signal detection theory was applied: the psychophysical threshold. The standard method for measuring the threshold of say, hearing, is to present tones of varying intensities and to observe the probability of the tone being correctly detected, with the intensity that yields 50% correct detection being designated as the threshold. But no matter how accurate an observer is, there is always some chance of missing a target or of "hearing" a tone when none is present. It is an assumption of SDT that the detection stage of the perceptual system introduces noise into the process, and that there is always some probability that a noise-alone event will be identical to some signal-plus-noise event. That being the case, no detection system is guaranteed to avoid making errors of commission (recognizing a signal when there is only noise) or errors of omission (failing to respond when a signal was present).

In applying SDT to psychophysical experiments, one recognizes that if subjects are acting in their best interest, they will adopt a response strategy that is sensitive to such things as the relative frequency or the prior probability of signal and noise, and on the consequences of different kinds of errors. Thus in deciding whether or not to respond, "signal" subjects must take into account various strategic considerations, including the "costs" of each type of error – that is, subjects must make decisions taking into account their utilities. If, for example, the perceived cost of an error of omission is higher than that of an error of commission, then the best strategy would be to adopt a bias in favor of responding positively. Of course, given some fixed level of sensitivity (i.e., of detector noise), this strategy will inevitably lead to an increase in the probability of errors of commission. It is possible, given certain assumptions, to take into account the observed frequency of both types of error in order to infer how sensitive the detector is, or, to put it differently, how much noise is added by the detector. This analysis leads to two independent parameters for describing performance, a sensitivity parameter d', which measures the distance between the means of the distribution of noise and of the signal-plus-noise (in standard units), and a response bias parameter, β which specifies the cutoff criterion along the distribution of noise and signal-plus-noise at which subjects respond that there was a "signal."

This example of the use of signal detection theory in the analysis of psychophysical threshold studies serves to introduce a set of considerations that puts a new perspective on many of the experiments of the 1950s and 1960s that are typically cited in support of the continuity view. What these considerations suggest is that although cognition does play an important role in how we describe a visual scene (perhaps even to ourselves), this role may be confined to a post-perceptual stage of processing. Although signal detection theory in its usual form is not always applicable (e.g., when there are several different responses or categories each with a different bias – see Broadbent 1967), the idea that there are at least two different sorts of processes going on has now become part of the background assumptions of the field. Because of this, a number of new methodologies have been developed over the past several decades to help distinguish different stages, and in particular to separate a decision stage from the rest of the total visual process. The results of this sort of analysis have been mixed. Many studies have located the locus of cognitive influence in the "response selection" stage of the process, but others have found the influence to encompass more than response selection. We shall return to this issue later. For the present we will describe a few of the results found in the literature to provide a background to our subsequent discussion.

One example is the simple phenomenon whereby the information content of a stimulus, which is a measure of its a priori probability, influences the time it takes to make a discriminative response to it. Many experiments have shown that if you increase the subjective likelihood or expectation of a particular stimulus you decrease the reaction time (RT) to it. In fact the RT appears to be a linear function of the information content of the stimulus, a generalization often called Hicks's Law. This phenomenon has been taken to suggest that expectations play a role in the recognition process, and therefore that knowledge affects perception. A variety of experimental studies on the factors affecting reaction time have been carried out using various kinds of stage analyses (some of which are summarized in Massaro 1988). These studies have suggested that such independent variables as frequency of occurrence, number of alternatives, predictability, and so on, have their primary effect on the response selection stage since, for example, the effect often disappears with overlearned responses, responses with high "stimulus-response compatibility" (such as reading a word or pressing the button immediately beside the stimulus light), or responses that otherwise minimize or eliminate the response-selecting decision aspect of the task. As an example of the latter, Longstreth et al. (1985) gave subjects the task of pressing a single response button upon the presentation of a digit selected from a specified set, and holding the button down for a time proportional to the value of the digit. In such a task, no effect of set size was observed – presumably because the decision could take place after the stimulus was recognized and a response initiated but while the response was in progress – so no response-selection time was involved in the reaction time measure.

Signal detection theory itself has frequently been used to assess whether context affects the stimulus evaluation or the response selection stage. Some of these studies have concluded that it is the response selection stage that is affected. For example, Farah (1989) reviewed a number of studies of priming and argued that priming by semantic relatedness and priming by perceptual features behave differently and that only the latter resulted in a d' effect. Since attention primed by meaning is unable to alter sensitivity, the data support the independence of pre-semantic visual processing.

Samuel (1981) used the SDT methodology directly to test the independence or impenetrability thesis as it applies to a remarkable perceptual illusion called the "phonemic restoration effect" in which observers "hear" a certain phoneme in a linguistic context where the signal has actually been removed and replaced by a short burst of noise. Although Samuel's study used auditory stimuli, it nonetheless serves to illustrate a methodological point and casts light on the perceptual process in general. Samuel investigated the question of whether the sentential context affected subjects' ability to discriminate the condition in which a phoneme had been replaced by a noise burst from one in which noise had merely been added to it. The idea is that if phonemes were actually being restored by the perceptual system based on the context (so that the decision stage received a reconstructed representation of the signal), subjects would have difficulty in making the judgment between noise-plus-signal and noise alone, and so the discrimination would show lower $d's$. In one experiment Samuel manipulated the predictability of the critical word and therefore of the replaced phoneme.

Subjects' task was to make two judgments: whether noise had been *added t*o the phoneme or whether it *replaced* the phoneme (added/replaced judgment); and which of two possible words shown on the screen was the one that they "heard." Word pairs (like "battle-batter") were embedded in predictable/unpredictable contexts (like "The soldier's/pitcher's thoughts of the dangerous battle/batter made him very nervous"). In each case the critical syllable of the target word (bat__) was either replaced by or had noise added to it. Samuel found no evidence that sentential context caused a decrement in d', as predicted by the perceptual restoration theory (in fact he found a surprising increase in d' that he attributed to prosodic differences between "replace" and "added" stimuli), but he did find significant β effects. Samuel concluded that although subjects reported predictable words to be intact more than unpredictable ones, the effect was due to response bias since discriminability was unimpaired by predictability.[8]

Other studies, on the other hand, seem to point to a context effect on d' as well as β. For example, Farah's (1989) analysis (mentioned above) led to a critical response by Rhodes and Tremewan (1993) and Rhodes et al. (1993) who provided data showing that both cross-modality priming of faces and semantic priming of the lexical decision task (which is assumed to implicate memory and possibly even reasoning) led to d' as well as β differences. Similarly, McClelland (1991) used a d' measure to argue in favor of a context effect on phoneme perception and Goldstone (1994) used it to show that discriminability of various dimensions could be changed after learning to categorize stimuli in which those dimensions were relevant. Even more relevant is a study by Biederman et al. (1982) which showed that sensitivity for detection of an object in a scene, as measured by d', was better when that object's presence in that scene was semantically coherent (e.g., it was harder to detect a toaster that was positioned in a street than one in a kitchen). This led Biederman et al. to conclude that meaningfulness was assessed rapidly and used early on to help recognize objects. We shall see later that these studies have been criticized on methodological grounds and that the criticism reveals some interesting properties of the SDT measures.

The distinction between a mechanism that changes the sensitivity, and hence the signal-to-noise ratio of the input, and one that shifts the acceptance criterion is an important one that we wish to retain, despite the apparently mixed results alluded to above. However, we need to reconsider the question of whether the stages that these measures pick out are the ones that are relevant to the issue of the cognitive penetrability of visual perception. There is reason to question whether d' and β measure precisely what we have in mind when we use the terms sensitivity and criterion bias to refer to different possible mechanisms for affecting visual perception. This is an issue to which we will return in section 4.2.

There are other methodologies for distinguishing stages that can help us to assess the locus of various effects. One widely used measure has been particularly promising because it does not require an overt response and therefore is assumed to be less subject to deliberate utility-dependent strategies. This measure consists of recording scalp potentials associated with the occurrence of specific stimuli, so-called "event-related potentials" or ERPs. A particular pattern of positive electrical activity over the centroparietal scalp occurs some 300 to 600 msecs after the presentation of certain types of stimuli. It is referred to as the P300 component of the ERP. A large number of studies have suggested that both the amplitude and the latency of the P300 measure vary with certain cognitive states evoked by the stimulus. These studies also show that P300 latencies are not affected by the same independent variables as is reaction time, which makes the P300 measure particularly valuable. As McCarthy and Donchin (1981) put it, "P300 can serve as a dependent variable for studies that require . . . a measure of mental timing uncontaminated by response selection and execution."

For example, Kutas et al. (1977) reported that the correlation between the P300 latency and reaction time was altered by subjects' response strategies (e.g., under a "speeded" strategy the correlations were low). McCarthy and Donchin (1981) also examined the effect on both RT and P300 latency of two different manipulations: discriminability of an embedded stimulus in a background, and stimulus-response (S-R) compatibility. The results showed that P300 latency was slowed when the patterns were embedded in a background within which they were harder to discriminate, but that the P300 latency was not affected by decreasing the S-R compatibility of the response (which, of course, did affect the reaction time). Other ERP studies found evidence that various manipulations that appear to make the perceptual task itself more difficult without changing the response load (or S-R compatibility) result in longer P300 latencies (Wickens et al. 1984). This research has generally been interpreted as showing that whereas both stimulus-evaluation and response-selection factors contribute to reaction time, P300 latency and/or amplitude is only affected by the stimulus evaluation stage. The question of exactly which properties of the triggering event cause an increase in the amplitude and latency of P300 is not without contention. There has been a general view that the amplitude of P300 is related to "expectancy." But this interpretation is itself in dispute and there is evidence that, at the very least, the story is much more complex since neither "surprise" by itself nor the predictability of a particular stimulus leads to an increase in P300 amplitude (see the review in Verleger 1988). The least controversial aspect of the ERP research has remained the claim that it picks out a stage that is independent of the response selection processes. This stage is usually identified as the stimulus evaluation stage. But in this context the stimulus evaluation stage means essentially everything that is not concerned with the preparation of an overt response, which is why a favorite method of isolating it has been to increase the stimulus-response compatibility of the pairing of stimuli and responses (thereby making the selection of the response as simple and overlearned as possible). When operationalized in this way, the stimulus evaluation stage includes such processes as the memory retrieval that is required for identification as well as any decisions or inferences not directly related to selecting a response.

This is a problem that is not specific to the ERP methodology, but runs through most of the stage analysis methods. It is fairly straightforward to separate a response-selection stage from the rest of the process, but that distinction is too coarse for our purposes if our concern is whether an intervention affects the visual process or the post-perceptual decision/inference/problem-solving process. For that purpose we need to make further distinctions within the stimulus evaluation stage so as to separate functions such as categorization and identification, which require accessing memory and making judgments, from functions that do not. Otherwise we should not be surprised to find that some apparently visual tasks are sensitive to what the observer knows, since the identification of a stimulus clearly requires both inferences and access to memory and knowledge.

4.2. Signal detection theory and cognitive penetration

We return now to an important technical question that was laid aside earlier: What is the relation between stages of perception and what we have been calling sensitivity? It has often been assumed that changes in sensitivity indicate changes to the basic perceptual process, whereas changes in criteria are a result of changes in the decision stage (although a number of people have recognized that the second of these assumptions need not hold, e.g., Farah 1989). Clearly a change in bias is compatible with the possibility that the perceptual stage generates a limited set of protohypotheses which are then subject to evaluation by cognitive factors at a post-perceptual stage, perhaps using some sort of activation or threshold mechanism. But in principle it is also possible for bias effects to operate at the perceptual stage by lowering or raising thresholds for features that serve as cues for the categories, thereby altering the bias for these categories. In other words, changes in β are neutral as to whether the effect is in the perceptual or post-perceptual stage; they merely suggest that the effect works by altering thresholds or activations or acceptance criteria. Because in principle they may be either thresholds for categories or for their cues, a change in β does not, by itself, tell us the locus of the effect.

The case is less clear for sensitivity measures such as d', as the argument among Farah (1989), Rhodes et al. (1993), and Norris (1995) shows. Norris (1995) argued that differences in d' cannot be taken to show that the perceptual process is being affected, because d' differences can be obtained when the effect is generated by a process that imposes a criterion shift alone. Similarly, Hollingworth and Henderson (in press) argued that applications of SDT can be misleading depending on what one assumes is the appropriate false alarm rate (and what one assumes this rate is sensitive to). These criticisms raise the question of what is really meant by sensitivity and what, exactly, we are fundamentally trying to distinguish in contrasting sensitivity and bias.

To understand why a change in d' can arise from criterion shifts in the process, we need to examine the notion of sensitivity itself. It is clear that every increase in what we think of as the readiness of a perceptual category is accompanied by some potential increase in a false alarm rate to some other stimulus category (though perhaps not to one that plays a role in a particular model or study). Suppose there is an increase in the readiness to respond to some category P, brought about by an increase in activation (or a decrease in threshold) of certain contributing cues for P. And suppose that category Q is also signaled by some of the same cues. Then this change in activation will also cause an increase in the readiness to respond to category Q. If Q is distinct from P, so that responding Q when presented with

P would count as an error, this would constitute an increase in the false alarm rate and therefore would technically correspond to a change in the response criterion.

But notice that if the investigator has no interest in Q, and Q is simply not considered to be an option in either the input or the output, then it would never be counted as a false alarm and so will not be taken into account in computing d' and β. Consequently, we would have changed d' by changing activations or thresholds. Although this counts as a change in sensitivity, it is patently an interest-relative or task-relative sensitivity measure.

Norris (1995) has worked out the details of how, in a purely bias model such as Morton's (1969) Logogen model, priming can affect the sensitivity for discriminating words from nonwords as measured by d' (either in a lexical decision task or a two-alternative forced-choice task). Using our terminology, what Norris's analysis shows is that even when instances of what we have called Q (the potential false alarms induced by priming) do not actually occur in the response set, they may nonetheless have associative or feature-overlap connections to other nonwords and so the priming manipulation can affect the word – nonword discrimination task. As Norris (p. 937) puts it, "As long as the nonword is more likely than the word to activate a word other than the target word, criterion bias models will produce effects of sensitivity as well as effects of bias in a simple lexical decision task." Norris attributes this to an incompatibility between the single-threshold assumption of SDT and the multiple-threshold assumption of criterion-bias models such as the Logogen model (Morton 1969). Although this is indeed true in his examples, which deal with the effects of priming on the lexical decision task, the basic underlying problem is that any form of activation or biasing has widespread potential false alarm consequences that may be undetected in some experimental paradigms but may have observable consequences in others.

Hollingworth and Henderson (in press) have argued that the Biederman et al. (1982) use of d' suffers from the incorrect choice of a false alarm rate. Recall that Biederman et al. used a d' measure to show that the semantic coherence of a scene enhances the sensitivity for detecting an appropriate object in that scene. In doing so they computed d' by using a false alarm rate that was pooled over coherent and incoherent conditions. This assumes that response bias is itself not dependent on whether the scene is coherent or not. Hollingworth and Henderson showed that subjects adopt a higher standard of evidence to accept that an inconsistent object was present in the scene than a consistent object – that is, that the response bias was itself a function of the primary manipulation. When they used a measure of false alarm rate relativized to each of the main conditions, they were able to show that semantic coherence affected only response bias and not sensitivity. By eliminating this response bias (as well as certain attentional biases), Hollingworth and Henderson were able to demonstrate convincingly that the semantic relationship between objects and the scene in which they were presented did not affect the detection of those objects.

The issue of selecting the appropriate false alarm rate is very general and one of the primary reasons why observed differences in d' can be misleading if interpreted as indicating that the mechanism responsible for the difference does not involve a criterion shift. Consider how this manifests itself in the case of the TRACE model of speech perception (McClelland & Elman 1986). What networks such as the TRACE interactive activation model do is increase their "sensitivity" for distinguishing the occurrence of a particular feature-based phonetic category F_i, from another phonetic category F_j in specified contexts. They do so because the weights in the network connections are such that they respond more readily to the combination of features described by the feature vector $\langle F_i; C_i \rangle$ and $\langle F_j; C_j \rangle$ than to feature vectors $\langle F_i; C_j \rangle$ and $\langle F_j; C_i \rangle$ (where for now the C's can be viewed as just some other feature vectors). This is straightforward for any activation-based system. But if we think of the F's as the phonemes being detected and the C's as some pattern of features that characterize the context, we can describe the system as increasing its sensitivity to F_i in context C_i and increasing its sensitivity to F_j in context C_j. Because the system only increases the probability of responding F_i over F_j in the appropriate context and responds equiprobably to them in other contexts, this leads mathematically to a d' rather than a β effect in this two-alternative situation.[9] But notice that this is because the false alarm rate is taken to be the rate of responding F_i in context C_j or to F_j in context C_i, which in this case will be low. Of course there will inevitably be some other F_k (for $k \neq i$), which shares some properties or features with F_i, to which the network will also respond more frequently in context C_i. In principle, there are arbitrarily many categories that share basic features with F_i, so such potential false alarms must increase. Yet if stimuli belonging to these categories fail to occur in either the input or output (e.g., if they are not part of the response set for one reason or another), then we will conclude that the mechanism in question increases d' without altering β.

As we have already noted, however, a conclusion based on such a measure is highly task-relative. Take, for example, the phonemic restoration effect discussed earlier in which sentences such as "The soldier's/pitcher's thoughts of the dangerous bat__ made him very nervous" are presented. If the "soldier" context leads to a more frequent report of an /el/ than an /er/ while the "pitcher" context does the opposite, this may result in a d' effect of context because the frequency of false alarms (reporting /er/ and /el/, respectively, in those same contexts) may not be increased. If, however, the perceptual system outputs an "en" (perhaps because /en/ shares phonetic features with /el/), this would technically constitute a false alarm, yet it would not be treated as such because no case of an actual /en/ is ever presented (e.g., the word "batten" is never presented and is not one of the response alternatives). (Even if it was heard it might not be reported if subjects believed that the stimuli consisted only of meaningful sentences.) A change in the activation level of a feature has the effect of changing the criteria of arbitrarily many categories into which that feature could enter, including ones that the investigator may have no interest in or may not have thought to test. Because each task systematically excludes potential false alarms from consideration, the measure of sensitivity is conditional on the task – in other words, it is task relative.

Notice that d' is simply a measure of discriminability. It measures the possibility of distinguishing some stimuli from certain specified alternatives. In the case of the original application of SDT the alternative to signal-plus-noise was noise alone. An increase in d' meant only one thing: the signal-to-noise ratio had increased (e.g., the noise added to the system by the sensor had decreased). This is a non-task-relative sense of increased sensitivity. Is there such a sense that is relevant for our purposes (i.e., for asking whether cognition can alter the sensitivity of perception to particular expected properties)? In the case of distinguishing noise from signal-plus-noise we feel that in certain cases there is at least one thing that could be done (other than altering the properties of the sensors) to increase the signal-to-noise ratio at the decision stage, and hence to increase d'. What we could do is filter the input so as to band-pass the real signals and attenuate the noise. If we could do that we would in effect have increased the signal-to-noise ratio to that particular set of signals. This brings us to a question that is central to our attempt to understand how cognition could influence vision: What type of mechanism can, in principle, lead to a non-task-relative increase in sensitivity or d' for a particular class of stimuli? – an increase in sensitivity in the strong sense.[10]

4.3. Sensitivity, filtering, and focal attention

One way to inquire what kind of mechanism can produce an increase in sensitivity (in the strong sense) to a particular class of stimuli is to consider whether there is some equivalent of "filtering" that operates early enough to influence the proto-hypothesis generation stage of visual perception. Such a process would have to increase the likelihood that the proto-hypotheses generated will include the class of stimuli to which the system is to be sensitized. There are two mechanisms that might be able to do this. One is the general property of the visual system (perhaps innate) that prevents it from generating certain logically possible proto-hypotheses. The visual system is in fact highly restricted with respect to the interpretations it can place on certain visual patterns, which is why it is able to render a unique 3-D percept when the proximal stimulus is inherently ambiguous. This issue of so-called "natural constraints" in vision is discussed in section 5.1.

The other mechanism is that of focal attention. In order to reduce the set of proto-hypotheses to those most likely to contain the hypothesis for which the system is "prepared," without increasing some false alarm rate, the visual system must be able to do the equivalent of "filtering out" some of the potential false alarm causing signals. As we remarked earlier, filtering is one of the few ways to increase the effective signal-to-noise ratio. But the notion of "filtering," as applied to visual attention, is a very misleading metaphor. We cannot "filter" just any properties we like, for the same reason that we cannot in general "directly pick up" any properties we like – such as affordances (see the discussion in Fodor & Pylyshyn 1981). All we can do in filtering is attempt to capitalize on some physically specifiable detectable property that is roughly coextensive with the class of stimuli to which the system is to be sensitized, and to use that property to distinguish those from other stimuli.

If the perceptual system responds to a certain range of property-values of an input (e.g., a certain region in a parameter space), then we need a mechanism that can be tuned to, or which can somehow be made to select a certain subregion of the parameter space. Unfortunately, regions in some parameter space do not in general specify the type of categories we are interested in – that is, categories to which the visual system is supposed to be sensitized, according to the cognitive penetrability view of vision. The latter are abstract or semantic categories such as particular words defined by their meaning – food, threatening creatures lurking in the dark, and so on – the sorts of things studied within the New Look research program.

A number of physical properties have been shown to serve as the basis for focused attention. For example it has been shown that people can focus their attention on various frequency bands in both the auditory domain (Dai et al. 1991) and in the spatial domain (Julesz & Papathomas 1984; Shulman & Wilson 1987), as well as on features defined by color (Friedman-Hill & Wolfe 1995; Green & Anderson 1956), shape (Egeth et al. 1984), motion (McCleod et al. 1991), and stereo disparity (Nakayama & Silverman 1986). The most generally relevant physical property, however, appears to be spatial location. If context can predict where in space relevant information will occur, then it can be used to direct attention (either through an eye movement or through the "covert" movement of attention) to that location, thus increasing the signal-to-noise ratio for signals falling in that region. Interestingly, spatially focused attention is also the property that has been most successfully studied in attention research. Many different experimental paradigms have demonstrated increased sensitivity to attended regions (or in many cases to attended visual objects irrespective of their locations). These have included several studies that use SDT to demonstrate a d' effect at attended loci (Bonnel et al. 1987; Downing 1988; Lupker & Massaro 1979; Muller & Findlay 1987; and Shaw 1984). Many other studies (not using SDT measures) have shown that attention to moving objects increases detection and/or recognition sensitivity at the locations of those objects (Kahneman & Treisman 1992; Pylyshyn & Storm 1988).

While spatially focused attention provides a measure of relevant selectivity, and can generally serve as an interface between vision and cognition (see sect. 6.4), it is more doubtful that we can meet the more general requirement for enhanced perceptual sensitivity – finding a modifiable physical parameter that results in an increased signal-to-noise ratio for the expected class of signals. This is what we would need, for example, to allow the context to set parameters so as to select a particular phoneme in the phonemic restoration effect. In this more general sense of sensitivity there is little hope that a true perceptual selection or sensitivity-varying mechanism will be found for cognitive categories.

5. Perceptual intelligence and natural constraints

We now return to the question of whether there are extrinsic or contextual effects in visual perception – other than those claimed as cases of cognitive penetration. People often speak of "top-down" processes in vision. By this they usually refer to the phenomenon whereby the interpretation of certain relatively local aspects of a display is sensitive to the interpretation of more global aspects of the display (global aspects are thought to be computed later or at a "higher" level in the process than local aspects, hence the influence is characterized as going from high-to-low or top-down). Typical of these are the Gestalt effects, in which the perception of some subfigure in a display is dependent on the patterns within which it is embedded. Examples of such dependence of the perception of a part on the perception of the whole are legion and are a special case of the internal regularities of perception alluded to earlier as item 2 in section 2.

In the previous section we suggested that many contextual effects in vision come about after the perceptual system has completed its task – that they have a post-perceptual locus. But not all cases of apparent top-down effects in perception are cases that can be explained in terms of post-perceptual processes. Such top-down effects are extremely common in vision, and we shall consider a number of examples in this section. In particular, we will consider examples that appear on the surface to be remarkably like cases of "inference." In these cases the visual system appears to "choose" one interpretation over other possible ones, and the choice appears remarkably "rational." The important question for us is whether these constitute cognitive penetration. We shall argue that they do not, for reasons that cast light on the subtlety, efficiency, and autonomy of the operation of visual processing.[11]

In what follows, we consider two related types of apparent "intelligence" on the part of the visual system. The first has seen some important recent progress, beginning with the seminal work of David Marr (Marr 1982). It concerns the way in which the visual system recovers the 3-D structure of scenes from mathematically insufficient proximal data. The second has a longer tradition; it consists in demonstrations of what Rock (1983) has called "problem-solving," wherein vision provides what appear to be intelligent interpretations of certain systematically ambiguous displays (but see Kanizsa 1985, for a different view concerning the use of what he calls a "ratiomorphic" vocabulary). We will conclude that these two forms of apparent intelligence have a similar etiology.

5.1. Natural constraints in vision

Historically, an important class of argument for the involvement of reasoning in vision comes from the fact that the mapping from a three-dimensional world to our two-dimensional retinas is many-to-one and therefore noninvertible. In general there are infinitely many 3-D stimuli corresponding to any 2-D image. Yet in almost all cases we attain a unique percept (usually in 3-D) for each 2-D image – though it is possible that other options might be computed and rejected in the process. The uniqueness of the percept (except for the case of reversing figures like the Necker cube) means that there is something else that must be entering into the process of inverting the mapping. Helmholtz, as well as most vision researchers in the 1950s through the 1970s, assumed that this was inference from knowledge of the world because the inversion was almost always correct (e.g., we see the veridical 3-D layout even from 2-D pictures). The one major exception to this view was that developed by J. J. Gibson, who argued that inference was not needed since vision consists in the "direct" pickup of relevant information from the optic array by a process more akin to "resonance" than to inference. (We will not discuss this approach here since considerable attention was devoted to it in Fodor & Pylyshyn 1981.)

Beginning with the work of David Marr (1982), however, a great deal of theoretical analysis has shown that there is another option for how the visual system can uniquely invert the 3-D-to-2-D mapping. All that is needed is that the computations carried out in early processing embody (without explicitly representing and drawing inferences from) certain very general constraints on the interpretations that it is allowed to make. These constraints need not guarantee the correct interpretation of all stimuli (the noninvertibility of the mapping ensures that this is not possible in general). All that is needed is that they produce the correct interpretation under specified conditions that frequently obtain in our kind of physical world. If we can find such generalized constraints, and if their deployment in visual processing is at least compatible with what is known about the nervous system, then we would be in a position to explain how the visual system solves this inversion problem without "unconscious inference."

A substantial inventory of such constraints (called "natural constraints" because they are typically stated as if they were assumptions about the physical world) has been proposed and studied (see, for example, Brown 1984; Marr 1982; Richards 1988; Ullman & Richards 1990). One of the earliest is Ullman's (1979) "rigidity" constraint, which has been used to explain the kinetic depth effect. In the kinetic depth effect, a set of randomly arranged moving points is perceived as lying on the surface of a rigid (though invisible) 3-D object. The requirement for this percept is primarily that the points move in a way that is compatible with this interpretation. The "structure from motion" principle states that if a set of points moves in a way that is consistent with the interpretation that they lie on the surface of a rigid body, then they will be so perceived. The conditions under which this principle can lead to a unique percept are spelled out, in part, in a uniqueness theorem (Ullman 1979). This theorem states that three or more (orthographic) 2-D views of four noncoplanar points that maintain fixed 3-D inter-point distances uniquely determine the 3-D spatial structure of those points. Hence if the display consists of a sequence of such views, the principle ensures a unique percept that, moreover, will be veridical if the scene does indeed consist of points on a rigid object. Since in our world all but a very small proportion of feature points in a scene do lie on the surface of rigid objects, this principle ensures that the perception of moving sets of feature points is more often veridical than not.[12] It also explains why we see structure from certain kinds of moving dot displays, as in the "kinetic depth effect" (Wallach & O'Connell 1953).

A Helmholtzian analysis would say that the visual system infers the structure of the points by hypothesizing that they lie in a rigid 3-D configuration, and then it verifies this hypothesis. By contrast, the natural constraint view says that the visual system is so constructed (perhaps through evolutionary pressures) that a rigid interpretation will be the one generated by early vision (independent of knowledge of the particular scene – indeed, despite knowledge to the contrary) whenever it is possible – that is, whenever such a representation of the 3-D environment is consistent with the proximal stimulus. This representation, rather than some other logically possible one, is generated simply because, given the input and the structure of the early vision system, it is the only one that the system could compute. The visual system does not need to access an explicit encoding of the constraint: it simply does what it is wired to do, which, as it happens, means that it works in accordance with the constraint discovered by the theorist. Because the early vision system evolved in our world, the representations it computes are generally (though not necessarily) veridical. For example, because in our world (as opposed to, perhaps, the world of a jellyfish) most moving features of interest do lie on the surface of rigid objects, the rigidity constraint and other related constraints will generally lead to veridical perception. Notice that there is a major difference between the "natural constraint" explanation and the inference explanation, even though they make the same predictions in this case. According to the Helmholtz position, if the observer had reason to believe that the points did not lie on the surface of a moving rigid object, then that hypothesis would not be entertained. But that is patently false: experiments on the kinetic depth effect are all carried out on a flat surface, such as a computer monitor or projection screen, which subjects know is flat; yet they continue to see the patterns moving in depth.

Another natural constraint is based on the assumption that matter is predominantly coherent and that most substances tend to be opaque. This leads to the principle that neighboring points tend to be on the surface of the same object, and points that move with a similar velocity also tend to be on the surface of the same object. Other constraints, closely related to the above, are important in stereopsis; these include the (not strictly valid) principle that for each point on one retina there is exactly one point on the other retina that arises from the same distal feature, and the principle that neighboring points will have similar disparity values (except in a vanishingly small proportion of the visual field). The second of these principles derives from the assumption that most surfaces vary gradually in depth.

An important principle in computer vision is the idea that computations in early vision embody, but do not explicitly represent, certain very general constraints that enable vision to derive representations that are often veridical in our kind of physical world. The notion of "our kind of world" includes properties of geometry and optics and includes the fact that in visual perception the world presents itself to an observer in certain ways (e.g., projected approximately at a single viewpoint). This basic insight has led to the development of further mathematical analyses and to a field of study known as "observer mechanics" (Bennett et al. 1989). Although there are different ways to state the constraints – for example, in terms of properties of the world or in terms of some world-independent mathematical principle such as "regularization" (Poggio et al. 1990) – the basic assumption remains that the visual system follows a set of intrinsic principles independent of general knowledge,[13] expectations, or needs. The principles express the built-in constraints on how proximal information may be used in recovering a representation of the distal scene. Such constraints are quite different from the Gestalt laws (such as proximity and common fate) because they do not apply to properties of the proximal stimulus, but to the way that such a stimulus is interpreted or used to construct a representation of the perceptual world. In addition, people like Marr who work in the natural constraint tradition often develop computational models that are sensitive to certain general neurophysiological constraints. For example, the processes tend to be based on "local support" – or data that come from spatially local regions of the image – and tend to use parallel computations, such as relaxation or label-propagation methods, rather than global or serial methods (e.g., Dawson & Pylyshyn 1988; Marr & Poggio 1979; Rosenfeld et al. 1976).

5.2. "Problem-solving" in vision

In addition to the types of cases examined above, there are other cases of what Rock (1983) calls "perceptual intelligence," which differ from the cases discussed above because they involve more than just the 2-D to 3-D mapping. These include the impressive cases that are reviewed in the book by Irvin Rock (1983) who makes a strong case that they involve a type of "problem solving." We argue that these cases represent the embodiment of the same general kind of implicit constraints within the visual system as those studied under the category of natural constraints, rather than the operation of reasoning and problem-solving. Like the natural constraints discussed earlier, these constraints frequently lead to veridical percepts, yet, as in the amodal completion examples discussed earlier (e.g., Fig. 1), they often also appear to be quixotic and generate percepts that are not rationally coherent. As with natural constraints, the principles are internal to the visual system and are neither sensitive to beliefs and knowledge about the particulars of a scene nor are themselves available to cognition.

Paradigm examples of "intelligent perception" cited by Rock (1983) are the perceptual constancies. We are all familiar with the fact that we tend to perceive the size, brightness, color, and so on, of objects in a way that appears to take into account the distance that objects are from us, the lighting conditions, and other such factors extrinsic to the retinal image of the object in question. This leads to such surprising phenomena as differently perceived sizes – and even shapes – of afterimages when viewed against backgrounds at different distances and orientations. In each case it is as if the visual system knew the laws of optics and of projective geometry and took these into account, along with retinal information from the object and from other visual cues as to distance, orientation, as well as the direction and type of lighting and so on. The way that the visual system takes these factors into account is remarkable. Consider the example of the perceived lightness (or whiteness) of a surface, as distinct from the perception of how brightly illuminated it is. Observers distinguish these two contributors of objective brightness of surfaces in various subtle ways. For example if one views a sheet of cardboard half of which is colored a darker shade of gray than the other, the difference in their whiteness is quite apparent. But if the sheet is folded so that the two portions are at appropriate angles to each other, the difference in whiteness can appear as a difference in the illumination caused by their different orientations relative to a common light source. In a series of ingenious experiments, Gilchrist (1977) showed that the perception of the degree of "lightness" of a surface patch (i.e., whether it is white, gray, or black) is greatly affected by the perceived distance and orientation of the surface in question, as well as the perceived illumination falling on the surface – where the latter were experimentally manipulated through a variety of cues such as occlusion or perspective.

Rock (1983) cites examples such as the above to argue that in computing constancies, vision "takes account of" a variety of factors in an intelligent way, as though it were following certain kinds of rules. In the case of lightness perception, the rules he suggests embody principles that include (Rock 1983, p. 279): "(1) that luminance differences are caused by reflectance-property differences or by illumination differences, (2) that illumination tends to be equal for nearby regions in a plane . . . and (3) that illumination is unequal for adjacent planes that are not parallel." Such principles are exactly the kind of principles that appear in computational theories based on "natural constraints."

They embody general geometrical and optical constraints, they are specific to vision, and they are fixed and independent of the particulars of a particular scene. Lightness constancy is a particularly good example to illustrate the similarities between cases that Rock calls "intelligent perception" and the natural constraint cases because there are at least fragments of a computational theory of lightness constancy (more recently these have been embedded within a theory of color constancy) based on natural constraints that are very similar to the principles quoted above (see, for example, Maloney & Wandell 1990; Ullman 1976).

Other examples are cited by Rock as showing that perception involves a type of "problem solving." We will examine a few of these examples in order to suggest that they too do not differ significantly from the natural constraints examples already discussed. The examples below are also drawn from the ingenious work of Irvin Rock and his collaborators, as described in Rock (1983).

A familiar phenomenon of early vision is the perception of motion in certain flicker displays – so-called apparent or phi motion. In these displays, when pairs of appropriately separated dots (or lights) are displayed in alternation, subjects see a single dot moving back and forth. The conditions under which apparent motion is perceived have been investigated thoroughly. From the perspective of the present concern, one finding stands out as being particularly interesting. One way of describing it is to say that if the visual system is provided with an alternative perceptible "reason" why the dots are alternatively appearing and disappearing (other than that it is one dot moving back and forth), then apparent motion is not seen. One such "reason" could be that an opaque object (such as a pendulum swinging in the dark) is moving in front of a pair of dots and is alternately occluding one and then the other. Experiments by Sigman and Rock (1974) show, for example, that if the alternation of dots is accompanied by the appearance of what is perceived to be an opaque form in front of the dot that has disappeared, apparent motion of the dots is not perceived (Fig. 2B). Interestingly, if the "covering" surface presented over the phi dots is perceived as a transparent surface, then the illusory phi motion persists (Fig. 2A). Moreover, whether or not a surface is perceived as opaque can be a subtle perceptual phenomenon since the phi motion can be blocked by a "virtual" or "illusory" surface as in Figure 3A, though not in the control Figure 3B.

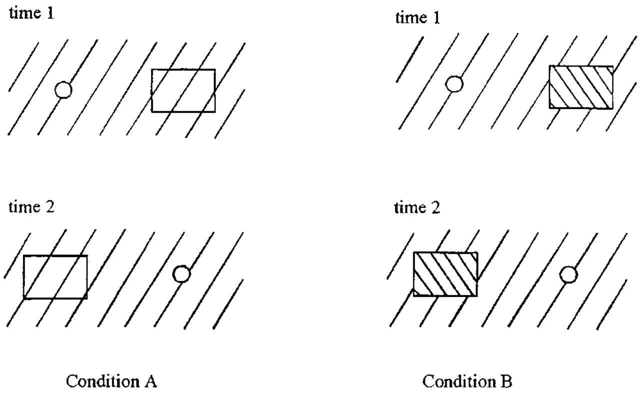


Figure 2. In the left figure (Condition A), the texture seen through the rectangle makes it appear to be an outline, so phi motion is perceived, whereas in the figure on the right (Condition B), the distinct texture on the rectangle makes it appear opaque, so phi motion is not perceived (after Sigman & Rock 1974).

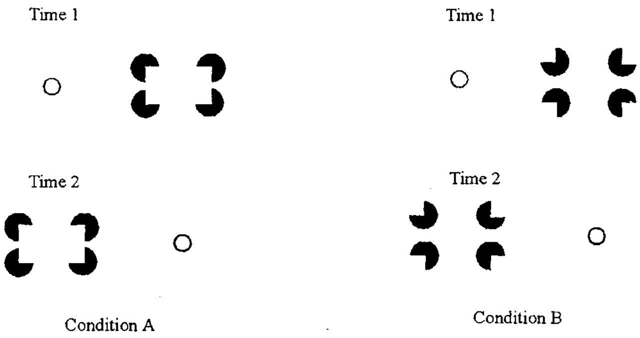


Figure 3. In the figure on the left (Condition A), the illusory rectangle appears to be opaque and to alternately cover the two dots so apparent motion is not perceived, whereas in the control display on the right (Condition B), apparent motion is perceived (after Sigman & Rock 1974).

There are many examples illustrating the point that the visual system often appears to resolve potential contradictions and inconsistencies in an intelligent manner. For example, the familiar illusory Kanizsa figure (such as the one shown in Fig. 3A) is usually perceived as a number of circles with an opaque (though implicit) figure in front of them which occludes parts of the circles. The figure is thus seen as both closer and opaque so that it hides segments of the circles. However, the very same stimulus will not be seen as an illusory figure if it is presented stereoscopically so that the figure is clearly seen as being in front of a textured background. In this case there is an inconsistency between seeing the figure as opaque and at the same time seeing the background texture behind it. But now if the texture is presented (stereoscopically again) so that it is seen to be in front of the Kanizsa figure, there is no longer a contradiction: subjects see the illusory opaque figure, as it is normally seen, but this time they see it through what appears to be a transparent textured surface (Rock & Anson 1979).

Such examples are taken to show that the way we perceive parts of a scene depends on making some sort of coherent sense of the whole. In the examples cited by Rock, the perception of ambiguous stimuli can usually be interpreted as resolving conflicts in a way that makes sense, which is why it is referred to as "intelligent." But what does "making sense" mean if not that our knowledge of the world is being brought to bear in determining the percept?

Embodying a natural constraint is different from drawing an inference from knowledge of the world (including knowledge of the particular constraint in question) in a number of ways. (1) A natural constraint that is embodied in early vision does not apply and is not available to any processes outside of the visual system (e.g., it does not in any way inform the cognitive system). Observers cannot tell you the principles that enable them to calculate constancies and lightness and shape from shading. Even if one takes the view that a natural constraint constitutes "implicit knowledge" not available to consciousness, it is still the case that this knowledge cannot in general be used to draw inferences about the world. nor can it be used in any way outside the visual system. (2) Early vision does not respond to any other kind of knowledge or new information related to these constraints (e.g., the constraints show up even if the observer knows that there are conditions in a certain scene that render the constraints invalid in that particular case). What this means is that no additional regularities are captured by the hypothesis that the system has knowledge of certain natural laws and takes them into account through "unconscious inference." Even though in these examples the visual process appears to be intelligent, the intelligence is compatible with it being carried out by neural circuitry that does not manipulate encoded knowledge. Terms such as "knowledge," "belief," "goal," and "inference" give us an explanatory advantage when it allows generalizations to be captured under common principles such as rationality or even something roughly like semantic coherence (Pylyshyn 1984). In the absence of such overarching principles, Occam's Razor or Lloyd Morgan's Canon dictates that the simpler or lower-level hypothesis (and the less powerful mechanism) is preferred. This is also the argument advanced by Kanizsa (1985) and explicitly endorsed by Rock (1983, p. 338).

Finally it should be pointed out that the natural constraints involved in examples of intelligent perception are of a rather specific sort that might reasonably be expected to be wired into the visual system because of their generality and evolutionary utility. The constraints invariably concern universal properties of space and light, augmented by certain simplifying assumptions generally true in our world. Theories developed in the natural constraint tradition are based almost entirely on constraints that derive from principles of optics and projective geometry. Properties such as the occlusion of features by surfaces closer to the viewer are among the most prominent in these principles, as are visual principles that are attributable to reflectance, opacity, and rigidity of bodies. What is perhaps surprising is that other properties of our world – about which our intuitions are equally strong – do not appear to share this special status in the early vision system. In particular the resolution of perceptual conflicts by such physical principles that solid objects do not pass through one another rarely occurs, with the consequence that some percepts constructed by the visual system fail a simple test of rationality or of coherence with certain basic facts about the world known to every observer.

Take the example of the Ames trapezoidal window which, when rotated, appears to oscillate rather than rotate through a full circle. When a rigid rod is placed inside this window at right angles to the frame, and the window-and-rod combination is rotated, an anomalous percept appears (described by Rock 1983, p. 319). The trapezoidal window continues to be perceived as oscillating while the rod is seen to rotate – thereby requiring that the rod be seen to pass through the rigid frame. Another example of this phenomenon is the Pulfrich double pendulum illusion (Wilson & Robinson 1986). In this illusion two solid pendulums constructed from sand-filled detergent bottles and suspended by rigid metal rods swing in opposite phase, one slightly behind the other. When viewed with a neutral density filter over one eye (which results in slower visual processing in that eye) one pendulum is seen as swinging in an ellipse while the other one is seen as following it around, also in an ellipse with the rigid rods passing through one another. From a certain angle of view the bottles also appear to pass through one another even though they appear to be solid and opaque (Leslie 1988). Interpenetrability of solid opaque objects does not seem to be blocked by the visual system.

6. Other ways that knowledge has been thought to affect perception

6.1. Experience and "hints" in perceiving ambiguous figures and stereograms

So far we have suggested that many cases of apparent penetration of visual perception by cognition are either cases of intra-system constraints, or are cases in which knowledge and utilities are brought to bear at a post-perceptual stage – after the independent perceptual system has done its work. But there are some alleged cases of penetration that, at least on the face of it, do not seem to fall into either of these categories. One is the apparent effect of hints, instructions, and other knowledge-contexts on the ability to resolve certain ambiguities or to achieve a stable percept in certain difficult-to-perceive stimuli. A number of such cases have been reported, though these have generally been based on informal observations rather than on controlled experiments. Examples include the so-called "fragmented figures," ambiguous figures, and stereograms. We will argue that these apparent counterexamples, though they may sometimes be phenomenally persuasive (and indeed have persuaded many vision researchers), are not sustained by careful experimental scrutiny.

For example, the claim that providing "hints" can improve one's ability to recognize a fragmented figure such as those shown in Figure 4 (and other so-called "closure" figures such as those devised by Street 1931) has been tested by Reynolds (1985). Reynolds found that providing instructions that a meaningful object exists in the figure greatly improved recognition time and accuracy (in fact, when subjects were not told that the figure could be perceptually integrated to reveal a meaningful object, only 9% saw such an object). On the other hand, telling subjects the class of object increased the likelihood of recognition but did not decrease the time it took to do so (which in this study took around 4 sec – much longer than any picture-recognition time, but much shorter than other reported times to recognize other fragmented figures, where times on the order of minutes are often observed). The importance of expecting a meaningful figure is quite general and parallels the finding that knowing that a figure is reversible or ambiguous is important for arriving at alternative percepts (perhaps even necessary, as suggested by Rock & Anson 1979; Girgus et al. 1977). But this is not an example in which knowledge acquired through hints affects the content of what is seen – which is what cognitive penetration requires.

Although verbal hints may have little effect on recogniz-

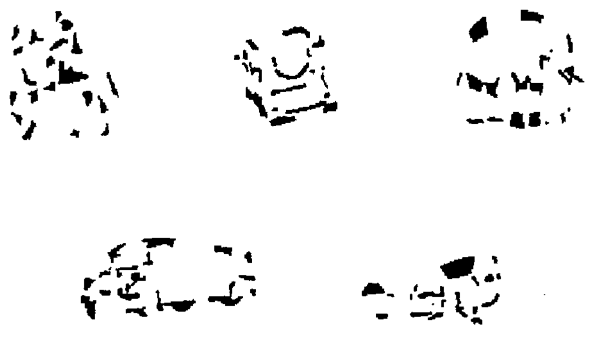


Figure 4. Examples of fragmented or "closure" figures, taken from Street's (1931) *A gestalt completion test.*

ing fragmented figures, other kinds of information can be beneficial. For example, Snodgrass and Feenan (1990) found that perception of fragmented pictures is enhanced by priming with moderately complete versions of the figures. Neither complete figures nor totally fragmented figures primed as well as an intermediate level of fragmentation of the same pictures. Thus it appears that in this case, as in many other cases of perceiving problematic stimuli (such as the random-dot stereograms and ambiguous figures described below), presenting collateral information within the visual modality can influence perception.

Notice that the fragmented figure examples constitute a rather special case of visual perception, insofar as they present the subject with a problem-solving or search task. Subjects are asked to provide a report under conditions where they would ordinarily not see anything meaningful. Knowing that the figure contains a familiar object results in a search for cues. As fragments of familiar objects are found, the visual system can be directed to the relevant parts of the display, leading to a percept. That a search is involved is also suggested by the long response latencies compared with the very rapid speed of normal vision (on the order of tenths of a second, when response time is eliminated; see Potter 1975).

What may be going on in the time it takes to reach perceptual closure in these figures may be nothing more than the search for a locus at which to apply the independent visual process. This search, rather than the perceptual process itself, may then be the process that is sensitive to collateral information. This is an important form of intervention from our perspective because it represents what is really a preperceptual stage during which the visual system is indeed directed, though not in terms of the content of the percept, but in terms of the location at which the independent visual process is applied. In section 6.4 we will argue that one of the very few ways that cognition can affect visual perception is by directing the visual system to focus attention at particular (and perhaps multiple) places in the scene.

A very similar story applies in the case of other ambiguous displays. When only one percept is attained and subjects know there is another one, they can engage in a search for other organizations by directing their attention to other parts of the display (hence the importance of the knowledge that the figure is ambiguous). It has sometimes been claimed that we can will ourselves to see one or another of the ambiguous percepts. For example, Churchland (1988) claims (contrary to controlled evidence obtained with naive subjects) to be able to make ambiguous figures "flip back and forth at will between the two or more alternatives, by changing one's assumptions about the nature of the object or about the conditions of viewing." But, as we have already suggested, there is a simple mechanism available for some degree of manipulation of such phenomena as figure reversal – the mechanism of spatially focused attention. It has been shown (Kawabata 1986; Peterson & Gibson 1991) that the locus of attention is important in determining how one perceives ambiguous or reversing figures such as the Necker cube. Gale and Findlay (1983) and Stark and Ellis (1981) showed that eye movements also determined which version of the well-known mother/daughter bistable figure (introduced by Boring 1930) was perceived. Magnuson (1970) showed that reversals occurred even with afterimages, suggesting that covert attentional shifts, without eye movements, could result in reversals. It has been observed that a bias present in the interpretation of a local attended part of the figure determines the interpretation placed on the remainder of the figure. Since some parts have a bias toward one interpretation and other parts have a bias toward another interpretation, changing the locus of attention can change the interpretation. In fact the parts on which one focuses generally have a tendency to be perceived as being in front – and also brighter. Apart from changes brought about by shifting attention, however, there is no evidence that voluntarily "changing one's assumptions about the object" has any direct effect on how one perceives the figure.

There are other cases in which it has sometimes been suggested that hints and prior knowledge affect perception. For example, the fusion of "random dot stereograms" (Julesz 1971) is often quite difficult and was widely thought to be improved by prior information about the nature of the object (the same is true of the popular autostereogram 3-D posters). There is evidence, however, that merely telling a subject what the object is or what it looks like does not make a significant difference. In fact, Frisby and Clatworthy (1975) showed that neither telling subjects what they "ought to see" nor showing them a 3-D model of the object provided any significant benefit in fusing random-dot stereograms. What does help, especially in the case of large-disparity stereograms, is the presence of prominent monocular contours, even when they do not themselves provide cues as to the identity of the object. Saye and Frisby (1975) argued that these cues help facilitate the required vergence eye movements and that in fact the difficulty in fusing random-dot stereograms in general is due to the absence of features needed for guiding the vergence movements that fuse the display. One might surmise that it may also be the case that directing focal attention on certain features (thereby making them perceptually prominent) can help facilitate eye movements in the same way. In that case learning to fuse stereograms, like learning to see different views of ambiguous figures, may be mediated by learning where to focus attention.

The main thing that makes a difference in the ease of fusing a stereogram is having seen the stereogram before. Frisby and Clatworthy (1975) found that repeated presentation had a beneficial effect that lasted for at least three weeks. In fact the learning in this case, as in the case of improvements of texture segregation with practice (Karni & Sagi 1995), is extremely sensitive to the retinotopic details of the visual displays – so much so that experience generalizes poorly to the same figures presented in a different retinal location (Ramachandran 1976) or a different orientation (Ramachandran & Braddick 1973). An important determiner of stereo fusion (and even more so in the case of the popular autostereogram posters) is attaining the proper vergence of the eyes. Such a skill depends on finding the appropriate visual features to drive the vergence mechanism. It also involves a motor skill that one can learn; indeed many vision scientists have learned to free-fuse stereo images without the benefit of stereo goggles. This suggests that the improvement with exposure to a particular stereo image may simply be learning where to focus attention and how to control eye vergence. None of these cases shows that knowledge itself can penetrate the content of percepts.

6.2. The case of expert perceivers

Another apparent case of penetration of vision by knowledge is in the training of the visual ability of expert perceivers – people who are able to notice patterns that novices cannot see (bird watchers, art authenticators, radiologists, aerial-photo interpreters, sports analysts, chess masters, and so on). Not much is known about such perceptual expertise since such skills are highly deft, rapid, and unconscious. When asked how they do it, experts typically say that they can tell that a stimulus has certain properties by "just looking." But the research that is available shows that often what the expert has learned is not a "way of seeing" as such, but rather some combination of a task-relevant mnemonic skill with a knowledge of where to direct attention. These two skills are reminiscent of Haber's (1966) conclusion that preparatory set operates primarily through mnemonic encoding strategies that lead to different memory organizations.

The first type of skill is illustrated by the work of Chase and Simon (1973), who showed that what appears to be chess masters' rapid visual processing and better visual memory for chess boards only manifests itself when the board consists of familiar chess positions and not at all when it is a random pattern of the same pieces (beginners, of course, do equally poorly on both). Chase and Simon interpret this to show that rather than having learned to see the board differently, chess masters have developed a very large repertoire (they call it a vocabulary) of patterns that they can use to classify or encode real chess positions (but not random positions). Thus what is special about experts' vision in this case is the system of classification that they have learned which allows them to recognize and encode a large number of relevant patterns. But, as we argued earlier, such a classification process is post-perceptual insofar as it involves decisions requiring accessing long-term memory.

The second type of skill, the skill to direct attention in a task-relevant manner, is documented in what is perhaps the largest body of research on expert perception: the study of performance in sports. It is obvious that fast perception, as well as quick reaction, is required for high levels of sports skill. Despite this truism, very little evidence of generally faster information processing capabilities has been found among experts (e.g., Abernethy et al. 1994; Starkes et al. 1994). In most cases the difference between novices and experts is confined to the specific domains in which the experts excel – and there it is usually attributable to the ability to anticipate relevant events. Such anticipation is based, for example, on observing initial segments of the motion of a ball or puck or the opponent's gestures (Abernethy 1991; Proteau 1992). Except for a finding of generally better attention-orienting abilities (Castiello & Umiltà 1992; Greenfield et al. 1994; Nougier et al. 1989), visual expertise in sports, like the expertise found in the Chase and Simon studies of chess skill, appears to be based on nonvisual expertise related to the learned skills of identifying, predicting, and attending to the most relevant places.

An expert's perceptual skill frequently differs from a beginner's in that the expert has learned where the critical distinguishing information is located within the stimulus pattern. In that case the expert can direct focal attention to the critical locations, allowing the independent visual process to do the rest. The most remarkable case of such expertise was investigated by Biederman and Shiffrar (1987) and involves expert "chicken sexers." Determining the sex of day-old chicks is both economically important and also apparently very difficult. In fact, it is so difficult that it takes years of training (consisting of repeated trials) to become one of the rare experts. By carefully studying the experts, Beiderman and Shiffrar found that what distinguished good sexers from poor ones is, roughly, where they look and what distinctive features they look for. Although the experts were not aware of it, what they had learned was the set of contrasting features and, even more important, where exactly the distinguishing information was located. This was demonstrated by showing that telling novices where the relevant information was located allowed them to quickly become experts themselves. What the "telling" does – and what the experts had tacitly learned – is how to bring the independent visual system to bear at the right spatial location, and what types of patterns to encode into memory, both of which are functions lying outside the visual system itself.

Note that this is exactly the way that we suggested that hints work in the case of the fragmented or ambiguous figures or binocular fusion cases. In all these cases the mechanism of spatially focused attention plays a central role. We believe that this role is in fact quite ubiquitous and can help us understand a large number of phenomena involving cognitive influences on visual perception (see sect. 6.4).

6.3. Does perceptual learning demonstrate cognitive penetration?

There is a large literature on what is known as "perceptual learning," much of it associated with the work of Eleanor Gibson and her students (Gibson 1991). The findings show that, in some general sense, the way people apprehend the visual world can be altered through experience. Recently there have been a number of studies on the effect of experience on certain psychophysical skills that are thought to be realized by the early vision system. For example, Karni and Sagi (1995) showed that texture discrimination could be improved with practice and that the improvement was long lasting. However, they also showed that the learning was specific to a particular retinal locus (and to the training eye) and hence was most probably due to local changes within primary visual cortex, rather than to a cognitively mediated enhancement. The same kind of specificity is true in learning improved spatial acuity (Fahle et al. 1995). The phenomenon referred to as "pop out" serves as another good example of a task that is carried out by early vision (in fact it can be carried out in monkeys even when their secondary visual area [V2] is lesioned; see Merigan et al. 1992). In this task the detection of certain features (e.g., a short slanted bar) in a background of distinct features (e.g., vertical bars) is fast, accurate, and insensitive to the number of distractors. Ahissar and Hochstein (1995) studied the improvement of this basic skill with practice. As in the case of texture discrimination and spatial acuity, they found that the skill could be improved with practice and that the improvement generalized poorly outside the original retinotopic position, size, and orientation. Like Fahle et al. they attributed the improvement to neuronal plasticity in the primary visual area. But Ahissar and Hochstein also found that there was improvement only when the targets were attended; even a large number of trials of passive presentation of the stimulus pattern produced little improvement in the detection task. This confirms the important role played by focal attention in producing changes, even in the early vision system, and it underscores our claim that early attentional filtering of visual information is a primary locus of cognitive intervention in vision.

Perceptual learning has also been linked to the construction of basic visual features. For example, the way people categorize objects and properties – and even the discriminability of features – can be altered through prior experience with the objects. Goldstone recently conducted a series of studies (Goldstone 1994; 1995) in which he showed that the discriminability of stimulus properties is altered by pre-exposure to different categorization tasks. Schyns et al. (1998) have argued that categorization does not rely on a fixed vocabulary of features but that feature-like properties are "created under the influence of higher-level cognitive processes . . . when new categories need to be learned."

This work is interesting and relevant to the general question of how experience can influence categorization and discrimination. The claim that a fixed repertoire of features at the level of cognitive codes is inadequate for categorization in general is undoubtedly correct (for more on this see Fodor 1998). However, none of these results is in conflict with the independence or impenetrability thesis as we have been developing it here because the tuning of basic sensory sensitivity by task-specific repetition is not the same as cognitive penetration as we understand the term (see, e.g., sect. 1.1). The present position is agnostic on the question of whether feature-detectors can be shaped by experience, although we believe that it is misleading to claim that they are "created under the influence of higher-level cognitive processes" since the role of the higher-level processes in the studies in question (learning how to categorize the stimuli) might plausibly have been limited to directing attention to the most relevant stimulus properties. As we saw above, in discussing the work of Ahissar and Hochstein, such attention is important for making changes in the early vision system.

6.4. Focal attention as an interface between vision and cognition

One of the features of perceptual learning that we have already noted is that learning allows attention to be spatially focused on the most relevant parts of the visual field. Spatial focusing of attention is perhaps the most important mechanism by which the visual system adjusts rapidly to an informationally dense and variable world. It thus represents the main interface between cognition and vision – an idea that has been noted in the past (e.g., Julesz 1990; Pylyshyn 1989). In recent years it has become clear that focal attention not only selects a subset of the available visual information, but it is also essential for perceptual learning (see sect. 6.3) and for the encoding of combinations of features (this is the "attention as glue" hypothesis of Treisman 1988; see also Ashby et al. 1996).

In addition to the single spatial locus of enhanced processing that has generally been referred to as focal attention (discussed in sect. 4.3 in terms of its filtering properties), our own experimental research has also identified an earlier stage in the visual system in which several distinct objects can be indexed or tagged for access (by a mechanism we have called FINSTs). We have shown (Pylyshyn 1989; 1994) that several spatially disparate objects in a visual field can be preattentively indexed, providing the visual system with direct access to these objects for further visual analysis. To the extent that the assignment of these indexes can itself be directed by cognitive factors, this mechanism provides a way for cognition to influence the outcome of visual processing by pre-selecting a set of salient objects or places to serve as the primary input to the visual system. Burkell and Pylyshyn (1997) have shown that several such indexes can serve to select items in parallel, despite interspersed distractors. Such a mechanism would thus seem to be relevant for guiding such tasks as searching for perceptual closure in fragmented figures (such as in Fig. 4), because that process requires finding a pattern across multiple fragments.

It has also been proposed that attention might be directed to other properties besides spatial loci, and thereby contribute to learned visual expertise. If we could learn to attend to certain relevant aspects of a stimulus we could thereby "see" things that others, who have not had the benefit of the learning, could not see – such as whether a painting is a genuine Rembrandt. As we have argued in section 4.3, unless we restrict what we attribute to such focusing of attention, we risk having a circular explanation. If attention is to serve as a mechanism for altering basic perceptual processing, as opposed to selecting and drawing inferences from the output of the perceptual system, it must respond to physically specifiable properties of the stimulus. As we have already suggested, the primary physical property is spatial location, although there is some evidence that under certain conditions features such as color, spatial frequency, simple form, motion, and properties recognizable by template-matching can serve as the basis for such pre-selection. Being a Rembrandt, however, cannot serve as such a pre-selection criterion in visual perception even though it may itself rely on certain kinds of attention-focusing properties that are physically specifiable.

On the other hand, if we view attention as being at least in part a post-perceptual process, so that it ranges over the outputs of the visual system, then there is room for much more complex forms of "perceptual learning," including learning to recognize paintings as genuine Rembrandts, learning to identify tumors in medical X-rays, and so on. But in that case the learning is not strictly in the visual system, but rather involves post-perceptual decision processes based on knowledge and experience, however tacit and unconscious these may be.

As a final remark it might be noted that even a post-perceptual decision process can, with time and repetition, become automatized and cognitively impenetrable, and therefore indistinguishable from the encapsulated visual system. Such automatization creates what I have elsewhere (Pylyshyn 1984) referred to as "compiled transducers." Compiling complex new transducers is a process by which post-perceptual processing can become part of perception. If the resulting process is cognitively impenetrable – and therefore systematically loses the ability to access long-term memory – then, according to the view being advocated in this paper, it becomes part of the visual system. Thus, according to the discontinuity theory, it is not unreasonable for complex processes to become part of the independent visual system over time. How such processes become "compiled" into the visual system remains unknown, although according to Newell's (1990) levels-taxonomy the process of altering the visual system would require at least one order of magnitude longer than basic cognitive operations themselves – and very likely it would require repeated experience, as is the case with most of the perceptual learning phenomena.

7. What is the input and output of the visual system?

7.1. A note about the input to early vision

We have sometimes been speaking as though the input to the early vision system is the activation of the rods and cones of the retina. But because we define early vision functionally, the exact specification of what constitutes the input to the early vision system must be left open to empirical investigation. For example, not everything that impinges on the retina counts as input to early vision. We consider attentional gating to precede early vision so in that sense early vision is post-selection. Moreover we know that the early vision system does receive inputs from other sources besides the retina. The nature of the percept depends in many cases on inputs from other modalities. For example, inputs from vestibular system appear to affect the perception of orientation (Howard 1982), and proprioceptive and efferent signals from the eye and head can affect perception of visual location. These findings suggest that certain kinds of information (primarily information about space) may have an effect across the usual modalities and therefore for certain purposes nonretinal spatial information may have to be included among inputs to the early vision system. (I suppose if it turns out that other modalities have unrestricted ability to determine the content of the percept we might want to change its name to "early spatial system," though so far I see little reason to suppose that this will happen.) For present purposes we take the attentionally modulated activity of the eyes to be the unmarked case of input to the visual system.

7.2. Categories and surface layouts

One of the important questions we have not yet raised concerns the nature of the output of the visual system. This is a central issue because the entire point of the independence thesis is to claim that early vision, understood as that part of the mind/brain that is unique to processes originating primarily with optical inputs to the eyes, is both independent and complex – beyond being merely the output of transducers or feature detectors. And indeed, the examples we have been citing all suggest that the visual system so defined does indeed deliver a rather complex representation of the world to the cognizing mind.

For Bruner (1957), the output of the visual system consists of categories, or at least of perceptions expressed in terms of categories. The idea that the visual process encodes the stimulus in terms of categories is not incompatible with the independence thesis, providing they are the right kinds of categories. The very fact that the mapping from the distal environment to a percept is many-to-one means that the visual system induces a partition of the visual world into equivalence classes (many-to-one mappings collapse differences and in so doing mathematically define equivalence classes). This is another way of saying that vision divides the perceptual world into some kinds of categories. But these kinds of categories are not what Bruner and others mean when they speak of the categories of visual perception. The perceptual classes induced by early vision are not the kinds of classes that are the basis for the claimed effects of set and expectations. They do not, for example, correspond to meaningful categories in terms of which objects are identified when we talk about perceiving as, for example, perceiving something as a face or as Mary's face and so on. To a first approximation the classes provided by the visual system are shape-classes, expressible in something like the vocabulary of geometry.

Notice that the visual system does not identify the stimulus in the sense of cross-referencing it to the perceiver's knowledge base, the way a unique internal label might. That is because the category identity is inextricably linked to past encounters and to what one knows about members of the category (e.g., what properties – visual and nonvisual – they have). After all, identifying some visual stimulus as your sister does depends on knowing such things as that you have a sister, what she looks like, whether she recently dyed her hair, and so on. But, according to the present view, computing what the stimulus before you looks like – in the sense of computing some representation of its shape, sufficient to pick out the class of similar appearances[14] – and hence to serve as an index into long-term memory – does not itself depend on knowledge.

According to this view, the visual system is seen as generating a set of one or more shape-descriptions that might be sufficient (perhaps in concert with other contextual information) to identify objects stored in memory. This provisional proposal was put forward to try to build a bridge between what the visual system delivers, which, as we have seen, cannot itself be the identity of objects, and what is stored in memory that enables identification. Whatever the details of such a bridge turn out to be, we still have not addressed the question of how complex or detailed or articulated this output is. Nor have we addressed the interesting question of whether there is more than one form of output; that is, whether the output of the visual system can be viewed as unitary or whether it might provide different outputs for different purposes or to different parts of the mind/brain. This latter idea, which is related to the "two visual systems" hypothesis, will be discussed in section 7.3.

The precise nature of the output in specific cases is an empirical issue that we cannot prejudge. There is a great deal that is unknown about the output – for example, whether it has a combinatorial structure that distinguishes individual objects and object-parts or whether it encodes nonvisual properties, such as causal relations, or primitive affective properties like "dangerous," or even some of the functional properties that Gibson referred to as "affordances." There is no reason why the visual system could not encode any property whose identification does not require accessing long-term memory, and in particular that does not require inference from general knowledge. So, for example, it is possible in principle for overlearned patterns – even patterns such as printed words – to be recognized from a finite table of pattern information compiled into the visual system. Whether or not any particular hypothesis is supported remains an open empirical question.[15]

Although there is much we do not know about the output of the visual system, we can make some general statements based on available evidence. We already have in hand a number of theories and confirming evidence for the knowledge-independent derivation of a three-dimensional representation of visible surfaces – what David Marr called the 2.5-D sketch. Evidence provided by J. J. Gibson, from a very different perspective, also suggests that what he called the "layout" of the scene may be something that the visual system encodes (Gibson would say "picks up") without benefit of knowledge and reasoning. Nakayama et al.

(1995) have also argued that the primary output of the independent visual system is a set of surfaces laid out in depth. Their data show persuasively that many visual phenomena are predicated on the prior derivation of a surface representation. These surface representations also serve to induce the perception of the edges that delineate and "belong to" those surfaces. Nakayama et al. argue that because of the prevalence of occlusions in our world it behooves any visual animal to solve the surface-occlusion problem as quickly and efficiently and as early as possible in the visual analysis and that this is done by first deriving the surfaces in the scene and their relative depth.

Although the evidence favors the view that some depth-encoded surface representation of the layout is present in the output of the early-vision system, nothing about this evidence suggests either (a) that no intermediate representations are computed or (b) that the representation is complete and uniform in detail – like an extended picture.

With regard to (a), there is evidence of intermediate stages in the computation of a depth representation. Indeed the time-course of some of the processing has been charted (e.g., Reynolds 1981; Sekuler & Palmer 1992) and there are computational reasons why earlier stages may be required (e.g., Marr's Primal Sketch). Also there is now considerable evidence from both psychophysics and from clinical studies that the visual system consists of a number of separate subprocesses that compute color, luminance, motion, form, and 3-D depth and that these subprocesses are restricted in their intercommunication (Cavanagh 1988; Livingston & Hubel 1987). In other words, the visual process is highly complicated and articulated and there are intermediate stages in the computation of the percept during which various information is available in highly restricted ways to certain specific subprocesses. Yet despite the clear indication that several types and levels of representation are being computed, there is no evidence that these interlevels and outputs of specialized subprocesses are available to cognition in the normal course of perception. So far the available evidence suggests that the visual system is not only cognitively impenetrable, but is also opaque with respect to the intermediate products of its process.

With regard to (b), the phenomenology of visual perception might suggest that the visual system provides us with a rich panorama of meaningful objects, along with many of their properties such as their color, shape, relative location, and perhaps even their "affordances" (as Gibson, 1979, claims). Yet phenomenology turns out to be an egregiously unreliable witness in this case. Our subjective experience of the world fails to distinguish among the various sources of this experience, whether they arise from the visual system or from our beliefs. For example, as we cast our gaze about a few times each second we are aware of a stable and highly detailed visual world. Yet careful experiments (O'Regan 1992; Irwin 1993) show that from one glance to another we retain only such sparse information as we need for the task at hand. Moreover, our representation of the visual scene is unlike our picture-like phenomenological impression, insofar as it can be shown to be nonuniform in detail and abstractness, more like a description cast in the conceptual vocabulary of mentalese than like a picture (Pylyshyn 1973; 1978). As I and many others have pointed out, what we see – the content of our phenomenological experience – is the world as we visually apprehend and know it; it is not the output of the visual system itself. Phenomenology is a rich source of evidence about how vision works and we would not know how to begin the study of visual perception without it. But like many other sources of evidence it has to be treated as just that, another source of evidence, not as some direct or privileged access to the output of the visual system. The output of the visual system is a theoretical construct that can only be deduced indirectly through carefully controlled experiments.[16] Exactly the same can be, and has been, said of the phenomenology of mental imagery – see, for example, Pylyshyn (1973; 1981).

7.3. Control of motor actions

In examining the nature of the output of the visual system we need to consider the full range of functions to which vision contributes. It is possible that if we consider other functions of vision besides its phenomenal content (which we have already seen can be highly misleading) and its role in visual recognition and knowledge acquisition, we may find that its outputs are broader than those we have envisaged. So far, we have been speaking as though the purpose of vision is to provide us with knowledge of the world. Vision is indeed the primary way that most organisms come to know the world and such knowledge is important in that it enables behavior to be detached from the immediately present environment. Visual knowledge can be combined with other sources of knowledge for future use through inference, problem-solving, and planning. But this is not the only function that vision serves. Vision also provides a means for the immediate control of actions and sometimes does so without informing the rest of the cognitive system – or at least that part of the cognitive system that is responsible for recognizing objects and for issuing explicit reports describing the perceptual world. Whether this means that there is more than one distinct visual system remains an open question. At the present time the evidence is compatible with there being a single system that provides outputs separately to the motor control functions or to the cognitive functions. Unless it is shown that the actual process is different in the two cases, this remains the simplest picture. So far it appears that in both cases the visual system computes shape-descriptions that include sufficient depth information to allow not only recognition, but also reaching and remarkably efficient hand positioning for grasping (Goodale 1988). The major difference between the information needed in the two cases is that motor-control primarily requires quantitative, egocentrically calibrated spatial information, whereas the cognitive system is concerned more often with more qualitative information in an object-centered frame of reference (Bridgeman 1995).

The earliest indications of the fractionation of the output of vision probably came from observations in clinical neurology (e.g., Holmes 1918) which will be discussed in section 7.3.2. However, it has been known for some time that the visual control of posture and locomotion can make use of visual information that does not appear to be available to the cognitive system in general. For example, Lee and Lishman (1975) showed that posture can be controlled by the oscillations of a specially designed room whose walls are suspended inside a real room and can be made to oscillate slowly. Subjects standing in such an "oscillating room" exhibit synchronous swaying even though they are totally unaware of the movements of the walls.

7.3.1. Visual control of eye movements and reaching. The largest body of work showing a dissociation between visual information available to high-level cognition and information available to a motor function involves studies of the visual control of eye movements as well as the visual control of reaching and grasping. Bridgeman (1992) has shown a variety of dissociations between the visual information available to the eye movement system and that available to the cognitive system. For example, he showed that if a visual target jumps during an eye movement, and so is undetected, subjects can still accurately point to the correct position of the now-extinguished target. In earlier and closely related experiments, Goodale (1988) and Goodale et al. (1986) also showed a dissociation between information that is noticed by subjects and information to which the motor system responds. In reaching for a target, subjects first make a saccadic eye movement toward the target. If, during the saccade, the target undergoes a sudden displacement, subjects do not notice the displacement because of saccadic suppression. Nonetheless, the trajectory of their reaching shows that their visual system did register the displacement and the motor system controlling reaching is able to take this into account in an on-line fashion and to make a correction during flight in order to reach the final correct position.

Wong and Mack (1981) and subsequently Bridgeman (1992) showed that the judgment and motor system can be given even conflicting visual information. The Wong and Mack study involved stroboscopically induced motion. A target and frame both jumped in the same direction, although the target did not jump as far as the frame. Because of induced motion, the target appeared to jump in the opposite direction to the frame. Wong and Mack found that the saccadic eye movements resulting from subjects' attempts to follow the target were in the actual direction of the target, even though the perceived motion was in the opposite direction (by stabilizing the retinal location of the target the investigators ensured that retinal error could not itself drive eye movements). But if the response was delayed, the tracking saccade followed the perceived (illusory) direction of movement, showing that the motor-control system could use only immediate visual information. The lack of memory in the visuomotor system has been confirmed in the case of eye movements by Gnadt et al. (1991) and in the case of reaching and grasping by Goodale et al. (1994). Aglioti et al. (1995) also showed that size illusions affected judgments but not prehension (see also Milner & Goodale 1995).

7.3.2. Evidence from clinical neurology. Clinical studies of patients with brain damage provided some of the earliest evidence of dissociations of functions that, in turn, led to the beginnings of a taxonomy (and information-flow analyses) of skills. One of the earliest observations of independent subsystems in vision was provided by Holmes (1918) who described a gunshot victim who had normal vision as measured by tests of acuity, color discrimination, and stereopsis, and had no trouble visually recognizing and distinguishing objects and words. Yet this patient could not reach for objects under visual guidance (though it appears that he could reach for places under tactile guidance). This was the first in a long series of observations suggesting a dissociation between recognition and visually guided action. The recent literature on this dissociation (as studied in clinical cases, as well as in psychophysics and animal laboratories) is reviewed by Milner and Goodale (1995).

Milner and Goodale (1995) have reported another remarkable visual agnosia patient (DF) in a series of careful investigations. This patient illustrates the dissociation of vision-for-recognition from vision-for-action, showing a clear pattern of restricted communication between early vision and subsequent stages, or, to put it in the terms that the authors prefer, a modularization that runs through from input to output, segregating one visual pathway (the dorsal pathway) from another (the ventral pathway). DF is seriously disabled in her ability to recognize patterns and even to judge the orientation of simple individual lines. When she was asked to select a line orientation from a set of alternatives that matched an oblique line in the stimulus, DF's performance was at chance. She was also at chance when asked to indicate the orientation of the line by tilting her hand. But when presented with a tilted slot and asked to insert her hand or to insert a thin object, such as a letter, into the slot, her behavior was in every respect normal – including the acceleration/deceleration and dynamic orienting pattern of her hand as it approached the slot. Her motor system, it seems, knew exactly what orientation the slot was in and could act toward it in a normal fashion.

Another fascinating case of visual processes providing information to the motor control system but not the rest of cognition is shown in cases of so-called blindsight. This condition is discussed by Weiskrantz (1995). Patients with this disorder are "blind" in the region of a scotoma in the sense that they cannot report "seeing" anything presented in that region. Without "seeing" in that sense patients never report the existence of objects in that region nor any other visual properties located in that part of his visual field. Nonetheless, such patients are able to do some remarkable things that show that visual information is being processed from the blind field. In one case (Weiskrantz et al. 1974), the patient's pupilary response to color and light and spatial frequencies showed that information from the blind field was entering the visual system. This patient could also move his eyes roughly toward points of light that he insisted he could not see and, at least in the case of Weiskrantz's patient DB, performed above chance in a task requiring reporting the color of the light and whether it was moving. DB was also able to point to the location of objects in the blind field while maintaining that he could see nothing there and was merely guessing. When asked to point to an object in his real blind spot (where the optic nerve leaves the eye and no visual information is available), however, DB could not do so.

Although it is beyond the scope of this paper, there is also a fascinating literature on the encapsulation of certain visual functions in animals. One particularly remarkable case, reported by Gallistel (1990) and Cheng (1986), shows the separation between the availability of visual information for discrimination and its availability for navigation. Gallistel refers to a "geometrical module" in the rat because rats are unable to take into account reflectance characteristics of surfaces (including easily discriminated texture differences) in order to locate previously hidden food. They can only use the relative geometrical layouts of the space and simply ignore significant visual cues to disambiguate symmetrically equivalent locations. In this case the output of the visual system either is selective as to where it sends its output or else there is a separate visuomotor subsystem for navigation.

Although it is still too early to conclude, as Milner and Goodale (1995) and many other researchers do, that there are two (or more) distinct visual (or visuomotor) systems, it is clear from the results sketched above that there are at least two different forms of output from vision and that these are not equally available to the rest of the mind/brain. It appears, however, that they all involve a representation that has depth information and that follows the couplings or constancies or Gogel's perceptual equations (see Rock 1997), so at least this much of the computations of early vision is shared by all such systems.

8. Conclusions: Early vision as a cognitively impenetrable system

In this article we have considered the question of whether visual perception is continuous with (i.e., a proper part of) cognition or whether a significant part of it is best viewed as a separate process with its own principles and possibly its own internal memory (see n. 7), isolated from the rest of the mind except for certain well-defined and highly circumscribed modes of interaction. In the course of this analysis, we have touched on many reasons why it appears on the surface that vision is part of cognition and thoroughly influenced by our beliefs, desires, and utilities. Opposed to this perspective are a number of clinical findings concerning the dissociation of cognition and perception, and a great deal of psychophysical evidence attesting to the autonomy and inflexibility of visual perception and its tendency to resolve ambiguities in a manner that defies what the observer knows and what is a rational inference. As one of the champions of the view that vision is intelligent has said, "Perception must rigidly adhere to the appropriate internalized rules, so that it often seems unintelligent and inflexible in its imperviousness to other kinds of knowledge." (Rock 1983, p. 340).

In examining the evidence that vision is affected by expectations, we devoted considerable space to methodological issues concerned with distinguishing various stages of perception. Although the preponderance of evidence locates such effects in a post-perceptual stage, we found that stage analysis methods generally yielded a decomposition that is too coarse to definitively establish whether the locus of all cognitive effects is inside or outside of vision proper. In particular, we identified certain shortcomings in using signal detection measures to establish the locus of cognitive effects and argued that although event-related potentials might provide timing measures that are independent of response-preparation, the stages they distinguished are also too coarse to factor out such memory-accessing decision functions as those involved in recognition. So, as in so many examples in science, there is no simple and direct method – no methodological panacea – for answering the question whether a particular observed effect has its locus in vision or in pre- or post-visual processes. The methods we have examined all provide relevant evidence but in the end it is always how well a particular proposal stands up to convergent examination that will determine its survival.

The bulk of this article concentrated on showing that many apparent examples of cognitive effects in vision arise either from a post-perceptual decision process or from a pre-perceptual attention-allocation process. To this end we examined alleged cases of "hints" affecting perception, of perceptual learning, and of perceptual expertise. We argued that in the cases that have been studied carefully, as opposed to reported informally, hints and instructions rarely have an effect, but when they do it is invariably by influencing the allocation of focal attention, by the attenuation of certain classes of physically specifiable signals, and in certain circumstances by the development of such special skills as the control of eye movements and eye vergence. A very similar conclusion was arrived at in the case of perceptual learning and visual expertise, where the evidence pointed to the improvement being due to learning where to direct attention – in some cases aided by better domain-specific knowledge that helps anticipate where the essential information will occur (especially true in the case of dynamic visual skills, such as in sports). Another relevant aspect of the skill that is learned is contained in the inventory of pattern-types that the observer assimilates (and perhaps stores in a special intra-visual memory) that helps in choosing the appropriate mnemonic encoding for a particular domain.

Finally, we discussed the general issue of the nature of the function computed by the early vision system and concluded that the output consists of shape representations involving at least surface layouts, occluding edges – where these are parsed into objects – and other details sufficiently rich to allow parts to be looked up in a shape-indexed memory in order to identify known objects. We suggested that in carrying out these complex computations the early vision system must often engage in top-down processing, in which there is feedback from global patterns computed later within the vision system to earlier processes. The structure of the visual system also embodies certain "natural constraints" on the function it can compute, resulting in a unique 3-D representation even when infinitely many others are logically possible for a particular input. Because these constraints developed through evolution they embody properties generally (though not necessarily) true in our kind of world, so the unique 3-D representation computed by early vision is often veridical. We also considered the likelihood that more than one form of output is generated, directed at various distinct post-perceptual systems. In particular we examined the extensive evidence that motor control functions are provided with different visual outputs than recognition functions – and that both are cognitively impenetrable.

ACKNOWLEDGMENTS
The author wishes to thank Jerry Fodor, Ilona Kovacs, Thomas Papathomas, Zoltan Vidnyanszky, as well as Tom Carr and several anonymous BBS referees for their comments and advice. Brian Scholl was especially helpful in carefully reading earlier drafts and providing useful suggestions. This work was initially supported by a grant from the Alfred P. Sloan Foundation and recently by the Rutgers Center for Cognitive Science.

NOTES

1. This thesis is closely related to what Fodor (1983) [see also *BBS* multiple book review: *The Modularity of Mind, BBS* (8) 1985] has called the "modularity of mind" and this article owes much to Fodor's ideas. Because there are several independent notions conflated in the general use of the term "module," we shall not use this term to designate cognitively impenetrable systems in this article.

2. Although my use of the term "early vision" generally corresponds to common usage, there are exceptions. For example, some people use "early vision" to refer exclusively to processes that occur in the primary visual cortex. Our usage is guided by an attempt to distinguish a functionally distinct system, regardless of its neuroanatomy. In placing focal attention outside (and prior to) the early vision system we depart somewhat from the use of the term in neuropsychology.

3. We sometimes use the term "rational" in speaking of cognitive processes or cognitive influences. This term is meant to indicate that in characterizing such processes we need to refer to what the beliefs are about – to their semantics. The paradigm case of such a process is *inference,* where the semantic property *truth* is preserved. But we also count various heuristic reasoning and decision-making strategies (e.g. satisficing, approximating, or even guessing) as rational because, however suboptimal they may be by some normative criterion, they do not transform representations in a semantically arbitrary way: they are in some sense at least quasi-logical. This is the essence of what we mean by cognitive penetration: it is an influence that is coherent or quasi-rational when the meaning of the representation is taken into account.

4. I use the technical term *content* (as in "the content of perception") in order to disambiguate two senses of "what you see." "I see a dog" can mean either that the thing I am looking at is a dog, regardless of how it appears to me, or that I see the thing before me as a dog, regardless of what it actually is. The second (opaque) sense of "see" is what I mean when I speak of the content of one's percept.

5. Bruner (1957) characterized his claim as a "bold assumption" and was careful to avoid claiming that perception and thought were "utterly indistinguishable." In particular, he explicitly recognized that perception "appear(s) to be notably less docile or reversible" than "conceptual inference." This lack of "docility" will play a central role in the present argument for the distinction between perception and cognition.

6. Note that not all cases of Gestalt-like global effects need to involve top-down processing. A large number of global effects turn out to be computable without top-down processing by arrays of elements working in parallel, with each element having access only to topographically local information (see, for example, the network implementations of such apparently global effects as those involved in stereo fusion [described in Marr 1982] and apparent motion [Dawson & Pylyshyn 1988]). Indeed many modern techniques for constraint propagation rely on the convergence of locally based parallel processes onto global patterns.

7. An independent system may contain its own proprietary (local) memory – as we assume is the case when recent visual information is stored for brief periods of time or in the case of the natural language lexicon, which many take to be stored inside the language "module" (Fodor 1983). A proprietary memory is one that is functionally local (as in the case of local variables in a computer program). It may, of course, be implemented as a subset of long-term memory.

8. A number of studies have shown a reliable effect due to the lexical item in which the phoneme is embedded (e.g., Connine & Clifton 1987; Elman & McClelland 1988; Samuel 1996). This is perfectly compatible with the independence of perception thesis since, as pointed out by Fodor (1983), it is quite likely that the lexicon is stored in a local memory that resides within the language system. Moreover, the network of associations among lexical items can also be part of the local memory since associations established by co-occurrence are quite distinct from knowledge, whose influence, through inference from the sentential context, is both semantically compliant and transitory. Since we are not concerned with the independence of language processing in this article, this issue will not be raised further.

9. Of course because it is really the bias for the $<F_i; C_i>$ pair that is being altered, the situation is symmetrical as between the F_is and the C_is so it can also be interpreted as a change in the sensitivity to a particular context in the presence of the phoneme in question – a prediction that may not withstand empirical scrutiny.

10. We cannot think of this as an "absolute" change in sensitivity to the class of stimuli because it is still relativized not only to the class but also to properties of the perceptual process, including constraints on what properties it can respond to. But it is not relativized to the particular choice of stimuli, or the particular response options with which the subject is provided in an experiment.

11. A view similar to this has recently been advocated by Barlow (1997). Barlow asks where the knowledge that appears to be used by vision comes from and answers that it may come from one of two places: "through innately determined structure [of the visual system] and by analysis of the redundancy in sensory messages themselves." We have not discussed the second of these but the idea is consistent with our position in this article, so long as there are mechanisms in early vision that can exploit the relevant redundancies. The early visual system does undergo changes as a function of statistical properties of its input, including co-occurrence (or correlational) properties, thereby in effect developing redundancy analyzers.

12. The "rigidity" constraint is not the only constraint operative in motion perception, however. In order to explain the correct perception of "biological motion" (e.g., Johansson 1950) or the simultaneous motion and deformation of several objects, additional constraints must be brought to bear.

13. There has been at least one reported case where the usual "natural constraint" of typical direction of lighting, which is known to determine perception of convexity and concavity, appears to be superseded by familiarity of the class of shapes. This is the case of human faces. A concave human mask tends to be perceived as convex in most lighting conditions, even ones that result in spherical shapes changing from appearing concave to appearing convex (Ramachandran 1990) – a result that leads many people to conclude that having classified the image as that of a face, knowledge overrides the usual early vision mechanisms. This could indeed be a case of cognitive override. But one should note that faces present a special case. There are many reasons for believing that computing the shape of a face involves special-purpose (perhaps innate) mechanisms (e.g., Bruce 1991) with a distinct brain locus (Kanwisher et al. 1997).

14. Begging the question of what constitutes similarity of appearance, we simply assume that something like similar-in-appearance defines an equivalence class that is roughly coextensive with the class of stimuli that receive syntactically similar (i.e., overlapping-code) outputs from the visual system. This much should not be problematic because, as we remarked earlier, the output necessarily induces an equivalence class of stimuli and this is at least in some rough sense a class of "similar" shapes. These classes could well be coextensive with basic-level categories (in the sense of Rosch et al. 1976). It also seems reasonable that the shape-classes provided by vision are ones whose names can be learned by ostension – that is, by pointing, rather than by providing a description or definition. Whether or not the visual system actually parses the world in these ways is an interesting question, but one that is beyond the scope of this essay.

15. One of the useful consequences of recent work on connectionist architectures has been the recognition that perhaps more cognitive functions than had been expected might be accomplished by table-lookup, rather than by computation. Newell (1990) recognized early on the important trade-off between computing and storage that a cognitive system has to face. In the case of the early vision system, where speed takes precedence over generality (cf. Fodor 1983), this could take the form of storing a forms table or set of templates in a special internal memory. Indeed, this sort of compiling of a local shape-table may be involved in some perceptual learning and in the acquisition of visual expertise (see also n. 7).

16. Needless to say, not everyone agrees on the precise status of subjective experience in visual science. This is a question that has been discussed with much vigor ever since the study of vision became an empirical science. For a recent revival of this discussion see Pessoa et al. (1998) and the associated commentaries.

Open Peer Commentary

Commentary submitted by the qualified professional readership of this journal will be considered for publication in a later issue as Continuing Commentary on this article. Integrative overviews and syntheses are especially encouraged.

Visual space is not cognitively impenetrable

Yiannis Aloimonos and Cornelia Fermüller

Computer Vision Laboratory, Center for Automation Research, University of Maryland, College Park, MD 20742-3275. **yiannis@cfar.umd.edu**
fer@cfar.umd.edu **www.cfar.umd.edu/~yiannis/**
www.cfar.umd.edu/~fer/

Abstract: Cognitive impenetrability (CI) of a large part of visual perception is taken for granted by those of us in the field of computational vision who attempt to recover descriptions of space using geometry and statistics as tools. These tools clearly point out, however, that CI cannot extend to the level of structured descriptions of object surfaces, as Pylyshyn suggests. The reason is that visual space – the description of the world inside our heads – is a nonEuclidean curved space. As a consequence, the only alternative for a vision system is to develop several descriptions of space–time; these are representations of reduced intricacy and capture partial aspects of objective reality. As such, they make sense in the context of a class of tasks/actions/plans/purposes, and thus cannot be cognitively impenetrable.

Researchers and practitioners of computational or computer vision should be sympathetic to Pylyshyn's viewpoint, expressed very eloquently in his target article. These disciplines are concerned with the geometry and physics of light and descriptions of the world that can be extracted from images – snapshots of the optic array. At the same time, these disciplines provide insights into the visual system through the use of geometry and statistics (Fermüller & Aloimonos 1998; Fermüller et al. 1997b). If what can be seen depends heavily on cognition – purpose, experience, will, and so on – it would make little sense to attempt to develop representations of space–time from images using only geometry and statistics as tools. Thus, as computational theorists, we welcome Pylyshyn's thesis and the simultaneous admission by a part of the psychological community that a large part of visual perception is cognitively impenetrable. This has been an assumption in our field all along, emphasized by terms such as early vision (Marr 1982) or visual front-end machinery (Koenderink et al. 1992).

The real questions, however, whose answers can lead to fruitful ways of understanding visual perception, are related to the exact nature of this cognitively impenetrable part of perception. Pylyshyn's arguments, which are not computational, lead to the conclusion that this part of perception provides a structured representation of three-dimensional (3-D) surfaces of objects serving as an index into memory, with somewhat different outputs being made available to motor control and other subsystems. In the remainder of this brief commentary, we present a number of computational arguments explaining why this cannot be true. We concentrate on the motion pathway (Zeki 1993) or, in general, on the process of recovering 3-D shape descriptions from multiple views (Fermüller & Aloimonos 1995a; 1995b).

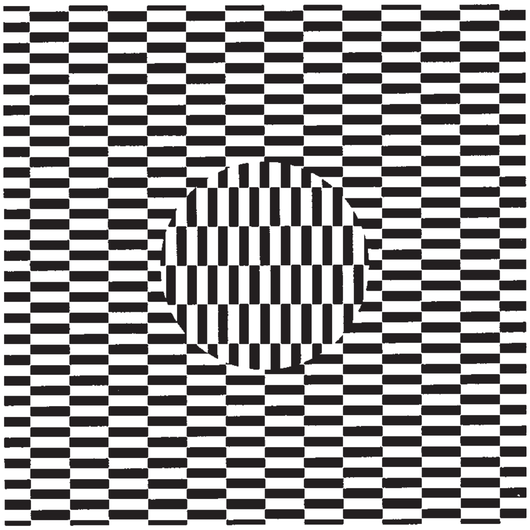


Figure 1 (Aloimonos & Fermüller). A pattern similar to one by Ouchi.

Examine the pattern in Figure 1, similar to one designed by the Japanese artist, Ouchi, showing the surprising property that small motions can cause illusory relative motion between the inset and background regions. The effect can be obtained with small retinal motions or a slight jiggling of the paper, and is robust over large changes in the patterns, frequencies, and boundary shapes. It has been shown (Fermüller et al. 1998b) that the cause of the illusion lies in the statistical difficulty of integrating local, one-dimensional motion signals into two-dimensional image velocity measurements. The estimation of image velocity generally is biased, and for the particular spatial gradient distributions of the Ouchi pattern the bias is highly pronounced, giving rise to a large difference in the velocity estimates in the two regions. This visual illusion demonstrates that not even image motion fields (Fermüller & Aloimonos 1997) can be computed exactly. From a mathematical viewpoint, the computation is biased and this bias is practically impossible to correct. Considering the statistical difficulties in image velocity estimation in conjunction with the problem of discontinuity detection in motion fields suggests that, theoretically, the process of optical flow computation should not be carried out in isolation but in conjunction with higher level processes such as 3-D motion estimation, segmentation, and shape computation, thereby providing more information for individual computations (Brodský et al. 1998). However, the inherent ambiguities in motion field estimation preclude the system from recovering veridical structure (Cheong et al. 1998). The only alternative is that the system extracts shape descriptions of reduced intricacy, lying between the projective and Euclidean layer (Aloimonos 1995; Aloimonos et al. 1995; Fermüller & Aloimonos 1996; Fermüller et al. 1998a). These descriptions do not encode the exact 3-D distances between features in the scene, that is, they cannot be used to reconstruct an exact Euclidean model of the world. As such, they make sense only in the context of a task.

Evidence for shape descriptions that do not encode exact distances in the world also comes from recent experiments aimed at finding how surfaces of objects are represented in our brains. It turns out that humans estimate a distorted version of their extrapersonal space. Over the years, a number of experiments have been performed to study depth perception due to stereo or motion, using tasks that involve the judgment of depth at different distances (Foley 1967; 1980; Johnston 1991; Ogle 1964; Tittle et al. 1995). In these experiments, it has been shown that in stereovision humans overestimate depth (relative to fronto-parallel size) at near fixations and underestimate it at far fixations, and that the orientation of an object in space has a strong effect on motion vision for the class of motions tested. A computational geometric model has been developed (Fermüller et al. 1997a) explaining why such distortions might take place. The basic idea is that, both in stereo and motion, we perceive the world from multiple views. Given the rigid transformation between the views and the properties of the image correspondence, the depth of the scene can be obtained. Even a slight error in the rigid transformation parameters causes distortion of the computed depth of the scene (Fermüller et al. 1997a). It has been shown that, for the case of motion or stereo, the transformation relating actual space to perceptual space is a Cremona transformation (Fermüller et al. 1997a).

To summarize, although the space in which we move and live is clearly Euclidean, our visual world does not have a Euclidean structure, that is, the representations that humans build in their heads about the shapes of the scenes they view do not reflect the metric structure of the 3-D world in the sense of encoding the exact distances between different points. The nature of the distortion and thus the global geometric structure of perceptual space remains largely unknown, and partly depends on the cues used by the system to estimate shape. Because perceptual space is a non-Euclidean curved one, the only alternative from a computational viewpoint is for the system to create several descriptions of its spatiotemporal environment that it can use to accomplish a few classes of tasks (such as those needed in navigation and recognition), which comprise its repertoire of actions/decisions/plans. These descriptions are thus not general (Fermüller & Aloimonos 1995c) and, as they make sense in the context of a class of tasks, they cannot be cognitively impenetrable. We suspect that cognitive impenetrability ends quite early, possibly after the computation of a number of spatiotemporal derivatives of the image function. This does not, of course, mean that the part of perception responsible for recovering descriptions of space–time requires cognition in order to be understood. It simply means that there are many such descriptions. Empirical and theoretical scientists of vision have to consider purpose, goals, and action in their investigations (Aloimonos 1993a). In computational vision, geometric/statistical studies of how the 3-D world is imprinted on images seem to be the only avenue for uncovering these representations (Aloimonos 1993b). Thus, although Pylyshyn's thesis does not appear correct, from a methodological viewpoint it is the most fruitful in the study of visual perception. Geometry and statistics will tell us how shape descriptions are to be recovered, but at some point during computation the system has to "stick its neck out" – that is, it has to commit to a particular set of assumptions coming from the context of a task. In this way, invariances of the distorted spaces can be estimated and the system can properly function (Fermüller & Aloimonos 1998).

ACKNOWLEDGMENT
Special thanks to Sara Larson for her editorial and graphics assistance.

Cognitive impenetrability, phenomenology, and nonconceptual content

José Luis Bermúdez
Department of Philosophy, University of Stirling, Stirling FK9 4LA, Scotland.
jb10@stir.ac.uk www.stir.ac.uk/philosophy/cnw/webpage1.htm

Abstract: This commentary discusses Pylyshyn's model of perceptual processing in the light of the philosophical distinction between the conceptual and the nonconceptual content of perception. Pylyshyn's processing distinction maps onto an important distinction in the phenomenology of visual perception.

Pylyshyn is rightly sceptical about the authority of subjective experience and introspective report in the study of visual experience. Nonetheless, there are important questions to be asked about how his model of perceptual processing maps onto the phenomenology of perception. Until they are tackled it is unclear what is at stake between his approach and the cognitivist approach to visual perception.

Pylyshyn is attacking Bruner's (1957) claim that "all perceptual experience is necessarily the end product of a categorization process." Against this he maintains that visual processing involves an inflexible and cognitively impenetrable early vision stage. It is not immediately clear, though, that these two proposals are incompatible. Bruner is making a claim about perceptual experience and it is perfectly possible that perceptual experience could be as he describes and still be the product of several stages of processing, including an early vision component as described by Pylyshyn. Indeed, what Pylyshyn says about perceptual experience sounds very much like Bruner: for example, "what we see – the content of our phenomenological experience – is the world as we visually apprehend and know it: it is not the output of the [early] visual system itself" (sect. 7.2).

The two positions will only be incompatible if the outputs of the early vision system form a distinct, isolable, and phenomenologically salient part of perceptual experience. That this is sometimes the case is implied by Pylyshyn's discussion of perceptual illusions in section 2. In the following, I sketch out a philosophical theory of perceptual experience that generalizes this.

We need to identify two different levels of perceptual content, one appropriate for the outputs of the early visual system and one for the outputs of the evaluation/decision systems. This can be done in terms of the distinction between conceptual and nonconceptual content (Bermúdez 1998a; Evans 1982; Peacocke 1992). The distinction between conceptual and nonconceptual content is a function of concept possession. A creature has perceptions with nonconceptual content to the extent that it represents the environment in ways that outstrip the concepts it possesses – either because its concepts are insufficiently fine grained or, more radically, because it lacks concepts at all. That is to say, it sees an object under a certain aspect which it does not have the resources to conceptualise. A creature has perceptions with conceptual contents, on the other hand, to the extent that its perceptual representations of the environment are determined by its classificatory and recognitional abilities.

The obvious practical and adaptive advantage in representing the environment nonconceptually is in the control and initiation of motor behaviour. Navigating successfully through the environment often does not require identifying and conceptualising the objects it contains. It is more important to have perceptual sensitivity to information about a limited range of object properties – position, motion, colour, relative size, texture, distance, and so on – as well as to the forms of kinaesthetic information, both exteroceptive and proprioceptive, discussed by Gibson (1979). Perceptual sensitivity to these properties can feed into motor behaviour, of course, without any ability to conceptualise the information picked up. Thus an infant can reach out towards an object which it perceives as being within reach even though it has no concept of distance or reaching. And, as Pylyshyn suggests in section 7.2 it is most likely representations of properties of this type that are computed by the early visual system. The nonconceptual nature of the outputs of the early vision system explains their cognitive impenetrability, on the plausible assumptions that beliefs and utilities are conceptual and can only cognitively influence other conceptual representations.

Nonconceptual content is a distinct and isolable component in visual perception. Our abilities to act upon the distal environment are underwritten by the nonconceptual component in visual perception. It is this that provides the egocentrically salient information about the layout, relative position and motion of objects in the environment. The dissociations between saccadic eye movements and perceptual report which Pylyshyn notes in section 7.3.1 show that the nonconceptual content controlling action is not determined by how things seem to the subject at the conceptual level. Moreover, there is good evidence for double dissociations in the neuropsychological disorders that Pylyshyn discusses in section 7.3.2 – visual disorientation and the associative agnosias. In visual disorientation (Holmes 1918) we see the conceptual content of perception preserved in the absence of the nonconceptual content, whereas the nonconceptual content is preserved at the expense of the conceptual content in the associative agnosias (Bermúdez 1998b). More significant, though, is evidence from developmental psychology. It is becoming clear that the perceptual experience of even the youngest infants is organised and structured in a way that reflects perceptual sensitivity to precisely those properties whose representation is the function of the early vision system (Spelke 1990). It is misleading to suggest, as many workers in the area do, that this aspect of infant cognition reflects infant mastery of the concept of an object (Bermúdez 1995). Information is being processed nonsemantically.

Once this picture of the phenomenology of visual perception is taken on board, the traditional debate about where perception starts and cognition begins starts to look simplistic. One should ask instead about the location of two boundaries: the boundary between the conceptual and the nonconceptual content of perception and the boundary between the conceptual content of perception and the content of beliefs based upon perception. Pylyshyn's discussion of cognitive impenetrability makes an important contribution to fixing the first boundary. It will be interesting to see whether the second can be fixed as cleanly.

The visual categories for letters and words reside outside any informationally encapsulated perceptual system

Jeffrey S. Bowers
Department of Experimental Psychology, University of Bristol, Bristol BS8-1TN, England. **j.bowers@bris.ac.uk**

Abstract: According to Pylyshyn, the early visual system is able to categorize perceptual inputs into shape classes based on visual similarity criteria; it is also suggested that written words may be categorized within early vision. This speculation is contradicted by the fact that visually unrelated exemplars of a given letter (e.g., *a/A*) or word (e.g., *read/READ*) map onto common visual categories.

Pylyshyn, like Fodor (1983) before him, is uncommitted about the exact boundaries of the cognitively impenetrable early vision system but argues that this system computes as much as possible, given its inputs and the visual knowledge internal to the system. These computations would include the derivation of a three-dimensional representation of visible surfaces (the 2.5-D sketch of David Marr), and the determination of perceptual categories that organize inputs into shape classes, expressible in something like a vocabulary of geometry. That is, inputs are categorized on the basis of visual similarity, which in the case of objects, might be coextensive with basic-level categories. Pylyshyn also raises the possibility that printed words may map onto perceptual categories within early vision, again based on the visual similarity of word exemplars (sect. 7.2). These categories would essentially function as a proprietary (local) memory embedded within the early visual system.

In this commentary, I want to summarize evidence that the perceptual categories for words are not structured on the basis of geometrical similarity and must hence lie outside early vision. These findings contradict the claim that visual word identification is mediated by a modular system (e.g., Polk & Farah 1997; Seidenberg 1987), and further, highlight the complexity of identifying to boundaries of early vision. Indeed, to the extent that perceptual word categories are non-encapsulated, it raises questions as to whether other perceptual categories, such as basic level object categories, reside in an encapsulated system.

The difficulty in arguing that perceptual categories for words – so-called orthographic codes – are encapsulated within an early visual system is that different exemplars of a given letter or word are often unrelated in their visual form, and nevertheless, they map onto the appropriate perceptual category. For example, the visual patterns *A/a* or *READ/read* map to the same abstract letter and word codes, respectively. Clearly, perceptual categories that map together these forms cannot be learned in an encapsulated system that only has access to "bottom-up" visual input.

Evidence for the existence of abstract orthographic codes comes from a variety of sources. Coltheart (1981), for example, describes a conduction aphasic patient who could not name individual letters or name pseudowords (e.g., *nega*), but who nevertheless was able to match upper/lower pseudowords that were perceptually dissimilar (e.g., *NEGA/nega*) without difficulty. Given that these items are (1) meaningless, (2) perceptually dissimilar in upper/lower case, and (3) unpronounceable by the patient, Coltheart concluded that the patient must have accessed abstract orthographic codes in order to perform the task, what he called "abstract letter identities." Consistent with this conclusion, McClelland (1976) reported that the word superiority effect (WSE) is equally large for words presented in case uniform and mixed conditions; for example, the words *FADE* and *fAdE* were both better identified than the matched pseudowords *GADE* and *gAdE* in a task in which participants were required to identify briefly presented targets. Given that mixed-case words are unfamiliar visual patterns, these results suggest that the WSE is mediated by word representations coded in an abstract fashion. In addition, Bowers (1996) found long-term priming to be equally robust for words repeated in the same and different case, even though the different-case words were perceptually dissimilar at study and test (e.g., *READ/read*). This cross-case priming was attributed to orthographic knowledge, since the priming was dramatically reduced following a study/test modality shift in which words were studied auditorily and tested visually. Bowers and Michita (1998) extended this finding, observing robust priming between the Hiragana and Kanji scripts of Japanese, and this priming was again modality specific, indicating that these unrelated visual patterns map to common lexical-orthographic representations (for additional evidence in support the existence of abstract orthographic knowledge, see Besner et al. 1984; Bowers et al. 1998; Rayner et al. 1980, among others).

Taken together, these findings strongly support the conclusion that orthographic knowledge is organized into abstract perceptual categories that cannot be learned on the basis of the visual properties of the input. Accordingly, it is necessary to assume that a nonvisual "teacher" acts on the orthographic system in order to organize the perceptual representations of words. One possible account of this teacher is outlined in Bowers and Michita (1998). Briefly, it is argued that there are bi-directional connections between orthographic knowledge on the one hand, and phonological and lexical-semantic codes on the other, consistent with some recent experimental findings and models of reading (e.g., Stone et al. 1997). On this hypothesis, phonological and lexical-semantic codes act together as an external teacher to construct abstract orthographic codes, based on associationist learning principles. As a result, the perceptual categories for words get structured in such a way that they are consistent with our background knowledge of the sounds and meanings of words.

To see how phonology may act as a teacher and penetrate into the visual system, consider the two arbitrarily related visual letters (e.g., *a/A*), as depicted in Figure 1. In this figure, the different visual patterns map onto the same phonological representation, and because of bi-directional connections between orthography and phonology, both orthographic patterns are consistently co-activated within the orthographic system, via feedback. It is this co-activation that makes it possible to learn arbitrary perceptual mappings. More specifically, the learning process might proceed as follows. When the child learns that visual pattern "A" maps to sound /ei/, bi-directional connections develop such that the presentation of "A" leads to the activation of /ei/, and conversely, the presentation of the sound /ei/ activates the orthographic pattern "A." Similarly, when the child learns that the visual pattern "a" maps onto /ei/, bi-directional connections develop. As a result, when one of the visual patterns is presented, for example "A," it activates /ei/, which in turn activates "A" and "a," given the learned feedback connections. This co-activation, when combined with associative learning principles, provides a simple mechanism for acquiring abstract orthographic representations. That is, a learning rule would associate representations that are consistently co-activated within the orthographic system, which would include such items as upper and lower case letters.

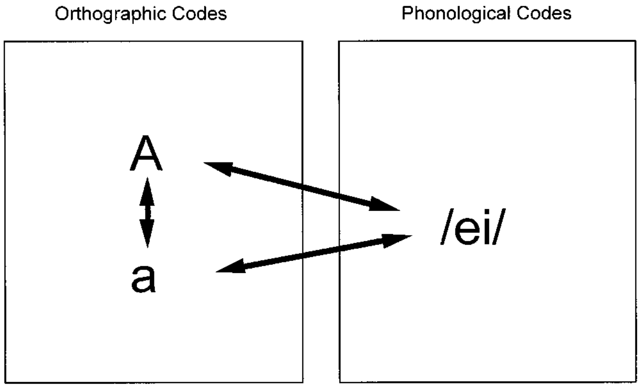


Figure 1 (Bowers). The visual patterns A and a each map onto the phonological code /ei/, which results in the coactivation of *A* and *a* each time one of the patterns is presented, via feedback. Associative learning principles within the orthographic system act to map together these coactive patterns 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/
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 that 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 divison 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 claim that perception, or specifically, early vision, is unlike cognition and cannot use higher order processes such as inference. Complex perceptual processes are not inferential, they are merely compiled, built-in procedures. Not only is vision not cognition, it is not even like it. Cognition is the sole site of reasoning and rational problem solving.

This is serious cognocentrism. Pylyshyn confuses what people know with what cognition knows and what perception really ought to know. This does not demonstrate that perception and cognition use different procedures. If I make a decision despite something only you know, you do not call me irrational, you call me ignorant. And the same holds for perception. It is not irrational, it is just ignorant of the knowledge available to cognition.

You could analyze a picture cognitively, say, deciding whether a dark patch was a shadow, a dark object, or a marking on a lighter object by checking whether there were any possible objects in the scene for which this could be a shadow, whether the region was uniformly darker than its surround all along its border, whether there was a light source consistent with the shadow, or alternatively, whether the dark area itself could be recognized as a known object. If this were done on a numerical representation of the image, to disable any help from the visual system, we would note that this is a possible task for cognition but that it would be extremely slow. The visual system performs the very same, highly sophisticated and complex steps at great speed, interrogating knowledge bases, verifying image support, and selecting the eventual interpretation of the image from among several alternatives. Two things distinguish this process from cognition. It is extremely fast, whereas cognition is slow and its knowledge base is independent of the knowledge base used for cognition.

Clearly, perception and cognition have access to different knowledge bases – things known and reportable consciously have only indirect influence on perceptual outcomes. The details of the knowledge that drives perception cannot be reported verbally. However, this separation of knowledge bases is not limited to cognition and perception; it is often found within cognition itself. As only one example: religious beliefs are, almost by definition, held independently of rational analysis of the physical world – they are cognitively impenetrable. By choice.

Let us look at what Pylyshyn means when he says cognition. Cognition, he says, is present when a system's output "can be altered in a way that bears some logical relation to what the person knows." So if a person "knows" that the two lines of the Müller-Lyer illusion are the same length and yet persists in seeing them as different lengths, that percept is cognitively impenetrable. But who is the "person" who "knows" this fact about the two lines? The "person" cannot include a visual system because the visual system believes the lines are different and reports this as the percept. In truth, it can only be the verbally responsive, conscious part of the person that "knows" the lines are equal. Pylyshyn has linked cognition and cognitive-style processes solely to consciousness, to reportable knowing. [See also Puccetti: "The Case for Mental Duality" *BBS* 4(1) 1981.]

Not that cognition is restricted to conscious events. Clearly, much of the flow of cognition consists of inaccessible, unreportable gaps, memory retrievals, intuitions, rapid routines which return conscious results but whose details cannot be inspected. But the path of cognition is marked by a sequence of conscious states, like Hansel's trail of bread crumbs through the forest. Unless Pylyshyn defines it differently; what he claims we know appears to be only what we can tell another person.

In this case, Pylyshyn has not shown that vision is cognitively impenetrable, bereft of cognitive-style processes like inference and rationality. He has shown only that vision is impenetrable to consciousness. This fact alone does not constrain the nature of the processes used by vision. It does not rule out inference unless we accept that inference is solely a property of consciousness, but there would be no grounds for that rather strange assertion.

There are undoubtedly profound differences between vision and cognition but Pylyshyn has not identified differences in process, only differences in access to knowledge. What is needed is a description of the specific procedures which are unavailable to unconscious vision. If none can be named, no differences should be assumed.

Even feature integration is cognitively impenetrable

Dale J. Cohen[a] and Michael Kubovy[b]

Department of Psychology, [a]University of North Carolina at Wilmington, Wilmington, NC 28403; Department of Psychology, [b]University of Virginia, Charlottesville, VA 22903 **www.uncwiledu/people/cohend/**
cohend@uncwil.edu **www.virginia.edu/~mklab/kubovy**
kubovy@virginia.edu

Abstract: Pylyshyn is willing to assume that attention can influence feature integration. We argue that he concedes too much. Feature integration occurs preattentively, except in the case of certain "perverse" displays, such as those used in feature-conjunction searches.

Attention plays a central role in Pylyshyn's approach to the relation between cognition and perception. He claims that many of the effects generally attributed to the influence of cognition on early vision are actually due to the influence of attention on perception. We agree, but we think he has not gone far enough. He believes that attention is routinely called upon to perform the task of encoding combinations of features – the "attention as glue" hypothesis of Treisman (e.g., Treisman 1988) and others (sect. 6.4). Thus Pylyshyn believes that through attention, cognition can influence the perceptually essential task of feature integration.

We believe that feature integration *never* requires attention and that the evidence regarding the role of attention in feature integration stems from an overgeneralized interpretation of data obtained in visual search experiments. Consider a typical visual search of this type. Denote the target, say a black square, by T_{bs}, and two distractors, a black circle and a red square, by D_{bc} and D_{rs}. When observers are asked to search for a single T_{bs} target among N tokens of the D_{bc} distractor – called a single-feature search – the magnitude of N has only small, if any, effect on the time it takes them to find T_{bs}. If, however, observers are asked to search for the T_{bs} target among N_{bc} tokens of D_{bc} and N_{rs} tokens of D_{rs} (a conjunction search), the magnitude of $N_{\mathrm{bc}} + N_{\mathrm{rs}}$ has a large effect on the time it takes them to find T_{bs}.

The common explanation of such findings runs as follows: different features, such as color and form, are independently processed in different modules. The feature that differentiates the distractor from the target (in our example, form) activates a single module; the other module is assumed to be silent because the stimulus does not vary with respect to the dimension to which it is sensitive. Because finding the target in this display does not require feature integration, the output of one module can guide the observer directly to the target. Hence, in single-feature search, targets can be located with only a small influence of N. According to feature integration theory, even though the observer may be guided directly to the target, recognition of the target requires focal attention, for a reason that will presently become apparent. In the conjunctive search, neither module alone can locate the target. Therefore finding the target requires observers to combine the output of the modules, which they can do only by applying focal attention to each item serially. Hence in conjunctive-feature search, targets are located with a large influence of $N_{\mathrm{bc}} + N_{\mathrm{rs}}$.

The data do not force these conclusions. The design of the visual search display forces the observer to perform a serial search but not for the reasons usually given. It is reasonable to believe that a central function of early vision is to define the spatial boundaries of objects, and that the multiple retinotopic maps found in the cortex are modules that specialize in finding spatial boundaries within a certain dimension. As we will show later, (1) these modules do not work independently, and (2) features are routinely integrated preattentively.

Specifically, feature boundaries around objects are generally coincident; for example, color and texture both change abruptly at the edge of an object. The stimuli used in conjunction searches are "perverse" in the sense that the output of the preattentive integration that is the centerpiece of our theory would be two objects coexisting in the same region of space, a paradox. Returning to the example we gave earlier, if in the display used for conjunctive search you drew a boundary between the black blobs and the red blobs and another between the square shapes and the round shapes, the boundaries would not coincide. Under these circumstances, and only under these circumstances, we believe that the visual system recruits attention to salvage a preattentively uninterpretable input. It does not follow however that attention is required for all feature integration. Thus in our view focal attention is not required to integrate features. Rather, attention – rather like the cavalry called in to save the threatened settlers – is called in only when normal preattentive processing is in trouble.

In a series of seven experiments studying preattentive vision we (Cohen 1997; Kubovy et al., in press) have collected considerable evidence in favor of our theory. In these studies we have shown that (1) when different feature dimensions form coincident boundaries, these boundaries are detected better than would be expected if the modules were independent; (2) when different feature dimensions form inconsistent boundaries, these boundaries are detected more poorly than would be expected if the modules were independent. These data show that the modules are not preattentively independent, and that they interact intelligently. When appropriate, the outputs of the two modules are synergistic; under other circumstances they are antagonistic. We have surmised that the mutual antagonism that we observed may be the signal that brings attention to bear on solving the perceptual puzzle posed by inconsistent boundaries.

In summary, we have erected an additional barrier between cognition and visual perception, we note that it is not unreasonable to think of a function as important as feature integration as being done preattentively and impenetrably by default.

What is the point of attempting to make a case for cognitive impenetrability of visual perception?

Boris Crassini,[a] Jack Broerse,[b] R. H. Day,[c] Christopher J. Best,[a] and W. A. Sparrow[a]

[a]*School of Psychology, School of Human Movement, Deakin University, Geelong, VIC 3217, Australia;* [b]*School of Psychology, University of Queensland, St. Lucia, Queensland 4072, Australia;* [c]*School of Psychological Science, La Trobe University, Bundoora, VIC 3083, Australia*
buc or cjb or sparrow@deakin.edu.au broerse@psy.uq.edu.au
r.day@latrobe.edu.au www.hbs.deakin.edu.au/psychology
www.uq.edu.au www.latrobe.edu.au

Abstract: We question the usefulness of Pylyshyn's dichotomy between cognitively penetrable and cognitively impenetrable mechanisms as the basis for his distinction between cognition and early vision. This dichotomy is comparable to others that have been proposed in psychology prompting disputes that by their very nature could not be resolved. This fate is inevitable for Pylyshyn's thesis because of its reliance on internal representations and their interpretation. What is more fruitful in relation to this issue is not a difficult dichotomy, but a different look at perception such as proposed by Gibson (1979).

When one begins with the assumption that the function of vision (early or otherwise) is to provide the perceiver with a "structured representation of the 3-D surfaces of objects sufficient to serve as an index into memory," a fundamental problem arises. The assumption of early vision as a representation-producing process involves the further assumption of a representation-interpreting process. As we know, however, this admits the further and thornier problems of infinite regress and solipsism (see Katz 1983). Pylyshyn seems to allow as much in his discussion of what he calls "Bruner's influential theory": If "poor folk" actually see (not metaphorically "see") a coin as larger than the same coin seen by "rich folk" then the misperceptions and consequent maladaptive behaviours that may occur in folk who are in various other deprived states do not bear contemplation. Given this unsatisfactory state of affairs it is appealing to be told that in the visual brain there are mechanisms that eliminate the more fanciful or even dangerous interpretations of representations "provided" by early vision. Even more appealing is the guarantee that these "smart" mechanisms are unable to be influenced by the perceiver's motivations, beliefs, and knowledge; that is, the mechanisms are cognitively impenetrable. However, it is our view that the solution to the problems inherent in theories of perception based on the generation and interpretation of representations provided by Pylyshyn's notion of cognitive impenetrability is problematic on several grounds which we explore in this commentary. We conclude that a better solution to the problem is provided by Gibson's approach to perception.

The origins of Pylyshyn's current version of cognitive impenetrability can be seen in his 1973 publication "What the mind's eye tells the mind's brain: A critique of mental imagery." In this he argued that "the picture metaphor underlying recent theoretical discussions of . . . [the relationship between perception and imagery] . . . is seriously misleading – especially as it suggests that the image is an entity to be perceived" (1973, p. 1). As an alternative Pylyshyn (1973) proposed that mental representations of the world were abstract rather than picture-like, and involved, for example, propositions in the form of pre-compiled subroutines, or rules, which *were* the interpretation. Part of Pylyshyn's argument rested on the claim that there was no necessary connection between a person's introspections and the nature of mental representations. An extension of this disconnection between conscious awareness and mental representation is at the heart of the earlier (Pylyshyn 1980; 1981) and present versions of cognitive impenetrability; that is, not only are mental representations opaque to inspection by the mind's eye, they are immune from influence by the mind.

Pylyshyn's choice of mental imagery as the vehicle to present his notion was not surprising given that imagination can properly be regarded as a quintessentially cognitive activity. More surprising, however, is the conspicuous absence of consideration of imagery in the present elaboration of cognitive impenetrability. This omission cannot be explained on the grounds that the problems at the heart of the imagery debate have been solved (see Tye 1991). In the 1980s psychophysical data initially interpreted as revealing functional equivalence between imagery and perception were "explained away" in terms of, for example, tacit knowledge (Pylyshyn 1980; Richman et al. 1979), or on methodological grounds (Broerse & Crassini 1981; 1984). More recently, claims of functional equivalence have been made based not only on psychophysical data, but also on brain-imaging data (e.g., Kosslyn et al. 1995). The problems, discussed over many years, inherent in interpreting psychophysical data taken to demonstrate equivalence between imagery and perception still remain. Furthermore, as Sarter et al. (1996) point out, interpretations of brain-imaging data are also problematic: although these data are typically taken to reveal something about the role of neural activity in "causing" mental states, Sarter et al. (1996) argue persuasively that such conclusions cannot be drawn, given the nature of brain-imaging data.

It may be that Pylyshyn's failure to engage in further discussion of the relationship between imagery and perception reflects an appreciation of the fact that the current imagery debate, like the debate about imageless thought at the start of this century (see Boring 1950), and the debate about the role of representations in perception, is more about faith than about fact. That is, given their nature, mental states do not afford objective measurement enabling the carrying out of an *experimentis crucis.* In addition, the imagery debate and indeed the debate about the role of representations in perception involve the positing of dichotomies of the type that Hinde (1974) describes as having "bedevilled" the history of behavioural science. In this context he makes special mention of the particularly problematic nature of dichotomies that involve sources of information, or types of behaviour, or underlying processes defined solely in negative terms. Consider, for example, the disputes that have occurred regarding the distinction between motives that are learned versus motives that are not learned (i.e., instincts); and the distinction between development that is based on interaction with the environment versus development in the absence of interaction (i.e., maturation). Hinde's (1974) view is that the disputes which have "plagued" research in these areas are fundamentally insoluble because those involved in the disputes have "adopted different theoretical approaches *and did not see that they differed primarily because they were interested in different questions*" (p. 37, our emphasis). To the list of such dichotomies can be added another: top-down interpretation of representations that is based on perceivers' beliefs, attitudes, and knowledge (i.e., top-down processes that are cognitively penetrable) *versus* top-down interpretation of representations that is not based on these mental phenomena (i.e., top-down processes that are cognitively impenetrable).

Our concern is that Hinde's pessimistic analysis of the earlier dichotomies can be generalised to this one. This is despite the evidence adduced by Pylyshyn to support his claim for "perceptual intelligence" in which "the visual system appears to 'choose' one interpretation over other possible ones, and the choice appears remarkably 'rational' . . . [but it does not involve] cognitive penetration." Pylyshyn's evidence relies heavily on the operation of "smart" mechanisms which through their operation produce "natural constraints on vision" that provide better guarantees of unique interpretation than do, for example, other cognitively-impenetrable top-down constraints such as the Gestalt principles of perceptual organisation. We are somewhat puzzled at Pylyshyn's implication that natural constraints such as Ullman's rigidity constraint are different from (better than?) earlier Gestalt principles. What is Ullman's rendition of a rigid object under rotation if it is not a generalised form of shape constancy (Broerse et al. 1994), and in turn, what is shape constancy if it is not a generalised form of size constancy and size-distance invariance (Koffka 1935)? Furthermore, as Pylyshyn indicates in the qualification he sets out in Note 12, "additional constraints must be brought to bear" to account for more complex cases of motion such as non-rigid biological motion. This qualification hints at the criticism of *post-hoc*-ness and tautology often aimed at the use of Gestalt principles, and their like, as explanatory devices. But more importantly, Pylyshyn's dichotomy between cognitively-penetrable top-down processes and cognitively-impenetrable top-down processes begs the question of what experiment could be designed to differentiate between these processes? Our answer is that any such attempt would inevitably lead to the kinds of dispute that characterise the imagery debate, and is ultimately futile.

The solution is to ask a different question. Rather than beginning with the assumption of a dualism between the environment to be perceived and the perceiver, adding to this a further dualism between the environment and a representation of the environment "inside" the perceiver, and then adding yet another dualism between top-down processes that are or are not cognitively penetrable, a more fruitful approach is to eschew all three dualisms. It is not enough, as Katz (1983) pointed out, to replace pictures in the mind's eye with propositions in the mind's mind; both pictorial and abstract forms of representation are faced with the problem begging an assumed interpreter. In Katz's (1983) words: "The regress can only be avoided if the whole organism is made the interpreter, and representations are given their appropriate place; in the external world, not inside heads" (p. 269). This, of course, is Gibson's approach in which perception is considered an achievement of action rather than a process of interpreting internal representations. The difference between Pylyshyn's and Gibson's approaches to perception is made clear by Gibson's distinction between direct perception in an environment, and indirect perception of pictures (i.e., representations). According to Gibson (1979) the latter "requires two kinds of apprehension" (p. 291); a direct perceiving of the surfaces comprising the picture, and an indirect (interpreted) awareness of what these surfaces depict. But Gibson's distinction is another story.

Constraining the use of constraints

James L. Dannemiller[a] and William Epstein[b]

[a]*Department of Psychology, University of Wisconsin-Madison, Madison, WI 53706;* [b]*Department of Psychology, University of Virginia, Charlottesville, VA 22903.* **psych.wisc.edu/faculty/bio/dannemil/er.html**
jldannem@facstaff.wisc.edu **we7v@faraday.clas.virginia.edu**

Abstract: Pylyshyn uses constraints to solve many of the problems associated with the inverse problem in vision. We are sympathetic to such an approach, and indeed, we think that in many cases constraints allow tractable solutions to otherwise insoluble problems. We argue, however, that Pylyshyn has been too quick to assume that certain perceptual phenomena can be explained by appealing to constraints embodied in the visual machinery. For several more complex perceptual phenomena it is not clear how one proceeds to look for constraints once simple constraints like rigidity have been abandoned.

Rock (1983; 1997) has proposed that the visual system operates in certain situations like an intelligent problem solver bringing to bear knowledge of the world in interpreting ambiguous stimulation. If this were the case, Pylyshyn's thesis that visual perception is discontinuous with cognition would be seriously compromised. Pylyshyn argues instead that many of Rock's parade cases of intelligent perception (e.g., Rock 1983, Ch. 7) should be taken as examples of the power of natural constraints to reduce the solution space. It is supposed that the visual system uses constraints embedded in its operation (mainly by evolution) to arrive at unique and generally correct distal attributions for occurrent states of optical stimulation which are in fact compatible with innumerable candidate distal states. Such circumstances, which are the rule, are difficult chiefly because inverse projection is an ill-posed problem; (in the absence of constraints) the pattern of dynamic 2-D retinal stimulation cannot generally be inverted to recover the unique 3-D situation that produced it.

We are mostly in agreement with Pylyshyn's views on the importance of constraints. However, Pylyshyn has been too quick to argue that cases of what Rock (1983) called "perceptual intelligence" or "problem solving" by the visual system can be explained by appealing to constraints embodied in visual processing. Although we are unconvinced that the results of Rock's experiments constitute evidence of "cognitive penetration" of vision to use Pylyshyn's term, we do think that Rock's experiments show that a noncognitive constraint-based account of complex perceptual phenomena will not be easy to achieve. Consider the case of the perception of structure from motion and the rigidity constraint.

Following Ullman (1979), Pylyshyn observes that with the rigidity constraint in place there is a noncognitive computational solution to the problem of how the visual system generates a veridical representation of the 3-D shape of a rotating object from ambiguous 2-D input. The rigidity assumption allows 2-D input from a rotating rigid object to be used to recover the 3-D shape of the object that produced the input. Perceptually, however, it is not at all clear that the human visual system uses this rigidity constraint. Todd (1984), for example, showed that the curvature of nonrigid objects in motion could be perceived just as accurately as the curvature of rigid objects in motion. Rigidity played little if any role in determining how accurately observers were able to gauge the curvature of rotating and in some cases deforming objects. (Also see Braunstein & Andersen 1984 and Braunstein et al. 1990 for other evidence.) This case illustrates the fact that a constraint shown to be computationally sufficient for solving the inverse problem under restricted conditions in computer vision should not necessarily be invoked to explain similar aspects of human vision.

Next, consider a study by Rock and Smith (1981), which seems to show that the operation of the rigidity constraint can apparently be trumped by the simple addition of a visible rectangular aperture. The perception of a stick rotating rigidly about its center in 3-space (kinetic depth) was created by oscillating a luminous line in the frontoparallel plane behind a hidden rectangular aperture. The concomitant changes in the length and orientation of the line created the vivid impression of a stick moving rigidly in 3-space. Presenting the same 2-D dynamic stimulation with the aperture now visible, abolished the 3-D percept, replacing it with the percept of a line oscillating in orientation in the frontoparallel plane behind a rectangular aperture. The visual system appears to be operating with the rigidity constraint when the aperture is not visible, but this constraint is not used to interpret the same stimulation when the aperture is visible.

Pylyshyn proposes to set aside such apparent evidence of intelligent problem solving by the visual system and instead argues that the results can be assimilated to the principle that the visual system embodies natural constraints to resolve such ambiguities. In this case, it is not clear what the constraints are (occlusion?) that lead to the trumping of the rigidity constraint, but we are assured that such cases present no problems for this approach to explaining visual perception. But such cases should not be dismissed so easily. The dynamic stimulation that elicits perception of kinetic depth when the aperture is hidden is identical to the stimulation that elicits perception of oscillating 2-D motion when the aperture is visible. If rigidity is such a powerful constraint, and if the visual system uses this constraint to interpret such input (although see above), then what is it about the presence of an apparent aperture that nullifies the use of this constraint in interpreting the stimulation?

The findings reported by Rock and Smith (1981), when added to the negative results of the direct tests of the rigidity constraint by Todd (1984) and Braunstein et al. (1990) strongly suggest that in his treatment of the perception of structure from motion, Pylyshyn has rushed ahead of the empirical state of affairs. To be sure, the foregoing is not an argument against the constraint style of explanation. Only the heavy reliance on the rigidity constraint in the analysis of perception from optical motion configurations is brought under question. If we are wedded to the constraint-based explanation, we will need to move on from exclusive reliance on rigidity. The search for constraints is a difficult endeavor. Inasmuch as computational sufficiency is no warranty that such constraints actually play a role in human vision, the conclusion of the search cannot come in advance of relevant findings from the perception laboratory. The task is even more challenging than this. When multiple constraints may apply in a given situation (e.g., rigidity and occlusion), one must answer the question of how these constraints are sequenced or combined in the computation, and if they compete, how does the visual system "decide" which one should be allowed to determine the solution to the problem? These are questions that Pylyshyn was free to ignore by assimilating cases such as the one in Rock and Smith (1981) into others in which the role of natural constraints is more readily apparent.

We do not have answers to the questions posed above. We do think, however, that such questions will have to be confronted if we are to arrive at explanations of visual perception that do justice to the complexity of the phenomena that they are intended to explain. Arguing that all such cases can be treated safely as examples of the operation of natural constraints in vision without actually sketching out what those constraints are and how they might operate only puts off the hard work necessary to understand the promise and the limitations of such an approach.

ACKNOWLEDGMENTS
The writing of this manuscript was supported by NICHD R01 HD32927 and a University of Wisconsin Vilas Foundation Fellowship to JLD. Correspondence concerning this article should be addressed to James L. Dannemiller, Room 515 Waisman Center, University of Wisconsin-Madison, 1500 Highland Avenue, Madison, WI 53705-2280. Electronic mail may be sent via the Internet to jldannem@facstaff.wisc.edu.

Better theories are needed to distinguish perception from cognition

Michael R. W. Dawson and C. Darren Piercey
Biological Computation Project, University of Alberta, Edmonton, Alberta, Canada T6G 2E9. **mike@bcp.psych.ualberta.ca**
www.bcp.ualberta.ca/

Abstract: Pylyshyn argues that many of the methods used to study perception are too coarse to detect the distinction between perceptual and cognitive processing. We suggest that the reason for this is that the theories used to guide research in perception are at fault. More powerful theories – for instance, computer simulations – will be required to identify where perception ends and where cognition begins.

Pylyshyn's target article is the latest in a long discussion of the role of cognition in perception (Dretske 1969; 1981; Epstein 1977; Fodor 1983; Fodor & Pylyshyn 1981; Gogel & Tietz 1992; Hamlyn 1957; Heil 1983; Mace 1977; Michaels & Carello 1981; Pitcher 1971; Pylyshyn 1980; 1984; Schyns et al. 1998; Turvey et al. 1981; Ullman 1980; Wagemans 1988; Wright & Dawson 1994). Pylyshyn now adds a crucial idea to this extensive debate: the methods of experimental psychology are too coarse to explore the details of perception or cognition, or to even identify the boundary between the two. Pylyshyn's take-home message is that the experimental study of perception must adopt radically new methodologies before it can attempt to disentangle "data-driven" from "top-down" processing.

It is important to note that the methodological problem facing the study of perception does not arise because researchers have only a small handful of techniques used to collect data. Indeed, perception is studied with a wide range of methods capable of making subtle empirical distinctions (see, for example, sect. 4 of the target article). The problem with perception is that it is a discipline that is data rich, but theory poor. Unfortunately, these poor theories do not appropriately direct available research methods to collect the data required to flesh out the issues that the target article addresses.

To illustrate this, let us consider the study of apparent motion, a topic introduced in section 5.2 of the target article. Apparent motion occurs when two static images are presented successively, in different locations, under appropriate timing conditions. Instead of generating the experience of stationary flicker, the visual system generates the illusion of movement from the first image to the second. This apparent motion, known as *beta motion,* is extremely difficult to distinguish from real motion (Kolers 1972).

The discovery of beta motion led many researchers to believe that it was the result of high-level problem solving processes, because the appearance of objects as they "moved" had to be inferred by the visual system. New Look researchers tested this hypothesis by manipulating the "familiarity" or "meaningfulness" of displays (Jones & Bruner 1954). In general, familiar motion (e.g., a car "moving" over a bridge) was perceived more easily than unfamiliar motion (e.g., an oval "moving" over an arc). Similar results were found in more focused studies that manipulated depth cues to make motion more or less meaningful (Attneave 1974; Corbin 1942).

A problem with this whole approach was its guidance by a vague theory of motion perception (motion as the product of inference). As a result, these New Look studies examined a limited range of dependent measures. Modern research has been guided by a more sophisticated theory of apparent motion that recognizes that many different information processing problems have to be solved on the way to experiencing apparent motion (Petersik 1989; Wright & Dawson 1994). The results of the New Look studies depend a great deal on what property of experienced motion is being measured. If motion quality (i.e., the degree to which it is beta motion) is measured, then the New Look results hold. However, if lower-level properties of motion are measured (i.e., did an object move left or right, when movement in one direction is more "meaningful" than another), then no New Look effects are found at all. This has been shown for manipulations of apparent depth (Dawson & Wright 1989; Wright et al. 1987) and of shape (Dawson 1989). In other words, had New Look theories been rich enough to consider multiple roles of motion perception, and multiple stages of motion detection, the data collected would have shown that although some aspects of motion perception are cognitively penetrable, others are not.

Similar problems can be identified in the study of the components of apparent motion processing. For instance, the visual system must solve the *motion correspondence problem* before apparent motion can be experienced (Dawson 1991; 1998; Ullman 1979). A solution to this problem requires that the assignment of identity matches between elements in the first and second image of an apparent motion display; these matches indicate "what went where." For any apparent motion display, a number of different solutions to this problem are possible, but the visual system will only generate one. The main question of interest concerns the principles that are used to exclude possible motion correspondence matches.

Unfortunately, the search for such principles – particularly when a possible role for top-down effects exists – is not terribly constrained by theories about motion correspondence processing. For instance, it has been suggested that the visual system applies a variety of "rules of thumb" from "what is in effect a bag of tricks" (Ramachandran & Anstis 1986). These rules are studied by "watching the visual system" as it solves the correspondence problem (Sekuler et al. 1990). In essence this is done by manipulating a stimulus property to see if it affects the correspondence matches that are assigned. Unfortunately, if a stimulus property is found that affects the assigned matches, then there is no pressing needed to carefully integrate it into a theory of correspondence processing, or even to consider whether the result is indicative of data-driven or top-down processing. This is because the stimulus property can simply be described as another rule to be added, uncritically, into the bag of tricks.

In our view, a more careful and more productive approach to studying the principles of motion correspondence processing is to start with a theory that is rich enough and precise enough to be formalized. This requires a computational analysis of motion correspondence processing which will identify the natural constraints that could logically be involved (see sect. 5 of the target article). These constraints can then be incorporated into a working computer simulation which can be used to explore the necessity and sufficiency of a variety of rules of thumb. For example, various studies of motion correspondence processing have revealed that a number of specialized principles (e.g., least energy transformations) that have been applied to particular displays (e.g., the Ternus configuration) actually emerge from a smaller set of interacting natural constraints (Dawson 1991; Dawson et al. 1994; Dawson & Wright 1994; Sekuler et al. 1990).

What does this imply for the study of the continuity between perception and cognition? As the target article implies, most current theories in perception are not detailed enough to address this basic issue. Our methodological suggestion is to adopt the approach of synthetic psychology (Braitenberg 1984), and build computer simulations that instantiate detailed theories of perceptual processing. The question to ask with such models is simply this: How much can they do without cognition, and how much of this would appear to be cognitive to someone who viewed the model's behavior, but not its internal workings? The history of the natural computation approach in vision (Marr 1982) provides ample proof that this kind of approach can reveal a surprisingly complex set of perceptual phenomena that do not require cognitive processing. With such a theory in hand, one is much better able to direct experimental methods to find evidence, in human subjects, of where (and why) perception and cognition are distinct.

The cognitive impenetrability hypothesis: Doomsday for the unity of the cognitive neurosciences?

Birgitta Dresp

Laboratoire de Psychophysique Sensorielle, Université Louis Pasteur, 67000 Strasbourg, France. **bee@currif.u-strasbg.fr**

Abstract: The heuristic value of Pylyshyn's cognitive impenetrability theory is questioned in this commentary, mainly because, as it stands, the key argument cannot be challenged empirically. Pylyshyn requires unambiguous evidence for an effect of cognitive states on early perceptual mechanisms, which is impossible to provide because we can only infer what might happen at these earlier levels of processing on the basis of evidence collected at the post-perceptual stage. Furthermore, the theory that early visual processes cannot be modified by cognitive states implies that it is totally pointless to try to investigate interactions between consciousness and neurosensory processes.

Pylyshyn's target article discusses why many of the so-called top-down effects in psychophysical or perceptual tasks are to be related to changes occuring at post-perceptual stages of processing rather than at the level of basic perceptual mechanisms. The argument is that this is true for most experiments designed to bring top-down effects on early visual mechanisms to the fore. Pylyshyn's analysis provokes the question "What do we actually measure in psychophysics when we believe we are tapping early perceptual processes or sensory mechanisms?" Clearly, what we do measure is variations in behavior. On the basis of these variations, we hope to infer the nature and operating characteristics of early perceptual mechanisms and processes.

According to Pylyshyn, a cognitive penetration hypothesis is conformed only when a direct effect of prior knowledge, belief, or some kind of memory-based stimulus analysis on a perceptual mechanism or process can be unambiguously demonstrated. My objection to this rationale is that it sets an impossible criterion. How could we ever find unambiguous evidence of an influence of cognition on the visual or perceptual system, given that all we can possibly measure in psychophysics or experimental psychology is a behavioral correlate that we believe is a valid indicator of some underlying sensory or perceptual process? The post-perceptual level is the only one to which we have access, not only in psychophysics (see also the recent book by Baird, 1997), but in the behavioral sciences in general. This leads inevitably to the conclusion that Pylyshyn's cognitive impenetrability argument cannot be proven wrong.

I think, and many colleague psychophysicists would certainly agree, that the heuristic value of any theory that cannot be proven wrong is doubtful. What are we to do with a theory that cannot be challenged? If the author is right, his target article not only settles whether there is any influence of higher levels of consciousness on basic perceptual processes and sensations, it also leads to the conclusion that the epistemological unity of the neurosciences and the cognitive sciences is doomed. It further implies that there is no scientific answer to fundamental questions about how mental operations and sensory functions interact.

Does Pylyshyn want to suggest that neuroscientists should stick to the molecular processes and hardwired visual mechanisms while cognitive scientists take care of the perceptual processes? Speaking as a visual psychophysicist here, I should like to insist that psychophysical theory has always been based on the idea of a functional unity of brain mechanisms and representational processes in general. Why should the case of visual function be an exception? In my opinion, the monist view is a necessary working theory in the behavioral and brain sciences because any other view runs the danger of closing doors once and forever.

No reconstruction, no impenetrability (at least not much)

Shimon Edelman
School of Cognitive and Computing Sciences, University of Sussex at Brighton, Falmer BN1 9QH, United Kingdom. **shimone@cogs.susx.ac.uk**
www.cogs.susx.ac.uk/users/shimone

Abstract: Two of the premises of Pylyshyn's target article – surface reconstruction as the goal of early vision and inaccessibility of intermediate stages in the process presumably leading to such reconstruction – are questioned and found wanting.

According to Pylyshyn, a certain chunk of visual machinery, which he calls the early vision module, is not accessible to "cognitive" intervention; at most, only attentional control over the locus of application of that module can be exercised. This view is part of a general treatment of the mind as modular (Fodor 1983; see also multiple book review of Fodor's "The Modularity of Mind" *BBS* 8(1) 1985), seeking to divide it into various cognitive faculties (presumably, each of these is to be conquered by theorists later, when the genies are securely confined to their bottles). In vision research, this amounts to postulating that all the visual processing between the retinal image and the mythical 2.5-D Sketch (Marr 1982) is inaccessible to the rest of cognition.

For this claim to be of interest, the cognitively impenetrable early vision module must be charged with a substantial portion of the visual processing burden. If all the vision module was doing was moving a few small steps away from the raw image (say, detecting abrupt intensity transitions or "edges" in the stimulus image), the vision genie would be too puny to justify imprisonment.

The crucial issue, which the target article commendably puts squarely on the table in section 7, is, accordingly this: What is the output of the visual system? If there was ever a 64 thousand dollar question in vision research, this is surely it: if you know the answer, you know the nature of the internal representation of the visual world. Pylyshyn offers one possible answer: in section 7.2, he claims that "evidence favors the view that some depth-encoded surface representation of the layout [of the scene] is present in the output of the early-vision system."

Such surface representation must be both explicit and obligatory ("automatic"), as per the 2.5-D Sketch doctrine (Marr 1982), if the modularity thesis is to be worth defending in the context of vision. In other words, the system must maintain depth representation of all the visible surfaces at all times – or else suffer the consequences of its inability to salvage the intermediate representations, to which, alas, the mind has no access (according to Pylyshyn).

In fact, neither of these alternatives corresponds fully to psychophysical reality. On the one hand, the postulate of obligatory surface reconstruction is undermined (1) by the scarcity of empirical support, (2) by the ability of current models of recognition and categorization to do without surface representation as such, and (3) by the indications that arbitrarily sketchy representations are passed off routinely as the real thing by the visual system. On the other hand, cognitive control can easily be exerted over the system's response to seemingly arbitrary combinations of very low-level features such as individual dots and lines. Let us consider each of these issues in turn.

Although some experiments testing the idea of explicit surface reconstruction have been carried out, the interpretation of their results is debatable. Studies that actually claim to have demonstrated that under certain circumstances surfaces are represented explicitly tend to rely on the subject's report of the perceived surface (Treue et al. 1995), a technique prone to what Dennett (1991) terms an internal revision. For example, it is conceivable that the system labels a portion of the visual field as "that surface" while actually marking only a tiny minority of the relevant pixels – perhaps those near the experimenter's probe – as belonging to it (see Dennett 1991, p. 344).[1] In 1992, Nakayama and Shimojo summarized their psychophysical study of surface interpolation as follows: "We have suggested that sampled images can be associated with surfaces, not mentioning the representation of surfaces themselves. . . . Because we have no specific data to address this issue directly, we can only speculate." (pp. 1362–63). At present, the issue of explicit representation of surfaces is still undecided: the survey of Pessoa et al. (1998) ends with a series of six conclusions that by no means settle the debate one way or the other.[2]

A serious problem facing the hypothesis of surface reconstruction stems from the indications that the perceptual functions thought by Marr to be the culmination and the very purpose of reconstruction – object recognition and categorization – need not actually involve anything like the 2.5-D Sketch (Bülthoff et al. 1995; Edelman & Duvdevani-Bar 1997; Mel 1997). This begs the question of why, in perceptual scenarios that do not call for an interaction with object surfaces, the visual system should bother with the computationally intensive and error-prone reconstruction in the first place. Indeed, mounting evidence (some of it cited in the target article) suggests that little of the scene structure beyond its general layout (described in terms of object categories, not raw surfaces) may be retained even in short-term memory (Blackmore et al. 1995; O'Regan 1992; Simons 1996).

When the uncertainty of actual surface reconstruction is combined with Pylyshyn's belief that anything less than a reconstructed surface is not accessible to cognition,[3] common perceptual phenomena that are the mainstay of classical visual psychophysics become great puzzles. For instance, one wonders why it is so easy to have subjects respond to simple stimuli consisting of dots or lines – entities that, according to Pylyshyn's claim, are locked inside the early vision module (in Marr's terminology, these would be part of the Primal Sketch, a hypothetical stage preceding the surface reconstruction in the 2.5-D sketch; an example of a task involving such stimuli is vernier discrimination, mentioned in section 6.3 of the target article). Unless the "cognitively impenetrable" early vision module is assigned a more substantial role than chasing around a few points in the visual field, Pylyshyn's thesis loses much of its compelling quality.

NOTES

[1] Experiments that use a depth probe to assess the subject's percept of the surface at a chosen spot in the visual field (Bülthoff & Mallot 1988) cannot help interfering with the very process they aspire to measure. In these experiments, the visual system is effectively called upon to produce on demand an estimate of the perceived depth *at the probe;* the possibility that elsewhere in the visual field there is no representation of anything like surface depth or orientation is not ruled out.

[2] In the target article, Pylyshyn cites Pessoa et al., albeit in a different context.

[3] "There is no evidence that . . . outputs of specialized subprocesses are available to cognition in the normal course of perception" (sect. 7.2; "subprocesses" are the intermediate steps that presumably lead to the computation of the surface layout).

The cognitive impenetrability of visual perception: Old wine in a new bottle

Howard Egeth
Department of Psychology, Johns Hopkins University, Baltimore, MD 21218.
egeth@jhu.edu **www.psy.jhu.edu/~egeth/**

Abstract: Pylyshyn's argument is very similar to one made in the 1960s to the effect that vision may be influenced by spatial selective attention being directed to distinctive stimulus features, but not by mental set for meaning or membership in an ill-defined category. More recent work points to a special role for spatial attention in determining the contents of perception.

I find myself in agreement with at least one key point of Pylyshyn's analysis of the impenetrability of vision. However, in the spirit of full disclosure, it is only fair to add that this may be due in part to the fact that my colleagues and I made a very similar argument approximately 30 years ago. In this commentary, I supplement Pylyshyn's review of the literature by describing the critical studies that enabled us to conclude that vision could be affected by selective attention to distinctive features, but not by mental set for meaning or membership in an ill-defined category. I then describe recent work that clarifies the nature of attention to features.

Early research. Although the relevant literature can be traced back at least as far as Külpe (1904), we can leap a half century and go directly to the seminal experiment of Lawrence and Coles (1954; for reviews of the literature see Egeth 1967; Egeth & Bevan 1973; Haber 1966). Lawrence and Coles were interested in determining whether preparatory set could influence perception rather than just memory or response processes. Stimuli were pictures of everyday objects that were displayed tachistoscopically; the subject's task was to identify what was shown by selecting one of four verbal labels provided by the experimenter. Following Chapman (1932), they reasoned that perception must be more detailed than memory. Let us consider the possible consequences of presenting the alternatives before or after the stimulus presentation. If the verbal labels denote objects that are quite different from one another, then even the sketchy memory trace of the stimulus should contain enough information for the subject to choose among the four alternatives. Thus, for distinctive labels it should not matter much whether they are presented before or after the stimulus. In contrast, if the alternatives denote very similar objects, the memory trace may well lack information crucial to a correct identification if these alternatives are presented after the stimulus. If the distinctive alternatives are presented before the picture – and if there is a mechanism of perceptual selectivity – then it may be possible for this mechanism to be tuned selectively to the pictorial details that will permit discrimination between the similar alternatives. Presumably, such a mechanism could only be helpful if it were readied in advance of the tachistoscopic presentation of the stimulus. To summarize, the hypothesis of perceptual selectivity predicts an interaction between similarity of the alternatives to one another and time of presentation of the alternatives (Before vs. After).

In the Lawrence and Coles experiment the critical interaction was not significant. Even the main effect of Before vs. After instructions was not significant. The authors concluded that selectivity at the level of perception had not been demonstrated. This conclusion seems entirely reasonable. On reflection, just how *could* verbal labels have influenced perception, inasmuch as these labels only indicate broad classes of objects that may look quite different from one another? Egeth and Smith (1967) performed a conceptual replication of the Lawrence and Coles experiment in which both the stimuli and the alternatives were pictorial. For example, instead of presenting the four similar alternatives of church, school, house, and cabin, we presented a picture of a specific exemplar of each category to the subject; the stimulus was one of these four alternatives. In this study, the crucial interaction was significant, thus implying selectivity at the level of perception.

Pachella (1975) later confirmed the essential findings of both the Lawrence and Coles and Egeth and Smith experiments using signal detection methodology. In an interesting second study, subjects were familiarized with the verbal labels and pictures 24 hours before the test session. During the familiarization session, for each of the 200 stimuli, subjects were first shown a verbal label for 3 sec. followed by the corresponding picture for 10 sec. During the test session the alternatives were verbal labels only, not pictures. They were presented either before or after. In this experiment, alternatives shown before the stimuli yielded significantly better performance than those shown after, which indicates an effect of perceptual selectivity.

With pictorial alternatives (in the Egeth & Smith experiment and in Pachella's Exp. 1), or even with verbal alternatives when subjects know precisely what objects they denote (as in Pachella's Exp. 2), subjects in a Before condition have an opportunity to focus attention on critical features that can serve to identify the stimulus. The Before conditions permit the tuning of relevant perceptual analyzers. In other words, by using specific pictorial alternatives rather than generic verbal labels, we converted the mental set experiment of Lawrence and Coles into a selective attention experiment. Pylyshyn's argument seems to be entirely consistent with our results and theoretical conclusions.

Recent research. In section 4.3, Pylyshyn writes, "A number of physical properties have been shown to serve as the basis for focused attention." He has in mind here such properties as color, shape, motion, stereo disparity, and spatial location. However, not all properties are created equal. It now appears that if a subject is instructed to attend to, say, the red items in a display, this neither enhances the sensory quality of the red items, nor causes non-red items to be excluded from subsequent processing (e.g., Moore & Egeth 1998; Shih & Sperling 1996; see also Prinzmetal et al. 1998, for a discussion of the minimal effects of attention on the appearance of items). Attention does affect the *priority* of the red items; attending to red may mean simply that spatial attention is directed to those items earlier than they would be without the instruction. To put it differently, when we consider the allocation of attention, there may be a special role for spatial location. What this suggests about the early demonstrations that set could affect perception as long as the alternatives were pictorial (Egeth & Smith 1967; Pachella 1975), is that presenting pictorial alternatives before the stimulus permitted the subject to decide exactly *where* to look when the stimulus made its brief appearance.

ACKNOWLEDGMENT
The writing of this commentary was supported in part by grant MH57388 from the National Institute of Mental Health.

Perception and information processing

Angus Gellatly
Department of Psychology, University of Keele, Staffordshire, ST5 5BG, England. **psa31@keele.ac.uk**

Abstract: Perception and cognition can be understood either as conscious experience, thought, and behaviour or as bodily functions executed at the level of information processing. Whether or not they are cognitively penetrable depends on the level referred to. Selective attention is the mechanism by which cognition affects perception, theory influences observation and observational reports, culture biases experience, and current knowledge determines what inferences are made. *Seeing* must be distinguished from *seeing as.*

Pylyshyn's target article makes us think carefully about the perception-cognition relationship. The issues are of such complexity that he is able to cite Bruner (1957) arguing both that values and needs organise perception and also that perception lacks docility and is resistant to cognitive influences. Unsurprisingly, maintaining consistency on this treacherous ground proves equally difficult for Pylyshyn (and this commentator). Sometimes he focuses on the issue of cognitive penetration of perception, sometimes on the continuity of perception and cognition, which may not be the same thing. He variously refers to "perception," "perceptual representations," "vision," "early vision," and "visual apprehension," sometimes as synonyms, sometimes seemingly as contrasts. He attributes to Bruner the view that values and needs determine perception "down to the lowest level of the visual system," but the phrase is not a quotation and one may wonder whether Bruner really intended to propose top-down influence on, say, receptor functioning. A brief commentary can hardly hope to resolve these difficulties, but the following points seem relevant to the target article:

1. To claim there are functions of the visual system that are not cognitively penetrable is not different from claiming that parts of the brain involved in autonomic control execute nonpenetrable functions. Some of these autonomic functions – control of pupil dilation, regulation of diurnal rhythms, melatonin secretion in response to light – occur within the visual system broadly understood as the processing of retinal signals (Stoerig 1996). Pylyshyn might wish to restrict consideration to the geniculo-striate pathway, its projections, and the functions they compute. However, it is increasingly clear, not least from the study of "blindsight," that other visual pathways have an input to those functions (Cowey & Stoerig 1991), and that the geniculo-striate system cannot be treated as an isolated unit. [See also Campion, Latto & Smith: "Is Blindsight an Effect of Scattered Light, Spared Cortex, and Near-Threshold Vision?" *BBS* 6(3) 1983.] The general point is that the visual system is an organ of the body, and bodily functions tend not to be cognitively penetrable. (Though blood pressure can be raised by selective recall of past outrages; stress, mediated by beliefs ["The boss dislikes me," "The company is downsizing"], may influence physiological functioning; and rats can learn/be conditioned to suppress their own immune response.)

2. What *are* cognitively penetrable are voluntary behaviours and thoughts (internal behaviour) and, consequently, conscious experience. What you expose your sense organs to determines conscious experience, and Pylyshyn is surely right to emphasise the role of selective attention in such phenomena as perceptual learning, reversal of ambiguous figures, and so on. Selective attention is also the mechanism that mediates cultural influences on perceptual experience, including the experience of colour (Gellatly 1995).

3. Pylyshyn (sect. 1) criticises philosophers of science, such as Kuhn, for arguing that "every observation was contaminated by theory." More often, however, they argued for theory laden "facts" and observational reports rather than empirical observations. Seeing that the needle is pointing to 2 is an observation; stating that "the voltmeter is reading 2 volts" is an observational report. Of course, the siting of the boundary between observation and report is always open to dispute. This is a central concern of Pylyshyn's article, and one whose attendant difficulties were familiar to introspective psychologists under the title of "the stimulus error." However, observation is also theory laden in the sense that allocation of attention is theory driven; we make those observations thought to be theoretically significant (and fundable).

4. Pylyshyn (sect. 2) disputes that perception and cognition are continuous. It needs to be clear whether we are talking about perception and cognition understood as conscious experience, thought or behaviour on the one hand; or, on the other hand, as information processing operations (Van der Heijden 1992). Pylyshyn gives as examples of the cognitively penetrable inference making (Note 3), solving crossword puzzles, assigning referents to pronouns, reading comprehension, and attributing a cause to a noise (sect. 1.1, para. 1). This is a mixed bag. Examining individual examples may show that putative penetrability arises from attention and response effects of the kind Pylyshyn (sect. 4) notes in relation to vision. For example:

5. Pylyshyn states that "The paradigm case of such a [rational] process is *inference*" (Note 3). But "inference" is *not* a process; it is a classificatory term like "search" (Gellatly 1988; 1992). Like searching, inference making can be carried out in different ways using a variety of information processing operations (Johnson-Laird 1983). Both searching and inferring can be rational in that exposure to new information may change the way one searches or the inferences one makes. But this is not cognitive penetration of information processing operations, it is a change in the use of automatic operations. Pylyshyn claims that "Embodying a natural constraint is different from drawing an inference. . . . Observers cannot tell you the principles that enable them to calculate constancies and lightness and shape from shading." But then Lewis Carroll (1895) long ago demonstrated the impossibility of stating a principle by which we get from premises to conclusion. We simply "see" it that way.

6. Pylyshyn (sect. 3.3) argues from agnosia that recognition occurs later than early vision. By the same token, it could be argued from achromatopsia that colour perception occurs subsequent to early vision, an inference most would wish to resist. Of course, it often seems that recognition occurs *after* perception, as in the Dalmation dog puzzle picture, and Pylyshyn (sect. 6.1) points out the role of attention in seeing the picture *as* a dog. Yet we can never escape *seeing as.* Prior to seeing the stimulus *as* a dog we were seeing it *as* a puzzle picture, or *as* black marks on white paper, or *as* part of the paraphernalia in a psychologist's laboratory. The prosopagnosic (sect. 3.3) who does not recognise his wife's face still sees it *as* a face, and may even retain implicit recognition of her identity (Bruyer 1991). Only explicit recognition of personal identity is lost.

7. With regard to the vision of experts, Pylyshyn (sect. 6.2) states: "rather than having learned to see the board differently, chess masters have developed a very large repertoire . . . of patterns which they can use to classify . . . . such a classification process is post-perceptual." Yet later he adds: "Compiling complex new transducers is a process by which post-perceptual processing can become part of perception. If the resulting process is cognitively impenetrable . . . it becomes part of the visual system. . . . How such processes become "compiled" into the visual system remains unknown" (sect. 6.4, last para.). This seems to be precisely the problem with drawing hard and fast lines as to what is and is not cognitively penetrable.

Is haptic perception continuous with cognition?

Edouard Gentaz[a] and Yves Rossetti[b]

[a]*Faculty of Psychology and Educational Sciences, University of Geneva, 1227 Carouge, Switzerland;* [b]*Space and Action-I.N.S.E.R.M.U.94, 69500 Bron, France.* **gentaz@fapse.unige.ch** **rossetti@lyon151.inserm.fr**

Abstract: A further step in Pylyshyn's discontinuity thesis is to examine the penetrability of haptic (tactual-kinesthetic) perception. The study of the perception of orientation and the "oblique effect" (lower performance in oblique orientations than in vertical–horizontal orientations) in the visual and haptic modalities allows this question to be discussed. We suggest that part of the visual process generating the visual oblique effect is cognitively impenetrable, whereas all haptic processes generating the haptic oblique effect are cognitively penetrable.

In his target article, Pylyshyn proposes that an important part of visual perception is cognitively impenetrable. This idea stands in sharp contrast with the growing popularity of top-down control of visual processes according to which "there is no such thing as immaculate perception" (Kosslyn & Sussman 1995). One important function that may provide examples of such impenetrable perception is goal directed action, because many features of visuomotor or other forms of sensorimotor control are clearly distinguishable from higher level cognitive functions, in both the visual and the somesthetic modalities (Rossetti 1998; 1999). A clear distinction can be observed between perception involved in the motor control of simple goal-directed actions and higher level perception required for cognitive representations and/or memorizing stimuli (Rossetti 1998; Rossetti & Procyk 1997). The former seem to provide examples of "immaculate perception" whereas the latter do not.

A further step in this discontinuity thesis would be to study the penetrability of the haptic (tactual-kinesthetic) perception. The aim of this commentary is to discuss whether there are in haptic perception, too, some perceptual processes that are separate and distinct from cognition. The existence of modular haptic processes would be a positive demonstration that part of haptic perception may be cognitively impenetrable, and would generalize the discontinuity thesis. The interest in examining the penetrability of the haptic perception stems from the fact that a lot of research, in comparing visual perception and haptic perception, has shown that haptic processes depend on cognition more than the visual processes do (for reviews, see Hatwell 1978; 1986; 1990; 1994; Millar 1994; Streri 1993). Visual information is imposed upon the brain (it is impossible to decide not to see), whereas most haptic information is extracted by exploratory movements organized by the cognitive system (it is possible to decide not to touch). Hence it is not surprising that an important portion of haptic perception is cognitively penetrable. However, if we observe that some haptic perceptual processes are cognitively impenetrable and we hypothesize that similar spatial processing is at work in haptic and vision, we can predict that at least part of visual perception is likewise cognitively impenetrable. By contrast, if we observe that all haptic processes are cognitively penetrable, this can only suggest that the nature of the links between perception and cognition varies according to perceptual modality.

Comparing the perception of orientation in the visual and haptic systems allows us to discuss this question. In the visual system, the vertical and horizontal are known to be perceived more accurately than oblique orientations. This anisotropy, which Appelle (1972) called the "oblique effect," has been found in a wide variety of perceptual tasks (Appelle 1972; Essock 1980; Howard 1982). The processes that generate the visual oblique effect seem to be multi-componential and occur at different levels of processing (for recent reviews, see Heeley et al. 1997; Saarinen & Levi 1995). The visual oblique effect is consistent with the discontinuity thesis presented in Pylyshyn's target article. Indeed, part of the visual process that generates the oblique effect seems to be modular because it is not possible to suppress the occurrence of the visual oblique effect; it is only possible to modify its magnitude (i.e., the magnitude of the performance difference between vertical-horizontal orientations and oblique orientations). For example, Heeley and Buchanan-Smith (1992) compared orientation discrimination thresholds (for vertical, horizontal, 45° and 135° oblique orientations) obtained with simultaneous or successive presentation of the stimulus. Simultaneous presentation improved the discrimination of four orientations and reduced the magnitude of the visual oblique effect (though it remained significant). On the other hand, Vogels and Orban (1985) examined whether the visual oblique effect was influenced by the practice of 5,000 trials on an identification task. The effect of practice improved discrimination thresholds for all orientations. The effect was much stronger when the subjects were trained in orientation discrimination at 45°–135° oblique orientations than at vertical–horizontal orientations. (Despite decreasing magnitude of the oblique effect, it remained significant after practice.) Taken together, these findings show that part of the visual process that generates the oblique effect is cognitively impenetrable.

The experiments on haptic perception of orientation and the oblique effect suggest that the haptic processes generating the oblique effect are multi-componential and occur at different levels of processing (for a review, see Gentaz, in press). However, data seem to be more consistent with the continuity thesis. Indeed, the haptic processes that generate the oblique effect do not seem to be modular because it is possible to suppress the occurrence of the haptic oblique effect by changing the stimulus encoding conditions (Appelle & Countryman 1986; Gentaz & Hatwell 1995; 1996; 1998). Blindfolded subjects were asked to explore a rod haptically and to reproduce its orientation. Gentaz and Hatwell (1996) showed that the "gravitational cues" provided by the antigravitational forces developed by the arm-hand system during the scanning were involved in the haptic oblique effect. These cues probably reinforced the vertical orientation relative to the oblique orientations. In the three planes of space (horizontal, frontal, and sagittal), the oblique effect was present when the forearm was free (unsupported) in the air, and antigravitational forces were elicited during scanning. By contrast, the oblique effect was absent in the horizontal plane when the forearm was supported while remaining in physical contact with the surface of the device. In this condition, minimal antigravitational forces were required to move the arm and hand. In addition, the oblique effect was lower in the frontal and sagittal planes when the forearm was lightened by a device consisting of a set of weights and pulleys. In the two latter conditions, the antigravitational forces were reduced during scanning. These observations showed that the occurrence of the haptic oblique effect was influenced by the encoding conditions of manual exploration because these conditions changed the gravity constraints of scanning. Unfortunately, the effect of practice on the haptic oblique effect has not been yet studied. However, these results may suggest that all haptic processes that generate the oblique effect are cognitively penetrable.

In sum, we propose that part of the visual process that generates the visual oblique effect is cognitively impenetrable whereas all haptic processes that generate the oblique effect are cognitively penetrable. Taken together, this suggests that the links between perception and cognition may depend on the perceptual modality: visual perception is discontinuous with cognition whereas haptic perception is continuous with cognition.

Neurophysiology indicates cognitive penetration of the visual system

Alexander Grunewald

Division of Biology, California Institute of Technology, Pasadena, CA 91125.
alex@caltech.edu vis.caltech.edu/~alex

Abstract: Short-term memory, nonattentional task effects and nonspatial extraretinal representations in the visual system are signs of cognitive penetration. All of these have been found physiologically, arguing against the cognitive impenetrability of vision as a whole. Instead, parallel subcircuits in the brain, each subserving a different competency including sensory and cognitive (and in some cases motor) aspects, may have cognitively impenetrable components.

Pylyshyn argues that the visual system is encapsulated from other cognitive processing. Only attention, according to him, is able to affect the early visual system. This is a very important thesis because it provides strong support for the study of vision in isolation. Without at least some notion of cognitive impenetrability, studying vision without considering all other mental processing becomes meaningless. In addition, if there is a lack of continuity between vision and cognition, there is hope that other modules in the brain are also discontinuous, legitimizing methods that may pick up on these discontinuities, such as brain imaging. Pylyshyn draws upon results from physiology, psychophysics, and computer vision to support his claim. In this commentary I wish to comment on the physiological evidence.

Pylyshyn states that cognitive penetration means that a subject's knowledge or beliefs can affect the visual percept (sect. 1.1). But where do we situate short-term memory? The decision to keep something in short-term memory is cognitive, as is the content thereof; hence short-term memories do contain a cognitive component. According to Pylyshyn's thesis, this would imply that 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 presemantically, 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 nonsemantic 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, Kolinsky et al. 1994). These contrasting findings suggest that a critical variable is whether or not the observers' attentional control is required, which is coherent with Pylyshyn's discontinuity hypothesis.

One aspect of the impact of formal instruction provided at school is like the case of experts such as chicken sexers or art authenticators: it is the specificity of training or instruction. Indeed, illiterates and ex-illiterates were equally poor in many of our visual experiments, as well as in explicit nonverbal auditory analysis (Morais et al. 1986), while they dramatically differed in metalinguistic studies. This may be explained by the fact that ex-illiterates received specific instruction in reading and writing, but did not have a normal school curriculum involving drawing, geography, and elementary geometry. Specificity of instruction effects is also supported by the observation that ex-illiterates develop mirror-images discrimination skills which are observed neither in illiterates (Verhaeghe & Kolinsky 1992) nor in readers of a written system (Tamil syllabary) that does not incorporate mirror image letters like "b/d" (Danziger & Pederson 1998).

However, in comparison with the problem-solving tasks and the difficult-to-perceive stimuli considered by Pylyshyn, the tasks and stimuli used in our experiments appear deceptively simple to any educated person. For example, the part-verification situation designed by Palmer (1977; cf. Gottschaldt 1926) was used by many authors to investigate hierarchical perceptual representations. Yet, whereas all groups were good at detecting salient parts, both pre-school children and unschooled adults (illiterates and ex-illiterates) often missed the more deeply embedded parts, which was not the case with second-grade children (Kolinsky et al. 1987; 1990). Thus, the effects of training or formal instruction are not limited to exceptional cases. A non-negligeable part of our phenomenal experience is modelled by school learning.

In a sense, all readers of Pylyshyn's paper are like Rembrandt experts. Although visual perception cannot be assimilated to visual cognition, the scope of visual cognition in the visual experience of hundreds of millions of human beings has probably been underestimated by contemporary psychology, which tends to substitute the educated mind for the biologically mature mind.

An even stronger case for the cognitive impenetrability of visual perception

Lester E. Krueger

Department Psychology, Ohio State University, Columbus, OH 43210-1222.
krueger.2@osu.edu **cog.ohio-state.edu/people/cfaculty/krueger.html**

Abstract: Pylyshyn could have strengthened his case by avoiding side issues and by taking a sterner, firmer line on the unresolved (and perhaps unresolvable) problems plaguing the sensitivity (d′) measure of top-down, cognitive effects, as well as the general (nearly utter!) lack of convincing evidence provided by proponents of the cognitive penetrability of visual perception.

Pylyshyn's case is strong, but could be made even stronger. Here is how it could be strengthened:

1. Numerator versus denominator sensitivity. First, in his section 4, Pylyshyn could have come down harder on the weasel word "sensitivity," and all that it may (mis)represent. In statistics, "sensitivity" refers to power; hit rate is the analog of power in signal detection theory (SDT), and in medical applications of SDT, "sensitivity" indeed refers to hit rate (Macmillan & Creelman 1991, p. 32). Normally, however, "sensitivity" in SDT is synonymous with d′, that is, the difference between the distribution means divided by the common standard deviation, when equivariance is assumed (Macmillan & Creelman, p. 373). This clearly provides a better definition of "sensitivity"; d′ excludes effects of response bias such as placebo effects that involve changes only in the criterion measure (β), whereas hit rate does not.

However, d′ cannot provide an unambiguous measure of a unitary "sensitivity" because it actually consists of two independent parts. Let us term the numerator component (i.e., the difference in mean psychological effects) "numerator sensitivity," and the denominator component (i.e., the common standard deviation) "denominator sensitivity." Normally, numerator sensitivity is disregarded, because it is assumed not to vary when the physical difference between the two distributions is held constant. Thus, a change in d′ is normally assumed to represent a change solely in denominator sensitivity, that is, the noise level or general microefficiency of processing, or what Pylyshyn terms "sensitivity in the strong sense" (sect. 4.2, last line). However, as Pylyshyn rightly implies, that is based on the very questionable presumption that attention is not a factor.

If shifts in attention are allowed, then numerator sensitivity (which Pylyshyn presumably would term "sensitivity in the weak sense") can no longer be disregarded, and as a consequence changes in d′ may be attributed just as well to changes in numerator ("weak") sensitivity as to changes in denominator ("strong") sensitivity. In this scenario, "sensitivity" based on d′ is no more totally free of nonpenetrating cognitive effects than is "sensitivity" based on hit rate.

I agree with Pylyshyn that attention to parts (spatial locations) and properties is indeed important, and that such attention may produce, at most, a shallow cognitive penetration of perception. "Top-down, cognitive processes may affect where a perceiver is attending in space, but they typically have little effect on what features are extracted or compared (Fodor 1985; Krueger 1989; Marr 1982)" (Krueger & Tsay 1990, p. 450). Attention shifts based on physical features may be used to debunk and explain away seemingly very early, almost magical access to cognition, such as in the case of the category effect (i.e., the pop-out of digit targets among letter distractors, and letter targets among digit distractors). When subtle, Clever-Hans type physical cues are eliminated, poof goes the category effect (Krueger 1984), which evidently occurs not because the letter-versus-digit meaning is acquired rapidly, but because people have overlearned the distinctive features that differentiate letters from digits, and can direct their attention accordingly.

2. Levelling versus sharpening. Second, it gets even worse for d′. Even if d′ were a pure measure of deep cognitive penetration (i.e., denominator sensitivity), it is not really obvious how cognition should affect it. Should d′ be increased or decreased by cognition? New Look investigators assumed that "anomalous or unexpected stimuli tend to be assimilated to their regular or expected counterparts" (sect. 1, para. 4), just as occurs in the case of proofreaders' errors. Pylyshyn takes for granted the assumption that, if cognition did penetrate perception, it ought to override the sensory evidence, that is, reduce d′. (Pylyshyn terms "surprising" the increase in d′ for identifying the type of distortion in sententially predictable words found in Samuel 1981, Experiment 3; Samuel, too, was surprised, and he dismissed this result as cognitively shallow, reflecting merely a decreased load on the perceptual system through predictability.) Yes, sensory evidence may be overridden, thus reducing d′. That is evident in the case of the subjective contours visible when one resolves the fragmented figures shown in Pylyshyn's Figure 4, as well as in the compromise reactions (e.g., a red spade is reported as purple) obtained in the incongruous playing card study of Bruner and Postman (1949).

However, in these cases cognition is at odds with an impoverished or distorted stimulus. When cognition is in harmony with the stimulus, one might expect it to increase, not decrease, d′. That fits Proctor's (1981) priming principle, which posits facilitation at feature extraction or comparison (i.e., an increase in d′) owing to the prior presentation of a (priming) character (for supporting evidence see, for example, Chignell & Krueger 1984; Proctor & Rao 1983). Such priming evidently is cognitively shallow or a top-down effect entirely within vision, because it only occurs on physical matches, not on name matches, but it nevertheless provides an instructive model for what deeper cognitive penetration might do (i.e., increase d′).

Even with distorted stimuli, however, there is very good reason to expect cognition to sharpen distortions (i.e., increase d′) rather than override or level them. According to the notion of schema-with-correction (Attneave 1954; Woodworth 1938), a correction (anomalous or unexpected feature) ought to "stick out like a sore thumb" from the cognitively based schema. In perception proper, it is indeed the deviations from the schema (standard or prototypical stimulus) that pop-out (Treisman & Gormican 1988).

Cognition, like perception (Arnheim 1986), might both sharpen and level, with the relative balance depending on the circumstances (e.g., on whether the dimensions involved are separable or integral; Treisman 1988). If so, then proponents of cognitive penetrability might seize the occasion to argue that d′ reflects denominator sensitivity and that the null effects on d′ usually obtained simply reflect cases in which there is a nearly equal balance between sharpening and levelling. That would be a self-defeating move, however, because unless they could justify those assumptions, the move would render their position unfalsifiable, and thus unscientific.

3. Side issues. Third, Pylyshyn dilutes his case by dwelling too heavily on side issues, and he sometimes hits the wrong note in his excursions. He speaks of "feedback from global patterns computed later within the visual system to earlier processes" (sect. 8, last para. of target article), and he says that "global aspects are thought to be computed later . . . than local aspects" (sect. 5, para. 1). The local elements certainly come first at the very earliest (receptor) level, but not at later levels, where the faster magnacelluar pathways and dorsal stream, which handle the coarse aspect, outpace the slower parvocelluar pathways and ventral stream, which handle the fine aspect. Attention may be required for the integration or binding of features across separable dimensions (Treisman 1988), but not across integral dimensions, and global processing dominance (global precedence, global interference) is evident in some tasks (Navon 1977).

Pylyshyn reports that "Hernandez-Péon et al. 1956, showed that the auditory response in a cat's cochlear nucleus was attenuated when the cat was attending to a visual stimulus" (sect. 3.2, para. 2). Alas, this work has been vigorously criticized (Moray 1970; Neisser 1967), and it garnered only two citations in the 1997 Science Citation Index. Was the early criticism inapt or misguided, and is there some reason (new evidence? new analyses?) to resurrect this work?

4. The dog that did not bark. Fourth, Pylyshyn disregards what is arguably the best way of all to dismiss cognitive penetration of perception: accept the null hypothesis. "The fact that so many investigators have tried (and failed) to find such effects suggests that they do not exist and that perception is, to all intents and purposes, cognitively impenetrable" (Krueger 1989, p. 770). True, accepting the null hypothesis should never be done lightly, because some null results may be due to inapt or insensitive experimental designs, but in the present case its acceptance seems quite justified, even overdue. The New Look movement has had 50 years to make its case, but it has failed to do so. If robust, unmistakable top-down effects of cognition on feature extraction and the like existed, then surely they would have been established by now. It is time to move on.

Binary oppositions and what focuses in focal attention

Cyril Latimer

Department of Psychology, University of Sydney, Sydney, NSW 2006, Australia. **www.psych.usyd.edu.au/staff/cyril**
cyril@psych.usyd.edu.au

Abstract: Pylyshyn makes a convincing case that early visual processing is cognitively impenetrable, and although I question the utility of binary oppositions such as penetrable/impenetrable, for the most part I am in agreement. The author does not provide explicit designations or denotations for the terms penetrable and impenetrable, which appear quite arbitrary. Furthermore, the use of focal attention smacks of an homunculus, and the account appears to slip too easily between the perceptual, the cognitive, and the neurophysiological.

Binary oppositions. Newell (1973) was critical of binary oppositions as means of formulating research questions and directing research in cognition, and Pylyshyn's penetrable/impenetrable perception, exogenous/endogenous attention, and perceptions/cognitions are good examples of such oppositions. Newell's main concern was that such an approach fosters oversimplification in classification and hypothesis testing and directs attention away from the more difficult, pertinent and complex questions of mechanism and causality. What do we learn from the supposed distinction between exogenous and endogenous attention, and how exactly does this binary classification of attentional acts enhance our knowledge of them? Early this morning on my walk, I stepped to the edge of a pond, causing a cygnet to swim quickly to the shelter of its parents. What do we learn from classifying the cygnet's behavior as controlled by exogenous attention?

Surely we wish to know what mechanism underlies the cygnet's behavior, and how and under what influences the mechanism developed. Similarly, what do we learn from the penetrable/impenetrable distinction? Newell's position was clear, and he emphasized the priority of careful theorizing and modeling rather than oversimplified classification and experimentation.

Designations and denotations of penetrable/impenetrable. Pylyshyn does not provide clear and unambiguous designations for his terms penetrable and impenetrable. We are not told exactly what the properties of the penetrable are and what attributes set them apart from the impenetrable. Additionally, the denotations of the terms appear quite arbitrary when we consider the examples of penetrable and impenetrable referred to in the paper. Observers cannot articulate the principles of how they calculate constancies and lightness and shape from shading (sect. 5.2), but solving a crossword puzzle is a cognitively penetrable function (sect. 1.1). It is not clear just how these examples of the penetrable and the impenetrable differ. When I perceive the edges formed by a door in a wall, I agree that, while I may be capable of describing my perception of the edges, I am not aware of the perceptual processes mediating my perception. Similarly, when I solve a cryptic crossword item, I may be aware of how I derived the solution from the clue and my knowledge of language, but I am not aware of the processes (cognitive, perceptual, and neurophysiological) that mediated my solution. For these possible processes of search and comparison, I turn hopefully to current theories of language, word recognition, lexical memory, and access. In what sense is the calculation of shape from shading impenetrable and the arrival at crossword solutions penetrable?

Focal attention. Attempts to characterize and explain attention are dogged by homunculi and Pylyshyn's argument that focal attention is the means whereby cognition may affect lower-order perception is no exception. Without becoming enmeshed in the details of the various spotlight, zoom-lens, and activity-distribution conceptions of attention (most of which covertly involve some active, intentional, decision-making agent that determines the focus and allocation of attention), it is possible, within Pylyshyn's own terms to have a more simple and conceptually defensible conception of attention. Pylyshyn invokes Occam's razor or Lloyd Morgan's Canon on supposedly intelligent visual processes (sect. 5.2), but could do so with greater effect on focal attention, perhaps demonstrating in the process how attention itself may be conceived of more convincingly as lower-order and impenetrable. Why cannot the effects of focal attention be thought of as emerging simply from the interaction of learning, feature weighting, and some innate or preprogrammed predispositions or primary drives? Rather than an attentional system that actively seeks out differentiating features, why cannot the system passively develop optimal sets of feature weights through repeated exposure to stimulus sets and learning rules as, for example, in connectionist systems that do not necessarily have to scan inputs in order to effect recognition.

Objects and their differentiating features that become heavily weighted through learning and experience will naturally achieve ascendancy in any processes of comparison, recognition or choice. The effects of such learning on attention are clearest in cases of attentional bias such as Stroop interference (Stroop 1935) or top-right visual bias (Avrahami 1998; Latimer & Stevens 1993). These biases occur without necessarily involving covert attentional shifts or overt eye movements, and models for their explanation do not necessarily require active, front-end scanning mechanisms and their inevitable problems of intention and agency – see Cohen et al. (1990). Indeed, Pylyshyn suggests a role for learning in so-called expert perception such as chicken sexing (sect. 6.2), but seeks to explain this expertise in terms of knowing where to look and focus attention rather than the simpler, perhaps parallel, operation of optimal feature weights on elements and attributes of the visual array. Pylyshyn may seek to defend his active conception of focal attention by reference to studies demonstrating the importance of eye fixations in illusions and the interpretation of ambiguous figures (sect. 6.1), but without elaborate procedures for coupling attention and eye fixation (Latimer 1988), eye movements and fixations can be invalid indices of focal attention.

Relationships between neurophysiology, perceptions, and cognitions. It is not clear in Pylyshyn's account how perceptions, cognitions, and neurophysiological events are related, and he slips easily between these domains during the course of his argument. It is, of course, unfair to ask Pylyshyn to provide immediate answers to the venerable questions of reduction and of how the mental can affect the physiological. Nonetheless, we need some explicit account of how the penetrable are penetrated and why the impenetrable cannot be penetrated, which highlights again Newell's (1973) plea for clearly specified theory and mechanism rather than descriptive binary oppositions.

Better ways to study penetrability with detection theory

Neil A. Macmillan

Department of Psychology, Brooklyn College of CUNY, Brooklyn, NY 11210.
nmacmill@broadway.gc.cuny.edu

Abstract: Signal detection theory (SDT) is best known as a method for separating sensitivity from bias. If sensitivity reflects early sensory processing and bias later cognition, then SDT can be use to study penetrability by asking whether cognitive manipulations affect sensitivity. This assumption is too simple, but SDT can nonetheless be helpful in developing specific methods of how sensory and cognitive information combine. Two such approaches are described.

In sections 4.1 and 4.2 of the target article, Pylyshyn asks whether signal detection theory (SDT) can be used to assess cognitive impenetrability, and shows that past attempts to apply this psychophysical methodology have been flawed. My goal here is to suggest an SDT approach that might be more successful.

The naive argument rejected by Pylyshyn is that because d' is a perceptual measure and β a decisional one, β effects arise later in processing than d' effects. Thus cognitive penetrability is shown if changes in cognitive variables influence d' rather than β. Pylyshyn attacks this logic on several grounds, including the existence of models in which a single primitive affects both sensitivity and bias, influences that make d' a "task-relative" statistic, and problems in applying SDT to complex experimental designs. Indeed, although SDT has long been used to characterize and distinguish among psychological processes, it is not as naturally suited to examining stages of processing as, say, response-time techniques. But no other method provides an easy solution to the problems Pylyshyn outlines, and it is worth exploring the detection-theoretic approach.

I consider two SDT approaches that might be brought to bear on the penetrability issue. Both depend on a more-or-less specific model, but offer the benefit of quantitative predictions.

1. Local and global sensitivity. The sensitivity index d' is "invariant" over changes in response bias in many applications, but this *local* (two-stimulus) statistic is of limited value in the penetrability domain. According to one well-supported model, focal attention can alter local d' but not global measures of accuracy that include all possible signals. For divided-attention tasks, in which the observer must detect either of two distinct signals, such as a light flash and a tone burst, it has been proposed that $\Sigma(d')^2$ is a fixed value over changes in attention to the two signals. Lindsay et al. (1968) found just this result for several (though not all) pairs of dimensions in the original test of this hypothesis. The finding of a constant "multichannel processing capacity" (Lindsay et al., p. 115) itself supports the impenetrability of basic perceptual processing by the attention module: the observer can, through reallocation of attention, improve sensitivity for lights at the expense of tones, or tones at the expense of lights, but cannot affect total sensitivity. To test cognitive penetrability from above, one could ask whether providing a new semantic interpretation of the signals affects total sensitivity. By using a global rather than a local measure, one can minimize the difficulty that the semantic effect may be to raise activation in a region not currently being tested, an issue raised in section 4.2.

2. Strong and weak sensitivity. Pylyshyn distinguishes "sensitivity in the weak sense," which is "task relative," from "sensitivity in the strong sense," which is not. As "d' is simply a measure of discriminability" (sect. 4.2, para. 9), it can take on different values in different paradigms, and it is in this sense that sensitivity measures are weak. Is there such a thing as strong sensitivity, that is, a measure that is independent of task?

There is not, but a model that directly considers task effects can be used to make a distinction, more to the point for the penetrability question, between *basic sensitivity* that reflects sensory processing only, and sensitivity that depends on *trace* and *context* memory as well. Such a model was proposed by Durlach and Braida (1969) to account for a variety of experiments measuring listeners' ability to distinguish tones differing in intensity; the model has since been applied to many other continua, for example, several dimensions that distinguish speech sounds (Macmillan et al. 1988). Basic sensitivity is considered to be the best performance of which the observer is capable, and is measured in tasks that minimize memory factors, that is, in fixed-discrimination designs. Sensitivity is poorer in other tasks because of variance added by time- and context-dependent memory. The theory thus denies the existence of a "global" sensitivity, in that sensitivity is always task-relative, but distinguishes sensory and memory processes through explicit modeling of how these factors combine to limit performance in several converging tasks.

The trace-context model has, in effect, been used by its developers to study the impenetrability of the perceptual module by range effects. They find that changes in stimulus range are well-described by the context variance component of the model, without changing the estimate of sensory variance. The model is derived on the assumption that context and trace variance contribute independently to performance, as would be true under the impenetrability assumption that range effects arise later than sensory ones and do not feed back.

Of course, such factors as context- and trace-coding may be included in "stimulus evaluation," in which case the trace-context theory itself is not the best approach to the penetrability question. The argument is not that this particular model is the tool of choice, but that explicit models about how cognitive and sensory factors combine in determining performance are required for assessing penetrability psychophysically.

Defining perception and cognition

Dennis J. McFarland[a] and Anthony T. Cacace[b]

[a]*Wadsworth Center for Laboratories and Research, Albany, NY 12201-0509.*
[b]*Departments of Surgery and Neurology, Albany Medical College, Albany, NY 12208.* **anthony_cacace@ccgateway.amc.edu**
mcfarlan@wadsworth.org

Abstract: Discussions of the relationship between perception and cognition often proceed without a definition of these terms. The sensory-modality specific nature of low-level perceptual processes provides a means of distinguishing them from cognitive processes. A more explicit definition of terms provides insight into the nature of the evidence that can resolve questions about the relationship between perception and cognition.

Pylyshyn argues, based on the cognitive impenetrability of visual perception, that vision is distinct from cognition. That is, the visual system is a distinct, "encapsulated" module. As in most discussions of the nature of perception, however, a definition of terms such as perception and cognition has not been provided. Thus, we must evaluate Pylyshyn's arguments based on our intuitive or subjective impressions of visual perception and cognition.

We have previously discussed the distinction between perception and cognition as it applies to the diagnosis of perceptual disorders (Cacace & McFarland 1998; McFarland & Cacace 1995). Within this context, we have defined perceptual disorders as information processing deficits that predominantly involve a specific sensory modality. This view is consistent with Pylyshyn's discussion of visual agnosia. It can be extended to encompass a discussion of visual perception in general by noting that perception is distinct from cognition by virtue of its modality-specific nature. This is consistent with Pylyshyn's statement that "perceptual principles, unlike the principles of inference, are responsive only to visually presented information."

This functional definition of perception leads us to conclude that visual perceptual processes are those that are predominantly determined by visual input (past and present), whereas cognitive processes are those processes that are supramodal in nature. Thus, to the extent that some stage of information processing is driven mainly by sensory information, we may conclude that it is perceptual. The contention that cognitive penetrability is not demonstrated by top-down effects in cases involving modulation in the visual system follows from the sensory modality-specific nature of perception. Pylyshyn's contention that early vision is absolutely impenetrable cognitively is a strong version of this position; we would suggest that perceptual processes need only be driven predominantly by input from a specific sensory modality.

From this point of view, data from studies involving the P300 component of sensory evoked potentials are not relevant sources of information concerning perceptual processes as this response can be evoked by surprising events in multiple sensory modalities (Naumann et al. 1992). Available evidence is consistent with the neural generators of the P300 being co-localized for auditory and visual stimuli. Thus, those components of sensory evoked potentials that are predominantly supramodal do not reflect the activity associated with perceptual processes, even if they can be shown to be dissociable from overt motor responses. The discussion thus centers on the extent of perception: How far into the information processing stream can one identify sensory modality-specific processes?

Attention is problematic because it can be demonstrated that there are cross-modal influences on discrimination performance (e.g., McFarland & Cacace 1997). As Pylyshyn notes, this question concerns whether such top-down influences determine how a visual pattern is interpreted: How selective is selective attention? To the extent that cross-modal attentional processes have a generalized, non-specific influence on perceptual processes, we may conclude that cognition has not penetrated perception, but only modulated it. This issue centers on how specific such cross-modal influences are demonstrated to be.

Adopting sensory modality-specificity as a requirement for identifying perceptual processes as distinct from cognition would provide a simple means of clarifying issues in this field. This criterion is objective and consistent with common use of the term perception. It thus points to studies involving supramodal tasks as critical in resolving issues such as those discussed by Pylyshyn.

Cognitive impenetrability of early vision does not imply cognitive impenetrability of perception

Cathleen M. Moore

Department of Psychology, Penn State University, University Park, PA 16802.
cmm15@psu.edu **gandalf.la.psu.edu/cathleen**

Abstract: Pylyshyn argues that early vision is cognitively impenetrable, and therefore – contrary to knowledge-based theories of perception – that perception is noncontinuous with cognition. Those processes that are included in "early vision," however, represent at best only one component of perception, and it is important that it is not the component with which most knowledge-based theories are concerned. Pylyshyn's analysis should be taken as a possible source of refinement of knowledge-based theories of perception, rather than as a condemnation of them.

Pylyshyn's analysis of the perception literature is insightful and is sure to provide a focus for much discussion in the field. The cognitive impenetrability that he defends, however, does not substantially undermine knowledge-based theories of perception, such as New Look (Bruner 1957) or Logic of Perception (Rock 1983). This is because "early vision" – the only aspect of visual processing that was defended as cognitively impenetrable – is only one component of visual perception, and it is not necessarily the one with which knowledge-based theories are concerned. Thus, no matter how successful an argument concerning the cognitive impenetrability of early vision may be, it would fail to undermine the basic tenets of most knowledge-based theories of perception.

Contrary to the premise that is implicit in Pylyshyn's argument, knowledge-based theories of perception do not require that all visual processes be cognitively penetrable. For example, if we take perception to be a process that is, in form, like the process of hypothesis testing (e.g., Gregory 1970; Rock 1983), then the system would require data (i.e., visual representations) on which to judge among the alternative perceptual hypotheses. Such a knowledge-based system could easily accommodate a cognitively impenetrable process (e.g., early vision) that serves to produce the entry-level representations on which later knowledge-based perceptual processes act. Thus, more generally, the identification of a subset of cognitively impenetrable stages within a broader domain of processing would not necessarily render that entire domain cognitively impenetrable.

I focussed on this issue – the dependence of perception on the visual data – with a purpose in mind. It is this issue that is ignored in the following arguments that are often made against knowledge-based theories of perception: (1) If perception is top-down and knowledge-based, then people should be able to perceive whatever they want to perceive simply by willing it, but they cannot. (2) If perception is top-down and knowledge-based, then people should be able to reverse illusions when they *know* that they are illusions, but they often cannot (target article, sect. 2, first para.). In light of the dependence of perception on the visual data, these putative "failures" are perfectly consistent with knowledge-based theories. If the visual data do not support a given perception – or if they more strongly support an alternative perception, as is often the case with illusions and the Kanizsa (1985) examples cited in the target article (sect. 2, last para.) – then the "intended" perception will not be experienced. The focus on the importance of the visual data is perhaps most explicit in Rock's theory. As Palmer describes in the Forward to *Indirect perception* (Rock 1997), Rock maintained that "every aspect of the solution [to the perceptual problem] must be supported by some aspect of the stimulus" (Palmer in Rock, 1997, p. xx). Much of Pylyshyn's argument ignores this critical aspect of knowledge-based theories of perception.

Despite what I perceive as a failure to undermine knowledge-based theories of perception, Pylyshyn's analysis may well provide insight into the following question, which is critical to these theories: What is the entry-level representation on which knowledge-based processes act? Put another way, at what point does "the stimulus," on which perception is dependent, come into the system? In order for a perception to be supported by a stimulus, that stimulus must be represented within the system. Rock, and other authors of knowledge-based theories, usually referred to the retinal image as the entry-level representation for later knowledge-based processes. Pylyshyn's argument, however, may force a shift from the retinal image as the entry-level representation to the output of early vision as the entry-level representation (cf., Julesz 1984; Nakayama et al. 1995; Treisman 1985). Such a shift, however, would not render the entire enterprise of knowledge-based perception untenable. Put another way, Pylyshyn's analysis can be viewed as a source of possible refinement for the front end of these knowledge-based theories, wherein early vision may yield relatively structured representations of surfaces, edges, and other low-level properties that may serve as the data for later knowledge-based perceptual processes. This, I believe, is the role of the surface representations that were revealed by the work of Nakayama and colleagues (e.g., Nakayama et al. 1995; target article, sect. 7.2, para. 6).

Finally, Pylyshyn's argument seems to beg the question in the end. Consider that one might question whether the processes that produce the entry-level representations should be considered part of perception at all; that is, perception may be defined as the point at which the interpretation of the visual data begins. In the context of Pylyshyn's argument, however, that would be begging the question, and it is not a necessary position to take to defend the cognitive penetrability of perception as a whole. Pylyshyn's argument, though, ends up begging the question in the same way by implicitly redefining "perception" as "early vision." Specifically, in his analysis, early vision is the only component of visual processing that was defended as cognitively impenetrable. Yet the conclusion is drawn that *perception* is cognitively impenetrable. That can only follow if perception has been redefined as early vision. While a case can be made to include early vision within the domain of perception, there is no logical foundation on which to build a case that perception *equals* early vision. Therefore, the strongest possible conclusion that could be drawn from Pylyshyn's argument is that one component of perception is cognitively impenetrable, and as I have argued, this does not present a problem for knowledge-based theories of perception.

In summary, Pylyshyn's analysis is elegant and useful. However, it should serve to refine, rather than undermine, knowledge-based theories of perception.

ACKNOWLEDGMENTS
Many thanks to Rich Carlson, Mark Marrara, and Toby Mordkoff for helpful comments on previous drafts of this commentary. Author was supported by NSF grant number SBR-9728628.

Seeing beyond the modules toward the subject of perception

Alva Noë[a] and Evan Thompson[b]
[a]*Department of Philosophy, University of California, Santa Cruz, Santa Cruz, CA 95064;* [b]*Department of Philosophy and Centre for Vision Research, York University, North York, Ontario, Canada M3J 1P3.* **anoe@cats.ucsc.edu**
www2.ucsc.edu/people/anoe/ **evant@yorku.ca**
www.yorku.ca/research/vision/evant.htm

Abstract: Pylyshyn's model of visual perception leads to problems in understanding the nature of perceptual experience. The cause of the problems is an underlying lack of clarity about the relation between the operation of the subpersonal vision module and visual perception at the level of the subject or person.

The central claim of Pylyshyn's target article is that "early vision" (following Marr 1982) "is impervious to cognitive influences" (sect. 1, para. 2). Although this claim is probably true, Pylyshyn embeds it within a questionable model of perception. According to this model, visual perception is a three-stage process: attentional mechanisms direct the eyes; early vision produces a model of the surface layout; and the organism recognizes, identifies, and interprets the patterns in the output of the vision module (sect. 1.1., para. 2). Cognition penetrates the first and third stages, but has no effect on the operation of the vision module itself.

The main point of our commentary is that this model leads to problems in thinking about the nature of perceptual experience. The cause of the problems is an underlying lack of clarity about the relation between the operation of the subpersonal vision module and visual perception at the level of the subject. We could sum up our concern in the question: Is there a place for the subject of perception in Pylyshyn's three-stage model?

Consider first Pylyshyn's claim that the phenomenology of vision is "an egregiously unreliable witness" to the subpersonal operation of the visual system: "Our subjective experience of the world fails to distinguish among the various sources of this experience, whether they arise from the visual system or from our beliefs" (sect. 7.2, last para.). It is true that our perceptual experience is silent about its subpersonal sources, but this does not mean that the phenomenology of vision misleads us about the nature of those sources. The point of visual experience is to bear witness to what goes on in the world, not to what goes on in the head when we perceive. A crucial feature of ordinary perceptual experience is its transparency: perception aims directly at the world and does not ordinarily involve beliefs about what goes on in the visual system (Pessoa et al. 1998; Noë et al., in press). To suppose that perceptual experience foists on perceivers a naïve model of the subpersonal organization of the visual system is to misrepresent the character of such experience.

Second, Pylyshyn holds that visual experience (or "visual apprehension") depends on the interpretation of the output of the vision module. This gives rise to two questions. First, what is the status of this output with respect to the perceptual subject? It would seem that the output of early vision must be inaccessible to consciousness, for conscious visual experience is cognitively saturated (sect. 1.1, para. 2). As Pylyshyn observes, visual experience seems to provide "a rich panorama of meaningful objects" (sect. 7.2, last para.). If the output of early vision were conscious, it would be cognitively penetrable. This means that all seeing, in Pylyshyn's strict sense, must be unconscious. This implication threatens to trivialize the case for the cognitive impenetrability of vision, for it is hardly surprising that the operation of an automatic, unconscious module is impervious to the subject's knowledge and beliefs.

The second question is: Who does the interpreting? Sometimes Pylyshyn implies that it is the person (as distinct from a subpersonal module) who interprets the output of early vision – for example, by drawing on background knowledge: "In order to recognize someone *as* Ms Jones, you must not only compute a visual representation of that person, but you must also judge her to be the very person known as Ms Jones" (sect. 1.1, para. 2). The problem, however, is that the output of early vision is inaccessible to the subject, and therefore cannot be an object of evaluation. Thus it is unclear what to make of Pylyshyn's third stage.

The three-stage model gives rise to an ambiguity in how we understand the phenomenon of seeing. In one sense, seeing is the operation of the vision module, and what is seen corresponds to what gets encoded in the cognitively impenetrable representation. In another sense, seeing is something the subject does, and what is seen is the world. Pylyshyn recognizes the distinction: "what we see – the content of our phenomenological experience – is the world as we visually apprehend and know it; it is not the output of the visual system itself" (sect. 7.2, last para.).

What is the relation between these two senses of what-is-seen? Is the first sort a proper part of the second? Does the content of the subject's visual experience contain a cognitively impenetrable component that derives from the early vision module? Or does the output of this module simply play a causal role in enabling visual perception at the personal level? Pylyshyn's position on these questions is not clear.

This brings us to the heart of our criticism. Pylyshyn, like other proponents of information-processing models of vision, treats vision as a subpersonal module that computes representations of the environment from retinally encoded information. His main burden in the target article is to demonstrate that the subject's goals and knowledge do not affect such computations. According to the enactive approach to perception, on the other hand, what Marr (1982) called the computational task of vision, is not the production of internal world-models, but rather the guidance of action and the enabling of active exploration (Noë, submitted; Noë et al., in press; Pessoa et al. 1998; Thompson 1995; Thompson et al. 1992; Varela et al. 1991; see also Ballard 1991, and Clark 1997). The subject of vision, in this way of thinking, is not the retina-early vision information-processing stream, but rather the whole environmentally situated animal, actively engaged in movement and exploration. Seeing is not believing, as Pylyshyn rightly insists (sect. 1.1, para. 1), but at the level of the whole animal, perception, cognition, and action are interdependent capacities. Pylyshyn is right that, as he puts it, phenomenology is just "another source of evidence, not . . . some direct or privileged access to the output of the visual system" (sect. 7.2, last para.). We wish to emphasize that the only way to understand the nature of the phenomenology of vision itself is by making the embodied and situated animal the object of perception.

How does low level vision interact with knowledge?

John R. Pani

Department of Psychology, University of Louisville, Louisville, KY 40292.
jrpani@louisville.edu **www.louisville.edu/~jrpani01/**

Abstract: Basic processes of perception should be cognitively impenetrable so that they are not prey to momentary changes of belief. That said, how does low level vision interact with knowledge to allow recognition? Much more needs to be known about the products of low level vision than that they represent the geometric layout of the world.

There are empirical reasons to question the very strong claims Pylyshyn makes for the independence of low level vision and belief systems (e.g., Peterson & Gibson 1994a). However, I wish to emphasize the many elements of truth in these claims and then to explore some of the interesting issues at the fringes of the case that has been made.

Cognitive penetrability occurs when there are changes in a process just because there has been a change in knowledge or belief. Penetrability is crucial to cognition, for it permits us to change our minds. If I am beginning my normal route home when I realize that a downpour will produce spring floods in the Ohio River Valley, I change my route to one likely to be drier. On the other hand, might there be subprocesses of cognition that evolution should have insulated from cognitive penetrability? Certainly there are, and these include most importantly the perception of physical properties of the world. It is much better for me if politics, religion, and bad advice cannot affect my ability to walk, grasp objects, navigate around obstacles, and see that it is raining. Impenetrability of basic processes of perception is more than a set of constraints imposed by computational efficiencies and limitations (or evolutionary accidents). It is just as much a set of constraints that preserve the reasons for having perceptual systems in the first place.

If cognitive impenetrability may be part of sound cognitive design, the further question arises, just how common is it? Pylyshyn claims that impenetrability is rare among cognitive subsystems, but I do not see why. Someone may convince us that a certain auditory signal is generated randomly, but if it has the same physical form as a sensible English sentence, we will understand not only the phonology but also the morphology, syntax, and semantics of the illusory sentence. No manner of change in belief about what the sound really is will interrupt the process of language comprehension. In like manner, we may cry real tears at the end of a television movie even though we know perfectly well that we have done nothing more than stare at the consequences of an electron gun exciting an array of phosphors. Decisions about how to behave are affected by this knowledge, but the processes of object, event, and scene recognition are largely unaffected. Thus, even cognitive abilities that clearly depend on knowledge may be quite impenetrable to momentary changes in knowledge. Finding cases of cognitive impenetrability seems less of a challenge than is specifying the sources and processes of cognitive penetrability.

Notwithstanding such considerations, Pylyshyn is at pains to separate low from high level vision on the basis of whether aspects of vision are impenetrable. Identifying a painting as a genuine Rembrandt must involve knowledge above and beyond a representation of its immediate physical characteristics. Hence, this would involve vision that is over and above low level vision. Pylyshyn grants, however, that visual recognition in this case may depend on being sensitive to a set of physical properties that individuate Rembrandts. In like manner, he suggests, there may be something to J. J. Gibson's (e.g., 1979) claims that people can perceive the affordances of spatial layouts immediately. Here we are at the fringes of the case that Pylyshyn makes. He spends relatively little time on explaining the transition between low and high level vision, the tone is speculative, and the implications for further development of a theory of vision are not obvious. Just what is the force of the claim that knowledge may become "compiled into" (sect. 6.4) low level vision and thence become part of it?

Pylyshyn has developed an analysis that isolates low level vision, but when he begins to examine the interface between vision and knowledge, they appear to become inseparable parts of a single information system. Knowledge of Rembrandts or affordances is linked in the process of recognition to the spatial properties that specify those things. One need only go to magazines to find an endless source of visual properties that depend on knowledge and low level vision working in a single system. How does an advertiser make a model look healthy, happy, attractive (handsome or pretty), alluring, confident, nice, or glamorous? What is it in a news photo that makes a scene look natural, squalid, modern, or dangerous? Much of the answer certainly involves the knowledge that people have, but it involves knowledge inseparable from sensitivity to spatial properties provided through low level vision. This information processing system is rather underestimated by the suggestion that low level vision makes available a description of lines and surfaces in 3-D space. It is like saying that the Eiffel Tower is a lot of girders (but that French culture recognizes it as a beautiful design).

A natural place to turn for help in clarifying the processes of vision is the concept of visual attention. Eye movements and attention are (the sole) sources of cognitive penetrability in low level vision, in Pylyshyn's view. But there is relatively little description in this paper of what attention does. For example, what is the relationship between the many phenomena of spatial organization (e.g., Pani 1997) and visual attention? Answering this question might say a great deal about the interaction of low level vision and knowledge.

It is generally agreed that there is a low level visual system, that it makes available a description of the physical world to other cognitive systems, and that it is more or less resistant to interference from belief (e.g., Marr 1982; Nakayama et al. 1995). Pylyshyn has written an admirably scholarly paper on this topic. It provides a permanent antidote to some rather extreme views about the influences of knowledge on perception expressed some years back (e.g., Bruner 1957; Pylyshyn 1973). Ultimately, however, what is not said commands more interest than what is. Enumerating limits on low level vision leads to the questions of just what information is obtained from low level vision and what processes obtain it. Although these questions are not answered in this paper, Pylyshyn has once again encouraged us to think about important issues.

Is perception of 3-D surface configurations cognitively penetrable?

Thomas V. Papathomas

Laboratory of Vision Research and Department of Biomedical Engineering, Rutgers University, Piscataway, NJ 08854. **papathom@zeus.rutgers.edu**
www.zeus.rutgers.edu/~papathom

Abstract: Among Pylyshyn's most important questions is determining the boundaries of early vision. A simple stimulus illustrates that, in addition to the dominant percept, most observers can perceive alternative interpretations of 3-D surface layout only after provided with suggestions. These observations may indicate that cognitive influences reach the stages of visual processing where 3-D surface configurations are resolved.

I applaud Pylyshyn's skillful attempt to update the thesis that the early vision module is impenetrable to cognitive influences (Fodor 1983; Pylyshyn 1984). Pylyshyn gives a useful definition of *early vision* (the first such operational definition, to my knowledge) as that part of the visual system that is cognitively impenetrable. Few perception or cognition researchers would argue that there is some stage in vision below which cognitive influences cannot penetrate. On the other hand, there is evidence that high-level cognitive processes affect the final perceptual outcome of some later processing stages (Gregory 1968; 1980; 1997; Johnston & Hawley 1994). Thus, the main issue is: Where are the boundaries of early vision, that is, up to what processing stage is vision not affected by cognitive influences? This commentary concentrates on the perceptual tasks of constructing three-dimensional (3-D) surface layout percepts from 2-D retinal projections. These tasks include assigning 3-D orientations and positions to individual surfaces, as well as resolving spatial relationships among surfaces. These tasks appear to be prone to top-down influences from extra-visual brain modules, that is modules that are outside of the early visual system (as this is defined in the target article).

Despite the many-to-one mapping to a 3-D surface from its 2-D retinal image, as detailed in section 5.1, it is remarkable that the visual system is able to recover the veridical 3-D percept most of the time. This 3-D percept, or schema, which will be denoted by P_d, dominates the alternative schemata most often. The evidence seems to favor the view that, to solve this problem, the early visual system arrives at a solution in an automatic fashion by virtue of having been wired to implement natural constraints, rather than to favor the competing view that the solution is achieved by unconscious inference.

However, there are revealing special cases for which there is an alternative stable percept, denoted by P_a, in addition to the dominant one, P_d. Most naive viewers fail to obtain P_a unless it is suggested to them; a few have to be urged to keep trying until they finally obtain P_a, others find it impossible. A prime example is "Mach's book" (Stadler & Kruse 1995), which involves a piece of hard paper folded in half along its long dimension to form a dihedral angle (say, 90°), resting upright on its short angled edge upon a plane horizontal surface such as a desk top. One alternative percept, P_a, is that of a set of slanted surfaces, forming a "roof" that is suspended in mid air over the desk top. P_a is quite difficult to obtain without suggesting the schema to naive viewers.

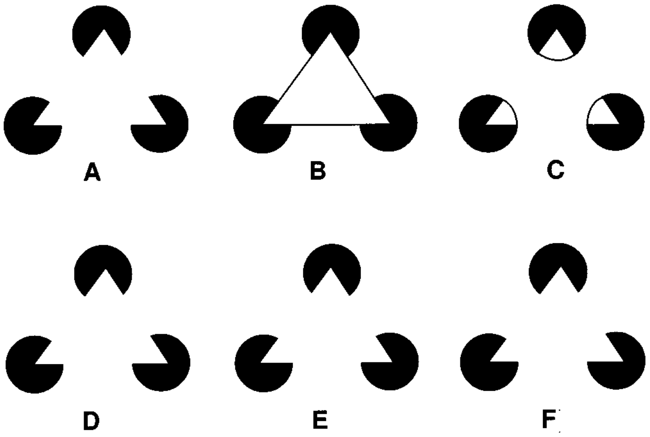


Figure 1 (Papathomas). A: This stimulus can give rise to more than one stable percept; B and C are schematic diagrams of the dominant percept and an alternative percept, respectively. The bottom panel, D, E, and F, consists of two stereoscopic pairs that illustrate the depth relationships among the surfaces in the two possible percepts B and C. If the stereoscopic pairs are uncross-fused, D and E produce a percept similar to that of B, while E and F elicit a percept similar to that of C. If the pairs are cross-fused, D and E produce a percept similar to that of C, while E and F elicit a percept similar to that of B.

We will examine a much simpler example, a classical "Kanizsa triangle" (Kanizsa 1976). The basic stimulus is shown in Figure 1A (Figs. 1B–1F are only used to elaborate the main point). Almost all observers viewing Figure 1A perceive the dominant schema P_d as shown schematically in Figure 1B: as a white triangle that occludes three black circles; observers perceive illusory continuations of the triangle's edges (added as thin lines in Fig. 1B), making the triangle appear "whiter" than the white background. This percept is so dominant as compared to alternative percepts, that the latter are obtained mostly after they are suggested. Suggestions for one possible alternative, P_a, can be made in at least three ways, listed below in ascending order of suggestive power (the level of suggestive power for achieving the percept P_a varies across subjects): (1) By verbal description of P_a: Think of the white area as a white vertical wall, and think of the black circles as portholes through which you can see the black sky behind; also, imagine a triangle being some distance behind the wall, showing partly through the portholes. (2) By a schematic drawing that matches the above verbal description, as shown in Figure 1C. (3) By actually showing a stereoscopic rendering of P_a, as can be obtained by cross-fusing Figures 1D and 1E, or uncross-fusing Figures 1E and 1F; in a dual manner, a stereoscopic rendering of P_d can be obtained by cross-fusing Figures 1E and 1F, or uncross-fusing Figures 1D and 1E. Parenthetically, the illusory continuations of the contours, hinted by the thin lines of Figures 1B and 1C, are perceived very easily in the stereoscopic versions of the two percepts.

There are two properties of this stimulus-percept set that make it particularly relevant to the central issue: (1) Most observers perceive P_a only after it is suggested to them. Moreover, observers report that they use cognitive influences to overcome the dominance of P_d, at least in their first several attempts to obtain P_a;

namely, they try to visualize the figure as a black sky showing through the portholes on a white wall. (2) When they obtain P_a, observers perceive the completion of the circles by illusory contours; these illusory curved contours in P_a (shown schematically by thin lines in Fig. 1C) are as vivid as their straight counterparts in P_d (thin lines in Fig. 1B).

The above two points taken together seem to suggest that schema-driven cognitive processes, in this case alternative interpretations, can influence visual stages that process 3-D surface layout. It may also appear, at first sight, that such cognitive signals can reach directly very early mechanisms responsible for the perception of illusory contours; neurophysiological evidence suggests that these early mechanisms involve neurons as early as V1 (Grosof et al. 1993) and V2 (von der Heydt et al. 1984; von der Heydt & Peterhans 1989). However, a more likely hypothesis is that cognitive factors influence only visual stages that are responsible for 3-D surface layout, and that this in turn may automatically, that is, by virtue of hard-wiring, influence the perception of illusory contours. This hypothesis would confine the perception of illusory contours within early vision, and would place a boundary on early vision at the stages of perceiving 3-D surface configurations.

Knowledge and intention can penetrate early vision

Mary A. Peterson

Department of Psychology, University of Arizona, Tucson, AZ 85721.
mapeters@u.ariziona.edu **www.arizona.edu/~mapeters**

Abstract: Although some types of cognition may not affect early vision, there is ample evidence that other types of cognition do. Evidence indicating that early vision is penetrable by direct manipulation of viewers' perceptual intentions and by knowledge of the structure of familiar objects is reviewed, and related to both the Pylyshyn target article and Fodor (1983).

Cognition comes in many varieties, including knowledge, beliefs, goals, and inferences. Pylyshyn argues that none of these varieties of cognition can penetrate early vision, the part of vision that computes 3-D shape descriptions. Early vision does seem impenetrable by beliefs and inferences, but it seems quite penetrable by some types of knowledge and perceptual goals.

Consider first the types of cognition to which early vision seems to be impenetrable. The New Look research on contributions to perception from beliefs and "needs" was discredited some time ago by research showing that the original results were due to response bias (Pastore 1949). Demonstrations previously taken to indicate that perception requires unconscious inference or problem solving have not met the same fate, but, as Pylyshyn states, many are now reinterpreted as reflecting the deployment of natural constraints in vision.

Consider knowledge and goals (or intentions) next. Empirical evidence indicates that subsets of these types of cognition can influence early vision. In what follows, I will discuss the research on each of these types of cognitive penetrability separately.

Intention. Julian Hochberg and I investigated whether viewers' perceptual intentions can affect the perceived organization of small 2-D and 3-D (wire) cubes. The cubes were biased toward one of the two possible interpretations that can be fitted to an ambiguous cube near one corner, but remained unbiased at the diagonally opposite corner (see Fig. 1). With the direct manipulation of viewers' intentions through instructions, we found quantitative evidence that intention can affect the depth organization fitted to both 2-D and 3-D cubes (Peterson & Hochberg 1983). Intention effects obtained from viewers instructed to fixate an ambiguous region of the cube did not differ when the nearby region was biased toward or against instructed intention, suggesting that eye movements were not mediating the results. Peterson and Gibson (1991) extended these conclusions to attentional movements.

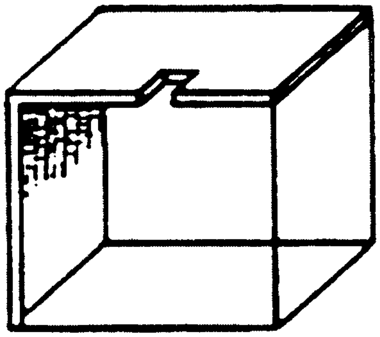


Figure 1 (Peterson). One of the cubes used by Peterson and Hochberg (1983). The cube is biased toward the interpretation that it is facing downwards and to the left in the upper left corner. Viewers fixating or attending to that part of the cube tend to see it facing downwards and to the left, regardless of their instructed intention. The cube is unbiased in the lower right corner. Viewers fixating or attending there can see that part of the cube as facing upwards and to the right just as readily as they can see it facing downwards and to the left.

These results did not seem susceptible to a response bias interpretation because indirect measures of perceived depth organization agreed with direct measures (Hochberg & Peterson 1987). The indirect measures used were ones that are perceptually coupled (Hochberg 1974) to perceived depth, so that when perceived depth reverses, they reverse as well. For example, for moving viewers, illusory concomitant motion is coupled to depth reversal and for stationary viewers, perceived direction of rotation of a moving cube is coupled to perceived depth organization. Perceptually coupled variables vary with what viewers really perceive, rather than with what they report seeing, if the two differ (Gogel 1976; Hochberg 1956; 1974).

To further test whether intention effects were perceptual rather than post perceptual, we measured the effects of instructed intention on the perceived depth organization of reversible stereograms (Peterson 1986). We chose stereograms for two reasons. First, for moving observers, differential illusory concomitant motion is coupled to changes in perceived depth in stereograms as in wire cubes. However, in stereograms, there is no relative retinal motion that might merely be registered in early vision, but interpreted later, as there is in wire cubes. Therefore, intention effects measured in reports about illusory concomitant motion in stereograms could be more confidently localized in perception. Second, with stereograms as stimuli, a Nonius fixation could be used to control both large and small (e.g., vergence) eye movements. We again found intention effects under these conditions, which allowed us to be reasonably confident that intention affected perception rather than post perceptual decisions or eye movements.

These experiments, demonstrating that viewers' instructed intentions could affect the perceived ordering of depth planes, provide strong evidence that intention can affect early visual processes. Next, I summarize the evidence that some forms of knowledge can affect early vision.

Knowledge. There are many types of knowledge. Familiarity is one type of knowledge. Experiments in my laboratory have shown that early visual processes entailed in depth segmentation are affected by knowledge embodied in memory representations of the structure of familiar objects.

To test for effects of familiarity on segmentation, we presented displays like Figure 2 in both upright and inverted orientations. The change in orientation did not change bottom-up factors known to influence depth segmentation in our displays (e.g., monocular and binocular depth cues, and Gestalt configural cues). However, a change in orientation from upright to inverted does

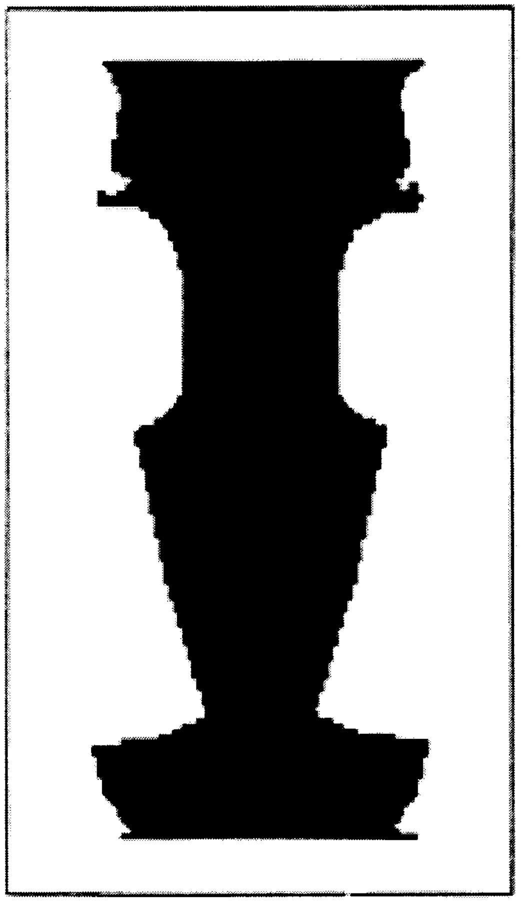


Figure 2 (Peterson). A stimulus drawn by Julian Hochberg and used by Peterson et al. (1991). Gestalt configural cues of symmetry, enclosure, and smallness of relative area (bottom-up cues to depth segmentation) favor the interpretation that the black center region lies in front of the white surround region at their shared borders. The white surround region depicts two silhouettes of standing women when it appears to be lying in front of the black region; hence, familiarity cues favor the interpretation that the white surround lies in front of the black center. The surround was more likely to appear in front of the center when this stimulus was viewed upright, as shown there, than when it was viewed in an inverted orientation, which can be seen by rotating the page by 180°.

delay outputs from object representations (Jolicoeur 1988). Therefore, influences from object representations can be revealed if depth segmentation varies with orientation so that regions denoting familiar objects were more likely to be seen in front of adjacent regions in upright compared to inverted displays. These are just the results we have obtained under both long and short exposure conditions, using both 2-D and 3-D displays (Gibson & Peterson 1994; Peterson & Gibson 1993; 1994a; 1994b; Peterson et al. 1991). From these results, we have argued that long term memories of object structure are accessed early in the course of visual processing, and that outputs from these memories serve as one more cue to depth segregation (Peterson 1994a; 1994b; 1999a).

Recent tests conducted with brain-damaged patients indicated that the object recognition processes that contribute to scene segmentation are unconscious (Peterson et al. 1998). Furthermore, tests of a visual agnosic patient demonstrated that conscious object recognition is not necessary for the manifestation of object recognition effects on depth segmentation (Peterson et al., under review). These results constitute strong evidence that one type of knowledge influences early vision.

Concluding remarks. It is important to point out that not all forms of knowledge and intention can affect perception, nor can knowledge and intention alter all aspects of perception (Peterson 1999b). The boundaries of the effects of knowledge and intentions on perception have yet to be firmly established. One possibility is that perception can be altered only by knowledge residing in the structures normally accessed in the course of perceptual organization (Peterson, in press; Peterson et al. 1991; 1996); and that intention operates through those structures (Peterson et al. 1991). Should these types of knowledge be subtracted from the sum of cognitive processes, as Pylyshyn might argue? My view is that such a strategy renders trivial a claim that early vision is cognitively impenetrable.

Cognitive penetration: Would we know it if we saw it?

Gillian Rhodes and Michael L. Kalish

Department of Psychology, University of Western Australia, Nedlands, Perth WA 6907, Australia. **gill@psy.uwa.edu.au** **www.psy.uwa.edu.au/**

Abstract: How can the impenetrability hypothesis be empirically tested? We comment on the role of signal detection measures, suggesting that context effects on discriminations for which post-perceptual cues are irrelevant, or on neural activity associated with early vision, would challenge impenetrability. We also note the great computational power of the proposed pre-perceptual attention processes and consider the implications for testability of the theory.

The core notion of cognitive impenetrability is that early vision (perceptual analysis prior to semantic interpretation and belief fixation) is insensitive to top-down influences of beliefs, background knowledge, and semantic context. To test Pylyshyn's claim we must therefore be able to determine whether the computations of early vision ("perception" from now on) are affected by such cognitive influences. Pylyshyn is pessimistic about the prospect of a "simple and direct" method (sect. 8, para. 2) for determining the locus of such influences. Nevertheless, some empirical headway must be possible, if the theory is not to be undermined by its untestability. Here we consider what might count as evidence for a perceptual locus.

Several theorists have proposed that changes in sensitivity, produced by semantic priming, would count (e.g., Farah 1989; Fodor 1983), at least when the required discrimination is a difficult perceptual one (Rhodes et al. 1993). Pylyshyn correctly points out that such changes do not necessarily indicate a perceptual change. He notes that a prime may have widespread false alarm consequences which, if not satisfactorily sampled in the experiment, will result in an apparent sensitivity change that is really due to a criterion shift. Given that criterion shifts can be mediated by changes at either a perceptual or postperceptual level, such a sensitivity change would be uninformative about the locus of priming (sect. 4.2, para. 1). Some semantic priming studies are vulnerable to this criticism. For example, in Rhodes and Tremewan's (1993) face priming study, the potential false alarm items were unfamiliar faces, which were not closely matched to the famous face targets. Similarly, in Masson and Borowsky's (1998) word priming study, potential false alarms came from nonwords that did not resemble the word targets. However, in Rhodes et al.'s (1993) word priming study, nonwords differed by only a single letter from their primed word counterparts. Therefore, the sensitivity changes obtained seem unlikely to be an artefact of inadequate sampling of potential false alarms (to nonwords that share features with primed target words).[1]

As Pylyshyn notes, other theorists have also challenged the interpretation of sensitivity changes as perceptual effects. Norris (1995) has suggested that, "As long as the nonword is more likely than the word to activate a word other than the target word, criterion bias models will produce effects of sensitivity as well as effects of bias in a simple lexical decision task".[2] In this case, hits will increase more than false alarms for primed target words, resulting in an increase in sensitivity that does not reflect any genuine effect of priming on perceptual analysis. Intuitively a nonword stimulus (e.g., "voad") must surely be less likely than the more closely matched word stimulus (e.g., "road") to activate the target lexical entry (ROAD), and must therefore, be more likely to activate some nontarget word. The problem with this analysis is that very often nonwords do not activate *any* word, and Norris's appeal to intuition ignores this large difference in the base rate with which nonwords and words activate lexical entries. The issue could be resolved by using a naming task to assess the probability with which words and nonwords activate nontarget words.

A more general problem is that in a multi-level processing system, sensitivity and bias can be measured for discriminations mediated by any level of the system, and changes in sensitivity will be evidence for a perceptual change only if the discrimination is based on perceptual, rather than post-perceptual or semantic cues. The challenge, therefore, is to find a task in which the discriminations are unambiguously based on a perceptual level of analysis. This is not straightforward, because we have no privileged access to computations at these levels. We can only measure overt responses, which may be contaminated by postperceptual processing and which are executed by response output systems. One approach might be to limit the opportunity for postperceptual processing, using a response deadline procedure. Another would be to use a discrimination for which semantic cues are uninformative, an approach advocated by Masson and Borowsky (1998). Neither lexical decisions nor face familiarity discriminations satisfy this condition, because in both cases the amount of semantic information activated by a target may provide a cue for the discrimination. What is needed are tasks requiring discriminations about some low-level feature of the target, such as judgments about the color of an oriented line within a letterstring, for which semantic information is completely irrelevant (cf. Reingold & Jolicoeur 1993). In such cases, perhaps top-down effects of semantic context on sensitivity would count as evidence of cognitive penetrability of perceptual analysis.

Another approach might be to determine whether neural activity in brain areas associated with perceptual analysis is modulated by semantic context. Initial results offer no evidence that semantic primes (from a different modality to the target) differentially activate perceptual areas. Schweinberger (1996) found identical scalp topography for the priming component of ERPs to name and face targets, suggesting that the locus of priming was nonmodality-specific, that is, postperceptual. In another domain, however, there is evidence for penetrability. Numerous studies have shown that topographically organized visual areas, including V1, can be activated in highly specific ways by top-down signals from propositionally encoded visual memories (for reviews see Kosslyn 1995; Pinker 1984; 1997). These imagery results suggest that perception is cognitively penetrable.

So far we have focussed on determining whether a cognitive effect has a perceptual or post-perceptual locus, a problem that has long been recognized (e.g., Fodor 1983). However, Pylyshyn suggests that we must also distinguish perceptual effects from pre-perceptual attentional effects. For example, the ability of experts to selective encode diagnostic parts of objects may be a pre-perceptual effect, in which attention is focussed on the spatial location that contains the relevant cues. In addition to spatial location, the proposed pre-perceptual attentional processes can be directed by cues such as shape, color, motion, and stereo disparity. We see two serious problems with this approach. First, some of these properties, such as stereo disparity and shape, are not available in the (visual) input to early vision (the attentionally modulated activity of the eye). Second, the proposed pre-perceptual system seems too powerful. If, as claimed, it can respond to any properties recognisable by template-matching, then it can compute any function through table look-up, or, equivalently, by implementing a multi-layer perceptron. Pre-perceptual processes thus seem to have access to complex representations, and to be able to perform Turing machine equivalent computations. Without clear guidelines to distinguish the functions of early vision from these powerful pre-perceptual processes, it will be difficult to test Pylyshyn's claims for the impenetrability of early vision.

NOTES

1. Note that these data cannot challenge penetrability, even if a perceptual locus is demonstrated, because the primes and targets were from the same modality. In this case, direct associative connections within the perceptual module could mediate the priming. Similar priming effects have, however, been obtained when picture primes were used (Masson & Borowsky 1998).

2. In Norris's model, semantic priming reduces the thresholds (criterion bias) of logogens for primed words. He interprets this as a postperceptual effect. Alternatively, if lexical entries are within a word recognition module, as Fodor (1983) suggests, then the locus would be perceptual.

Is visual recognition entirely impenetrable?

Azriel Rosenfeld
Computer Vision Laboratory, Center for Automation Research, University of Maryland, College Park, MD 20742-3275. **ar@cfar.umd.edu**
www.cfar.umd.edu/

Abstract: Early vision provides general information about the environment that can be used for motor control or navigation and more specialized information that can be used for object recognition. The general information is likely to be insensitive to cognitive factors, but this may not be entirely true for the information used in model-based recognition.

Early vision provides us with valuable information about our surroundings. At a general level, this information may deal with the spatiotemporal layout of visible surface patches – their relative directions and depths, and how these are changing with time. At a more specialized level, it may deal with the presence of particular configurations of surface patches that could arise when specific types of objects are seen (perhaps only in part) from specific viewpoints. The general information can be used to control eye movements (searching, tracking) or to constrain body motions (grasping, avoiding, path planning); while the specialized information can be used to recognize the presence of known objects.

Pylyshyn's target article contends that the layout information provided by early vision to aid in performing these tasks is independent of cognitive factors. This seems quite reasonable as regards control and constraint tasks, where the needed information is of a general nature. Recognition, however, requires comparing the visual information with stored descriptions or "models" of "known" objects in long-term memory. Since there are thousands of classes of known objects, and the distinctions between the classes may be quite subtle, it is difficult to regard this comparison process as merely "identification of the stimulus"; rather, the process seems likely to involve stages of hypothesizing and verifying where the verification stage(s) may require searching for highly specific types of information under the "top-down" guidance of a hypothesized model. These verification processes might well be cognitively penetrable.

Most of the "model-based" systems developed by computer vision researchers for three-dimensional object recognition (Binford 1982) deal with objects that have precise 3-D geometric definitions; only a few systems attempt to deal with "generic" classes of objects. They usually make use of low-level image features such as corners or junctions, and search for patterns of these features that could occur in some view of some known object. Even the initial detectability of the features may depend on expectations; and once some features have been detected and an object model has been hypothesized, features that confirm this hypothesis surely become easier to detect (though at this stage, selective attention, as well as response biases, may play a role). Similar remarks apply when the features are primitive regions of particular types, for example, blob-like or ribbon-like image regions that represent primitive object parts (Brooks et al. 1979; Marr 1982); here too, the segmentation processes involved in finding initial and confirming features may be expectation-dependent.

More recent work on 3-D object recognition has been "appearance-based," that is, approximate-viewpoint-specific. This approach is supported by evidence (see Edelman & Poggio 1992) that familiar objects are most rapidly recognized when they are seen from familiar viewpoints. Viewpoint-specific recognition could be based on some sort of elastic template matching of patterns of distinctive features. This process too could involve stages, with the matches found at earlier stages determining which matches are looked for, and where, at later stages. Here again, expected matches may be easier to detect.

Our ability to rapidly recognize unexpectedly encountered objects belonging to thousands of different, highly variable object classes (plants, animals, furniture, utensils, etc., or even letters of the alphabet) is not easy to account for. Apparently, we are able to rapidly extract from our visual images features that have sufficiently rich descriptions to be useful for indexing into long-term object memory, so that only a few of the many possible object classes become candidates for verification.

A framework for rapid viewpoint-specific object recognition, suggested in (Rosenfeld 1987), was based on the following steps:

1. The image is segmented into primitive regions (blobs, ribbons, etc.), and simple properties of these regions (e.g., low-order moments, which provide rich descriptions of the regions) are computed. This can be done very rapidly, for primitive regions having many different sizes, using pyramid-structured parallel processing (Rosenfeld 1989; 1990). The property values of relatively prominent (e.g., large) regions are computed at relatively high levels of the pyramid.
2. The property values are broadcast, most prominent first, to an "object memory." (Expectations could influence this step by modifying the order in which the property values are broadcast.)
3. The object memory is assumed to reside in a network of processors, each of which has stored in its local memory the definition of a class of objects. When each of these processors receives the broadcast data, it checks whether any portion of the data satisfies the constraints on region property values that define its particular class of objects.
4. If a processors' constraints are satisfied, it hypothesizes that an object of the given class may be present in the image, and it initiates appropriate verification steps. (Here again, expectations should facilitate verification.)

In summary: The "general" information about the environment provided by early vision, which is useful for such tasks as gaze control, motor control, and navigation, may be uninfluenced by cognitive factors; but this may not be true for the more specialized sorts of information that are needed for recognizing objects. Recognition, especially when there are many possible objects, seems to involve processes of hypothesization and verification. When object is unexpected, its initial hypothesization is a "bottom-up," data-driven process based on information that need not depend on cognitive factors. The subsequent verification steps, however, are "top-down" and model-driven; they look for visual information specified by the model, and this is often quite different from the initially extracted "default" information. It seems unlikely that this information is found solely through cognitive control of focal attention and stimulus identification; the process seems far more likely to involve cognitive penetration into the early vision system itself.

The future of vision needs more bridges and fewer walls

Thomas Sanocki
Department of Psychology, University of South Florida, Tampa, FL 33620.
sanocki@chuma.cas.usf.edu

Abstract: The commentator agrees with Pylyshyn's most general claims but sees problems with the more specific proposals about where the boundary between early vision and later processing might lie. The boundary cuts across current models of identification. Limitations in current research on scenic context effects preclude firm conclusions. High-level vision will benefit more from integrative work than from premature analysis.

Broad perspectives are essential for integrating the rapid advances of a field such as visual perception, and vision is ripe for perspective. Pylyshyn's target article provides perspective by providing a useful summary of evidence against old claims (of "new look" and strongly interactive models) that high-level cognitive entities such as intention and abstract knowledge directly influence early vision. The review should be informative for those outside of the field of vision but old news within.

More problematic are Pylyshyn's further claims, relating to the location of the impenetrable boundary that separates early vision from higher levels of visual processing. Pylyshyn's claims are premature because the study of intermediate and high-level vision is in its infancy and consequently a principled boundary cannot be drawn. Worse, the claims may be unhealthy because an emphasis on boundaries obscures crucial processes whose constrained interactions bridge early vision, late vision, and cognition. Two examples will be discussed.

Object identification. Identification is typically modeled with long term shape memory incorporated into the shape description process (e.g., Biederman 1987; Edelman 1998; McClelland & Rumelhart 1981; Ullman 1989). This is done because of ubiquitous familiarity effects – advantages in perceptibility for more familiar words, facial configurations, objects, and object orientations.

Because Pylyshyn seeks to separate early vision from long term shape memory, familiarity effects present a problem. It appears that Pylyshyn's only recourse is the proposal that shape description of familiar objects might be mediated by pre-compiled look-up tables in early vision (sect. 7.2). However, the complexity of such tables would be enormous. Further, each familiar item would have to be linked to corresponding long-term memories, requiring a bridging network of interconnection.

One model of word identification has a look-up-like table (the alphabetum of Paap et al. 1982). This stage is followed by a verification process that accesses memory. Interestingly, the motivation for this model was the need to integrate attention and memory with vision (see Becker 1976). Episodic approaches to identification also integrate memory and vision (e.g., Jacoby & Brooks 1984).

Objects and context. A crucial and related issue is the relation between processing of individual objects and of the larger scenic context. Pylyshyn argues that, when proper methodological measures are taken, there is no general effect of scenic context on sensitivity of object identification (e.g., Hollingworth & Henderson 1998). This conclusion is consistent with the idea that object identification is insulated from (impenetrable by) contextual processing. However, the problem of context effects is still relatively unexplored. The open issues make "continuum" a more reasonable hypothesis than "boundary."

One issue involves the overall validity of the scenic contexts. In a typical experiment (e.g., Hollingworth & Henderson 1998), the scenic contexts are paired on many trials with target objects that do not belong in (are inconsistent with) the context. This practice invalidates the contexts within the larger environment of the experiment because the scenic contexts no longer predict the objects within them. The effects of word and sentential contexts on word identification are much more facilitory when the contexts are valid in the experimental environment than when they are invalidated by using inconsistent targets (e.g., Norris 1987; Sanocki & Oden 1984; Tweedy et al. 1977). These effects are not easily attributable to post-identification processes because they obtain with naming latency (Norris 1987).

The influences of scenic contexts discussed in section 4.2 may be similar to the natural constraints of intermediate and low-level processing discussed in section 5: Like other types of natural constraints, scenic influences may "not guarantee the correct interpretation of all stimuli . . . [but could] produce the correct interpretation under specific conditions which frequently obtain in our kind of physical world" (sect. 5.1). In the same way that the appropriate conditions for studying algorithms based on the rigidity constraint involve rigid objects, the appropriate conditions for studying scenic context effects would involve objects that belong in the scenes.

Highly valid contextual environments can be created, as in studies of the influence of object contexts on the identification of object features (e.g., Sanocki 1993; 1997). The contexts were object features such as global outline and were presented immediately before target objects. Bias effects were precluded by limiting the forced choice identification response to items sharing the contextual features. The global object contexts increased sensitivity (accuracy) of the discriminations relative to various control contexts. Such methods could be extended to examine the question of whether scenic context facilitates discrimination between consistent objects.

Note that the line drawings used in scenic context research are highly tailored stimuli produced by artists. Careful depiction and positioning of objects greatly eases major problems such as evaluating edges and segmentation (see Sanocki et al. 1998). Representative stimuli require stronger contextual influences (Sanocki et al. 1998), which may include interactions that bridge knowledge and segmentation (e.g., Peterson & Gibson 1994a).

An observer's purpose is not only identification of individual objects. Localization is important and clear facilitory effects of scenic context on object localization occur (Sanocki & Epstein 1997). Observers also need to know if objects are consistent with their general expectations within a scene – such object–scene relations are processed very quickly (e.g., Biederman 1981). The rapid processing of spatial and semantic relations requires strong links between object processing, scene processing, and attention. Given such rapid relational processing, and given that objects are components of scenes after all, it seems more natural to model the processes as bridges than as barriers. Scene processing may well occur at a layer above object processing, but the similarities between the two make them seem more like part of a continuum than two sides of a principled boundary. Analysis is of course a useful tool, but the greater challenge is to understand the powerful relations between object perception, scene perception, and attention – a problem for which synthesis will be the more fruitful tool.

Color memory penetrates early vision

James A. Schirillo

Department of Psychology, Wake Forest University, Winston-Salem, NC 27109. **schirija@wfu.edu**

Abstract: Pylyshyn's concentration on form perception to demonstrate that early vision is cognitively impenetrable neglects that color perception is also part of early vision. Thus, the finding of Duncker (1939), Bruner et al. (1951), and Delk and Fillenbaum (1965) that the expected color of objects affects how they are perceived challenges Pylyshyn's thesis.

One of Pylyshyn's main objections to current theories of visual perception is that they tacitly accept the Gestalt principle that early vision has access to prior knowledge. Pylyshyn alludes to the constraint that identification requires previously established categories, suggesting that the necessary "vocabulary of geometry" is contained in "local vision-specific memory." This concentration on form perception throughout the thesis obscures the fact that early vision also generates the perception of color.

This may be why Pylyshyn failed to mention Duncker's (1939) seminal Gestalt work on "The influence of past experience upon perceptual properties." It is also possible, however, that Pylyshyn's extensive familiarity with Bruner's (e.g., 1957) work led him to conclude that Duncker's color memory is actually cognitive and thus separate from perceptual processes. Consequently, a brief review of both Duncker's and Bruner's main findings relating to color memory is warranted.

Duncker (1939) cut and reassembled a green leaf into the shape of a donkey and pasted it on a white background. On a second white background he pasted a second green leaf left in its original natural form. The two shapes were therefore identical in color and texture. Each shape was then viewed under red illumination to minimize its greenness. Observers subsequently adjusted a color wheel in a separate location under white illumination to match the color memory of each shape. Observers remembered the leaf as being approximately twice as green as the donkey which was remembered as being almost gray. Duncker claimed that the expectancy that leaves are green and donkeys are gray supported Hering's (1964) claim that it is color memory that produces color constancy.

Bruner et al. (1951) replicated Duncker's findings by matching successive presentations of a color wheel to several shapes cut from the same gray piece of paper. The shapes (a tomato, tangerine, lemon, broiled lobster claw, carrot, banana, oval and elongated ellipse) were displayed on a blue-green background. As expected, induction of all the shapes toward yellow-red resulted, yet the typically red shaped objects appeared more reddish, while the typically yellow shaped objects appeared more yellowish. However, when each shape was presented simultaneously alongside the color wheel they all appeared to be the same color. This is an example of where the New Look Psychology actually disagreed with the earlier Gestalt movement and agreed with Pylyshyn's current thesis. This may be why neither study was discussed.

However, a third study by Delk and Fillenbaum (1965) revisited Duncker's paradigm and demonstrated that it is possible to vary the perceived color of a shape. Fortuitously, Delk and Fillenbaum (1965) placed their color wheel directly behind the shape being examined. The shape could be either (1) a heart, apple, or pair of lips (all typically red), (2) an oval, circle, or ellipse (all typically neutral), or (3) a bell, mushroom, or a horse's head (all typically not red). The observers varied the color of the color wheel until the figure could no longer be distinguished from the background. This thresholding paradigm closely parallels the one used in signal detection theory as discussed by Pylyshyn, which makes this study particularly germane. Delk and Fillenbaum found that the typical red figures required a redder background before disappearing than either of the other two classes of figures, with the typical non-red figures requiring the least red in the background.

The level where shapes must be classified before determining the color of such objects is beyond what Pylyshyn would consider "local vision-specific memory." This suggests a direct influence of the past experience of color memory on current precepts. Delk and Fillenbaum (1965) have demonstrated that the cognitive categorization of shapes changes one's assumptions about their color, which resides in early vision.

The case for cognitive penetrability

Philippe G. Schyns
Department of Psychology, University of Glasgow, Glasgow G12 8RTS Scotland. **philippe@psy.gla.ac.uk** **www.gla.ac.uk/Acad/Psychology/**

Abstract: Pylyshyn acknowledges that cognition intervenes in determining the nature of perception when attention is allocated to locations or properties *prior* to the operation of early vision. I present evidence that scale perception (one function of early vision) is cognitively penetrable and argue that Pylyshyn's criterion covers not a few, but many situations of recognition. Cognitive penetrability could be their *modus operandi.*

People can place a visual stimulus in a number of distinct categories. For example, the top picture of Figure 1, might be recognized as a face, a male, an expressive face, or as "John," if this was his identity. These judgments, like many other categorizations, tend to use different cues from the visual input. This raises the general issue of whether attending to the visual cues that solve a categorization task can change the actual perception of the stimulus.

The pictures of Figure 1 (called "hybrid stimuli") were used to address this issue. They present simultaneously two faces to the visual system. In Figure 1, fine scale information (high spatial frequencies, HSF) represents an angry man in the top picture and a neutral woman in the bottom picture. Coarse scale information (low spatial frequencies, LSF) represents opposite interpretations of the same pictures – i.e., a neutral woman in the top picture and a smiling man in the bottom picture. If you blink, squint or move away from Figure 1 your perception should change.

Hybrids were used in several recognition experiments (Oliva & Schyns 1997; Schyns & Oliva 1994; 1998) which made two distinct points. The first one is that different categorization tasks can flexibly use the cues associated with a different spatial scale. For instance, Schyns and Oliva (1998) showed that deciding whether the faces of Figure 1 were expressive or not used fine scale cues whereas categorizing their specific expression (angry vs. happy vs. neutral) utilized coarse scale cues.

The second point is that people who engage in these tasks do not tend to perceive that the stimuli comprise two distinct faces or scenes. Instead, they only perceive the information associated with the spatial scale that is the basis of their categorizations. Shortly put, the perception of different visual contents (e.g., man vs. woman; smiling vs. neutral) arises as a side-effect of using visual information (here, spatial scales) that is diagnostic of a categorization. This cognitive penetrability of scale perception satisfies to Pylyshyn's own theoretical and methodological criteria for the reasons detailed below:

1. *Spatial scale filtering is part of early vision.* One early operation of the visual system is the decomposition of the input at multiple spatial scales (see de Valois & de Valois, 1990, for a review). Their temporal integration is very fast (below 50 msec) and their detection does not vary appreciably over a wide range of spatial frequencies at very brief (i.e., 30 msec) presentations (e.g., Hammett & Snowden 1995). Evidence that different categorizations modify the perception of spatial scales therefore suggests that cognitive demands penetrate one important function of early vision.
2. *Time course.* Hybrids are presented tachistoscopically for 50 msec, on the computer monitor, to adjust to the time course of the perception of all spatial scales.
3. *Hybrid stimuli and visual stimulation.* Hybrids, unlike band-filtered stimuli, do stimulate all spatial frequency channels and are therefore unbiased stimulations with respect to their subsequent orthogonal perceptions. The effect is not a bottom-up influence.
4. *Hybrid stimuli and response criteria.* Hybrids associate a different perceptual content to a different spatial scale in a "cue-conflict" situation, but this association is counterbalanced in the stimulus set (e.g., LSF equally represent male and female and so do HSF). The effect is not a response bias due to an imbalanced stimulus set. Hybrids are therefore well suited to explore the perception of scale-specific contents in response to various cognitive demands.

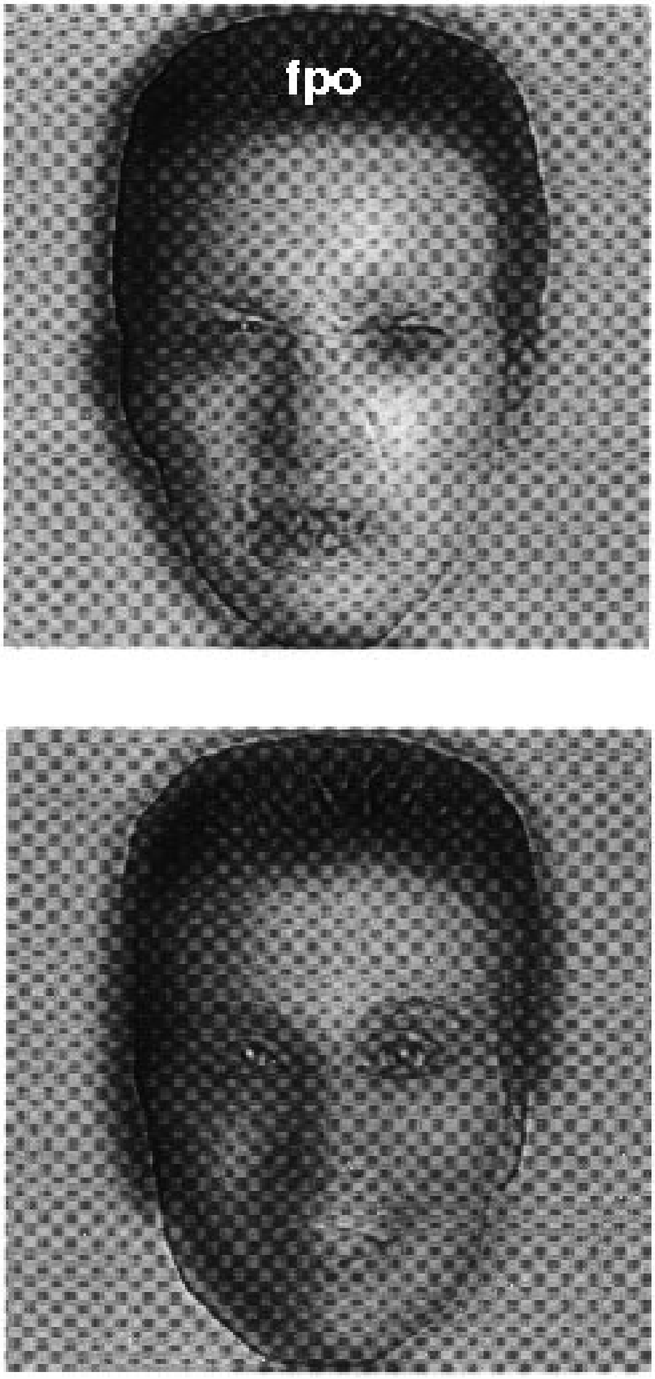


Figure 1 (Schyns). This figure, adapted from Schyns and Oliva (in press) illustrated *hybrid faces.* The top picture represents the High Spatial Frequencies (HSF) of an angry man; the bottom picture depicts a neutral woman. Low Spatial Frequencies (LSF) represent a neutral woman in the top picture and a smiling man in the bottom picture. If you blink, squint, or move away from the pictures, your perception should change.

The evidence with hybrids suggest a cognitive determination of scale perception. In agreement with Pylyshyn, I must stress that participants knew which categorization to perform; they could "allocate attention to visual properties *prior to* the operation of early vision." However, "allocating attention" is always an *explanans,* never an *explanandum.* With spatial scales, it could mean a selective adjustment of the filtering properties of LSF and HSF channels to optimize the encoding of information (e.g., contrast or orientation) in the task considered. Further studies are currently carried out to understand the parameters of spatial filtering that are under such cognitive influence.

Generalizing from these results, it would seem that the "allocation of attention to visual properties prior to the operation of early vision" does not reduce cognitive penetrability to a few situations of recognition. Instead, the most common situations concern the basic (*bird, car*) vs. subordinate (*sparrow, Mercedes*) categorizations of objects which are known to require different cues from the visual input (as LSF and HSF cues were best-suited to different face categorizations in the experiments just described). It is therefore conceivable that the common basic versus subordinate categorizations of identical objects would elicit distinct perceptual settings of early vision.

To conclude, I presented evidence of cognitive penetrability which support Pylyshyn's criterion. I then argued that the supposedly restrictive criterion of penetrability could in fact cover many situations of recognition. Two main issues are now in the empirical arena: (1) Is cognitive penetrability the exception or the rule of visual categorization? and (2) What parameters of early visual functions are under cognitive influence? Cognitive penetrability should not be settled *a priori.*

ACKNOWLEDGMENTS
The author wishes to thank Pascal Mammassian for helpful comments. The described research was partially funded by ESRC grant R000237412 awarded to Phillipe G. Schyns and a Fyssen Foundation fellowship awarded to Aude Oliva.

Perception, inference, and the veridicality of natural constraints

Manish Singh[a] and Donald D. Hoffman[b]
[a]*Perceptual Science Group, Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139;* [b]*Department of Cognitive Sciences, University of California, Irvine, CA 92697-5100.*
manish@psyche.mit.edu www-bcs.mit.edu/~manish
www.socsci.uci.edu/cogsci/personnel/hoffman/hoffman.html
ddhoff@uci.edu

Abstract: Pylyshyn's target article argues that perception is not inferential, but this is true only under a narrow construal of inference. A more general construal is possible, and has been used to provide formal theories of many visual capacities. This approach also makes clear that the evolution of natural constraints need not converge to the "veridical" state of the world.

Pylyshyn's target article distinguishes the natural constraints approach to vision from the inference approach. This distinction is real if one thinks of inference, as Pylyshyn does, as being necessarily deliberate and conscious, or at least as having unrestricted access to the individual's beliefs and goals. But clearly it is possible to think of inference more generally – e.g., as a mapping from premises to conclusions that preserves some relevant structure. Under such a formal definition, perception is very much an act of inference. Indeed, it is becoming increasingly popular to think of perception as an inductive inference (i.e., as an inference whose premises do not logically force its conclusions), that is appropriately modeled using Bayesian statistics (e.g., Bennett et al. 1996; Freeman 1996; Knill & Richards 1996). Such an approach has been used profitably to provide formal theories of many visual capacities (Brainard & Freeman 1997; Bulthoff & Yuilee 1991; Freeman 1994; Richards et al. 1996; Singh & Hoffman 1998). It has the advantage of allowing one to study perceptual inference and higher-level, cognitive (or "rational") inference within a common framework. Furthermore, it does not violate the cognitive impenetrability thesis because the premises for perceptual inference need not include the individual's beliefs and goals.

Such an inference based approach to perception also clarifies the role of natural selection in the evolution of natural constraints. For example, the target article tacitly assumes that, over the course of evolution, the constraints that human vision uses, and therefore its representations of the visual environment, become more and more veridical. The truth of this claim depends critically on how we understand the term "veridical." If by veridical we mean a resemblance relation – that, somehow, our visual representations begin to resemble the objective world – this is more than we can safely claim. We have no way to step outside ourselves – outside our perceptual constraints and representations – and evaluate to what extent these representations resemble the objective world. Every percept, observation, and judgment on our part is a conclusion (in the general sense of *inference* mentioned above) based on our current constraints and knowledge. In the language of Bayes,

$$P(S \mid I) = P(I \mid S)\, P(S) / P(I)$$

where I is an image, or set of images, presented to the visual system (such the P(I) is not zero), and S is a possible scene representation of the image. P(S | I) is the posterior probability of the scene S, given the image I. P(I | S) is the likelihood of obtaining the image I, given that the scene is S. And P(S) is the prior probability of S, reflecting the constraint embodied by the observer. This formalism makes clear that all we ever see and experience are our posteriors – based on our current priors. Over the course of evolution, our current posteriors may become our future priors. But this recursive updating by no means guarantees that the priors will eventually converge to the true probability measure that defines the world – which is precisely what would be required in order to claim that the natural constraints embodied in our visual system, and the resulting visual representations, eventually resemble the objective world. Examples such as the Ames trapezoidal window with the rod placed inside, the Pulfrich double pendulum (sect. 5.2, last para.), and other physically impossible percepts (see Fig. 1) certainly make a strong case for the cognitive impenetrability of perception. But they also underscore the lack of any resemblance relationship between our perceptual representations and the objective world.

However, a weaker interpretation of the term "veridical" is possible – one that is based not on resemblance, but on utility. In other words, the natural constraints that evolve do so because they endow the organism with a survival advantage, irrespective of whether the resulting representations bear a resemblance relationship to the objective world. Cockroaches are quite successful evolutionarily, for example, and this must be attributed to useful and sophisticated representations that they have developed. This does not mean, however, that these representations have con-

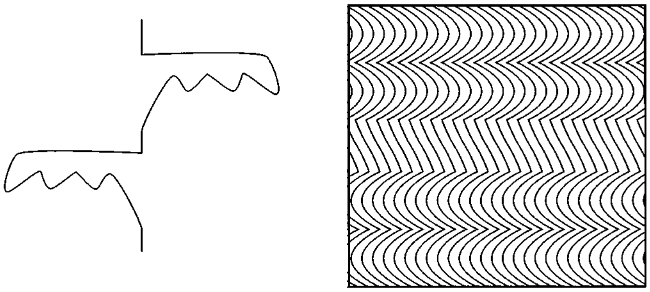


Figure 1 (Singh & Hoffman). In order to have a consistent physical interpretation, a plane curve must enclose physical material consistently on one side. Local geometric constraints can, however, override this requirement, leading to globally inconsistent figure-ground assignments (see Hoffman & Singh 1997). Such physically impossible percepts underscore the lack of any necessary resemblance relationship between our perceptual representations and the objective world.

verged or are converging, to the true state of the world. A simple way to clarify the difference between utility and resemblance is to think of an icon interface on a personal computer (Hoffman 1998). The folder, file, and trash icons on the computer screen provide a convenient, faithful, and useful way to interact with the diodes and transistors that form the circuitry of the computer. It would be a mistake, however, to claim that the shapes and colors of these icons bear any resemblance relation to the circuitry.

Furthermore, it is not sufficient to say that although perceptual representations do not capture all dimensions of the true state of the world, they do veridically represent (in a resemblance fashion) some of these dimensions – or at least they will do so eventually. It is also possible, for example, that the perceptual constraints *impose* certain dimensions on incoming data, which have no canonical dimensional structure to begin with. To take a toy example, let us say we have a "world" of words – a set with no canonical dimensions. Different organisms with different perceptual constraints could impose different dimensional structures on this set. For example, one organism could structure the set of words according to the alphabetical position of their last letter (dimension 1), and how close their first letter is to the letter "m" in the alphabet (dimension 2). Another organism might impose a very different structure on this set (based, perhaps, on the syntactical category of a word). The imposed structure in each case might, perhaps, enable the organism to interact usefully with its environment of words, but this would not mean that the organism has discovered the "true" dimensional structure of its environment – because there wasn't one to begin with.

Expert perceivers and perceptual learning

Paul T. Sowden

Department of Psychology, University of Surrey, Guildford, GU2 5XH, United Kingdom. **p.sowden@surrey.ac.uk**

Abstract: Expert perceivers may learn more than just where to apply visual processing, or which part of the output from the visual system to attend to. Their early visual system may be modified, as a result of their specific needs, through a process of early visual learning. We argue that this is, in effect, a form of long-term, indirect cognitive penetration of early vision.

Are early vision and visual perception the same? We start on a slightly pedantic point. The definition of visual perception (VP) is at the heart of the problem which Pylyshyn's target article addresses, and it is the very issue of its continuity or discontinuity with cognition that has caused such definitional problems. Many authors would include the cognitively mediated direction of attention and recognition processes, which clearly affect the phenomenal, conscious content of perception (e.g., sect. 6.1), as being part of VP. Pylyshyn does not attempt a definition directly, but argues for the existence of a distinct module which he terms early vision (EV) that is impervious to cognitive influences. He does, however, appear to acknowledge its distinction from VP, stating

> we will conclude that although what is commonly referred to as "visual perception" is potentially determined by the entire cognitive system, there is an important part of this process – which, following roughly the terminology introduced by Marr (1982), we will call early vision – that is impervious to cognitive influences (sect. 1, para. 2)

Confusingly a distinction is not adhered to or consistently brought out throughout the target article. In places, arguments are made which might support the view that EV is cognitively impenetrable, but which are presented as showing that *VP* is cognitively impenetrable. Without a clear position on the definition of VP, we feel that the case for the impenetrability of VP cannot be established. However, the case for the impenetrability of EV is argued convincingly; nevertheless, we next argue for an exception to the impenetrability of EV.

Early visual learning. The flexibility of our perceptual systems is important if we are to maximise the effectiveness with which we can interact with our environment. It has been argued that EV has been modified phylogenetically to match species specific needs by embodying a form of knowledge within the visual system (e.g., Pylyshyn's "natural constraints" sects. 5, 5.1; Barlow 1997). We believe that EV can also be modified ontogenetically, through a type of perceptual learning (Karni & Bertini 1997) that here we call early visual learning (EVL), to meet the needs of a specific individual. This process may be mediated, as Pylyshyn suggests (sects. 6.3, 6.4), by the direction of attention to the relevant stimulus properties. There is increasing evidence of the importance of attention to the relevant features in order for learning to occur (e.g., Ahissar & Hochstein 1993; Shiu & Pashler 1992) and in fact in some cases apparent EVL may be better explained as resulting from the operation of selective attention mechanisms (O'Toole & Kersten 1992; Sowden et al. 1996). Our own work suggests that expert perceivers (e.g., radiologists) learn which dimensions of visual analysis to attend to, and that, as a consequence, their analysis of them becomes enhanced through modifications *within EV*[1] (see Davies et al. 1994; Sowden et al., submitted). Further, we have shown that even language may shape basic perceptual sensitivity to a small extent, through our work on colour perception (see Davies et al. 1998), perhaps by operating to influence the direction of attention to linguistically salient dimensions of visual analysis, which then encourages subsequent EVL.

Thus far, the work described does not necessarily contradict Pylyshyn's view. As a result of acquiring knowledge, an individual modifies the dimensions of visual analysis to which they attend in accord with their particular needs. Then through a process of EVL modifications to the operation of EV take place.[2] However, whilst we agree this process may involve "shaping basic sensors, say by attenuating or enhancing the output of certain feature detectors (perhaps through focal attention)," we do not agree that this does not "alter the contents of perceptions in a way that is logically connected to the contents of beliefs, expectations, values, and so on" (sect. 1.1, para. 1) and which Pylyshyn consequently argues does not count as cognitive penetration. These modifications can clearly be connected to the individual's expectations, values, and so on (albeit mediated by attention), as shown by the learning of expert perceivers (see previous para.). Consequently, such learning could be said to indicate a form of indirect cognitive penetration. We feel it is worth describing this type of learning as a special form of cognitive penetration because such a description implies the long-term adaptiveness of the visual system to an individual's knowledge and needs.

In addition, it is not clear that attention can be afforded the degree of separation from EV that Pylyshyn suggests, because the activity of neurons in both striate and prestriate visual cortex appears to be moderated by attention (e.g., Motter 1993). Conceivably, the route to modification by EV may not be quite so indirect as we have suggested. On this point we await resolution of the degree to which attention is functionally and structurally separate from sensory processing.

So far we have argued for a special case of indirect cognitive penetration which operates on a relatively long-term time scale, where clearly an individual's cognitions have no immediate impact on their perception. However, we would like to finish on another point related to the performance of expert perceivers. Pylyshyn considers the case of chick sexing (Biederman & Shiffrar 1987) and argues that instruction teaches the sexer how "to bring the independent visual system to bear at the right spatial location" (sect. 6.2, para. 4). An alternative explanation is that the effect of instruction is to encourage the formation of a new "figure" which facilitates discrimination between the sex of the chicks. This could be analogous to the functional feature creation of Schyns and Rodet (1997) and Schyns et al. (1998) and, given the rapid time scale on which learning occurs in the case of chick sexing, may indicate that this process is not one of "tuning of basic sensory sensitivity by task specific repetition" (sect. 6.3, para. 3), but in fact indicates cognitive penetration of VP and conceivably of EV. Additional work is required to test this possibility.

NOTES

1. For instance, small, low luminance contrast features are important diagnostic signs in mammograms, a fact which expert radiologists have learned. Our data reveals that their sensitivity to such features is superior to novices', and that the process of learning to detect these features is not simply one of learning where, and to what to attend, but reflects changes within early vision.

2. In addition to the psychophysical data supporting EVL to which certain alternative interpretations have been proposed (e.g., Mollon & Danilova 1996) there is also good physiological evidence for change in early visual processing areas (e.g., Frégnac et al. 1988; Gilbert & Wiesel 1992).

Attentive selection penetrates (almost) the entire visual system

John K. Tsotsos

Department of Computer Science, University of Toronto, Toronto, Ontario, Canada M5S 3H5. **tsotsos@vis.toronto.edu**
www.cs.toronto.edu/~tsotsos

Abstract: Pylyshyn claims that if a system is cognitively penetrable, its function depends in a semantically coherent way to the organism's goals and beliefs. He rejects evidence of attentional modulation observed in neurons within the visual system, claiming that any modulation seen is not logically related to goals and behavior. I present some of this evidence and claim that it is connected in exactly the way Pylyshyn requires and thus it refutes his main thesis.

I will focus on Pylyshyn's main definition of cognitive penetrability: if a system is cognitively penetrable, then the function it computes is sensitive in a semantically coherent way to the organism's goals and beliefs. He adds

> Note that changes produced by shaping basic sensors, say by attenuating or enhancing the output of certain feature detectors (perhaps through focal attention) do not count as cognitive penetration because they do not alter the contents of perceptions in a way that is logically connected to the contents of beliefs, expectations, values, and so on, regardless of how the latter are arrived at (sect. 1.1, para. 1).

This is a convenient assumption because it dismisses the evidence that makes his thesis impossible.

Pylyshyn cites Moran and Desimone (1985) who trained monkeys to attend to one location or another within a single receptive field depending on task instructions. They write:

> The task used to focus the animal's attention on a particular location was a modified version of a 'match-to-sample' task. While the monkey held a bar and gazed at the fixation spot, a sample stimulus appeared briefly at one location followed about 500 msec later by a brief test stimulus at the same location. When the test stimulus was identical to the preceding sample, the animal was rewarded with a drop of water if it released the bar immediately, whereas when the test stimulus differed from the sample the animal was rewarded only if it delayed release for 700 msec (p. 783).

The monkey behavior and the cell responses led Moran and Desimone to conclude that when there were two stimuli within the receptive field, the response of the cell was determined by the properties of the attended stimulus; the attended stimulus is determined by the goals presented to the monkey. This basic result has now been extended to many other visual areas including V1, V2, V4, MT, MST, IT (Desimone & Duncan 1995; Treue & Maunsell 1996). Pylyshyn insists that the modulation must be logically connected with the organism's goals. If the above does not describe goal-specific responses, then Pylyshyn should elaborate on what he means.

Even more telling is the conclusion reached by Chelazzi et al. (1993) regarding the time course of attentive modulation. They recorded from IT neurons during a visual search task where the monkey is required to make an eye movement to foveate the target; correct foveation led to a juice reward. About 90–120 msec before the onset of eye movements, neural responses to nontargets are suppressed. This is clearly not a pre-perceptual effect; or is it post-perceptual, because the monkey has not yet acted on the percept. It is clearly an attentive effect that occurs during perception. This result goes against Pylyshyn's view of attentive selection.

Pylyshyn goes on to say there is no evidence that cells can be modulated by nonvisual information; the experiments of Haenny et al. (1988) do exactly this. A monkey is given the preferred orientation of visual stimuli using tactile stimulus; a disk with inscribed orientated gratings is placed under a table and away from view. Their conclusion was that such extra-retinal signals represent a prominent contributor to the activity of V4 cells in the behaving monkey. It seems abundantly clear that Pylyshyn cannot with a simple assumption eliminate the major source of trouble for his key claim.

One could attempt to use the "wiring diagram" of the visual system (Felleman & Van Essen 1991) to help Pylyshyn's cause. If there are portions of the visual system that are impenetrable, then it might also be the case that some internal layers of the system are unreachable from higher layers. The highest level visual areas in the hierarchy are areas 46, TH, and TF; area 46 is connected to no area earlier than MT and V4, TF is connected to no area earlier than V3 and VP, and area TH is connected to no area earlier than V4.

Area IT, the complex of visual areas just below these 3 areas in the hierarchy, has no direct connections to areas earlier than V4. Thus, it does seem that the higher level areas have no direct access to the earliest layers. However, the wiring diagram of Felleman and Van Essen is now dated and is no doubt incomplete. Whether or not these observations can actually lead to cognitive impenetrability of some sort is very unclear. If we simply acknowledge that connection lengths of arbitrary length are not optimal in the system, and allow one visual area as a transit station (V4), then all of the higher order areas have access to the entire set of visual areas. The hierarchy of visual areas spans 10 layers and area V4 is connected to at least one area in each layer. Thus, there is no issue regarding impenetrability since the higher order layers can affect processing in any earlier layer indirectly. In my model of visual attention (Tsotsos et al. 1995), attentional influences are transmitted layer to layer and do not depend on direct connections precisely because of the connection length constraint (Tsotsos 1990). The model is at least an existence proof that the strategy of attention penetrating all of the layers except for the retinal layers is realizable and does exhibit many of the characteristics of human attentive behavior.

Can we answer the unanswerable?

William R. Uttal

Department of Industrial and Management Systems Engineering, Arizona State University, Tempe, AZ 85287-5906. **aowru@asu.edu**

Abstract: Pylyshyn circumvents an even more fundamental question: Are the mechanisms of visual perception accessible to the theoretician? Neurophysiology, computer modeling, and psychophysics, as well as his definitions of visual phenomena suggest that he has asked an unanswerable question.

In raising the question of the penetrability or impenetrability of visual perception by cognitive factors, Pylyshyn ignores one of the most fundamental questions in psychological science. He argues that it is possible by means of converging psychophysical, neurophysiological, and computational evidence to validate the impenetrability hypothesis he puts forward. Several of us have argued that this question, like many other issues of representation, is actually unanswerable (Anderson 1978; Uttal 1998). However startling this may sound, even Pylyshyn comes close to partially agreeing when he suggests that there is no "methodological panacea" (sect. 8, para. 2) for at least one issue (does a particular observed effect have its locus "in vision or pre- or post-visual processes?" (sect. 8, para. 2).

None of the three bodies of data to which Pylyshyn refers, no matter how important or elegant they may be in their own domain, actually speaks to the problem he raises. First, computer models of vision operate within the constraints of technological capabilities, not those imposed by biopsychological realities. There can be no assurance that even the best fitting model operates by the same rules used by the organic visual system. Indeed, given the failure of computer vision systems to approximate the capabilities of human vision, one should argue that whatever is done in the computer modeling domain is, at best, analogous, but certainly not homologous. At a more fundamental level, it may be argued that *all* mathematical and computer models are neutral with regard to the underlying mechanisms. All they do is describe the overall function of a system; they cannot, for deep reasons of principle, ontologically reduce the system to its true components.

The neuroscientific data cited by Pylyshyn are equally inapplicable. Despite the enormous commitment of contemporary psychology to neuroreductive theories, there are formidable conceptual obstacles that should caution us against using many kinds of brain data to explain psychological mechanisms. Some of these obstacles are based on trivial logical errors. For example, given the widespread induction of activity in the visual parts of the brain by even simple attributes of the visual stimulus (for example, distance cues modulate activity in all visual areas; Dobbins et al. 1998), it seems inappropriate to associate any one fortuitously observed area of evoked activity with a particular psychological process. Similarly, there is widespread confusion between the transmission aspects of the brain and the representational aspects. The fact that some correlate of the stimulus is found in the peripheral portions of the visual system does not validate that locus as the site of psychoneural equivalence, nor does it reductively "explain" a visual phenomenon. The difficulties in using clinical neurophysiological data to support psychological theory are also well known (Uttal 1978).

This brings us to the last class of data marshaled by Pylyshyn: psychophysical findings share with computational models a common constraint; both are intrinsically neutral with regard to underlying mechanisms. I am sure that this will be among the most controversial points I make in this brief commentary, but readers are directed to the prescient and long overlooked article by Pachella (1974) for at least an introduction to some of the problems faced when one tries to use psychophysical data to infer underlying mechanisms. Like any other "black box," the inner workings of the brain cannot be uniquely resolved by input-output methods.

Next, note the argument that vision is not penetrated by cognitive processes is confounded by the extreme difficulty we have in defining the mental processes involved. More often than not, categories of perception and cognition are defined by the operations we use to carry out our experiments. Whether the words we use are actually biological realities (as opposed to hypothetical constructs) is still debated. Thus, Pylyshyn's approach is based on an unjustified dichotomy. He discriminates between vision and cognition as if they were two substances that could be separated from each other by some appropriate analytic technique. This premise may not be justified. Indeed, if one examines his definitions carefully, it appears that his respective views of "early vision" and impenetrability are so closely interconnected that they are circularly defined, if not synonymous. Thus, he has set up a taxonomic tautology that, despite superficial appearances, does not really support his thesis.

Finally, it must be pointed out that vision comes in many different varieties. It is not at all clear that there is a dichotomous distinction to be drawn, as Pylyshyn does, between the penetrable and the impenetrable. He is right that some visual responses are driven by constraints built into both the geometry of the external world and low level processes in our brains. However, he also enumerates many other visual phenomena that seem to be affected by their meaning or significance to a greater or lesser degree. The usual outcome of such controversies is a compromise on a continuum rather than such extreme alternatives as he proposes.

In sum, the question raised in Pylyshyn's target article is intrinsically unanswerable as are so many of the other seductively reductive questions common in contemporary cognitive neuroscience.

Segregation and integration of information among visual modules

Giorgio Vallortigara
Dipartimento di Scienze Filosofiche e Storico-Sociali, Laboratorio di Psicologia Sperimentale, Università di Udine, 33100 Udine, Italy.
giorgio.vallortigara@ifp.uniud.it

Abstract: It is argued that the alleged cases of cognitive penetration of visual modules actually arise from the integration of information among different modules. This would reflect a general computational strategy according to which constraints to a particular module would be provided by information coming from different modules. Examples are provided from the integration of stereopsis and occlusion and from computation of motion direction.

I would like to add to Pylyshyn's excellent review some considerations on penetrability among visual modules. Pylyshyn was careful at the start in distinguishing between top-down influences in early vision and genuine cases of cognitive penetration. I think the issue deserves more discussion because in several cases interactions between visual modules are confused with cognitive penetration. Although I fully subscribe to the view of cognitive impenetrability of visual modules (i.e., the fact that the result of their activity is unaffected by higher-order extra-visual knowledge), I do not think there is complete segregation of information in the visual modules. Moreover, I believe there are good (computational) reasons for a certain degree of interaction between modules. This view is in agreement with the original interpretation of modularity as a characteristic of the hierarchical organization of the biological systems: visual modules would be "nearly decomposable" systems, that is, systems in which there are rich interactions among elements belonging to the same component and few interactions among elements belonging to different components (Simon 1969b). Few, however, is not the same as none.

Let us consider some examples. Stereopsis and occlusion are very good candidates for the role of visual modules and it is widely accepted that their operations are impenetrable to high-order visual cognition. However, the visual system can integrate information from distinct modules when they provide contradictory outputs. If two stripes are arranged as a cross with the horizontal stripe occluding the vertical one, and the vertical stripe looks nearer to the observer with crossed disparity, two main percepts can occur. Either the vertical stripe splits into two unconnected halves that float separately in depth, or, after a short time, the horizontal stripe bends to allow the vertical stripe to come forward in depth (Vallortigara & Bressan 1994; Zanforlin 1982). This would be considered an example of "perceptual intelligence" by Rock and his followers (see also Rock 1984, p. 73). However, as Pylyshyn correctly points out, the fact that the visual system does its best to make sense of contradictory information does not imply that this would require cognition or cognitive penetrability. All we need is a less restrictive use of the notion of modularity of vision, one incorporating the idea that, together with the segregation of information, there is also a great deal of integration of information among visual modules.

A second example concerns a phenomenon which has been widely explored in recent years: the perception of the direction of visual motion. The accurate interpretation of moving images is difficult because any motion detection operation performed on a limited area of the image is intrinsically ambiguous. For example, when a translating homogenous edge is viewed through a small aperture (such as a neural receptive field), the component of motion parallel to the edge cannot be measured and, as a result, all motions having the same perpendicular component of translation will appear to be identical. Several solutions to this "aperture problem" have been proposed. A likely one suggests that the visual system determines the direction of motion using a two-step procedure: a first stage measures the velocities of oriented components, the second combines these components of motion to compute the full two-dimensional pattern of movement in the image (Adelson & Movshon 1982). What is common to this and other similar hypotheses is the idea that the operations of the motion module are relatively independent and can therefore be studied in isolation from other modules. However, there has been recent evidence against such a strict notion of modularity in the perception of visual motion. The basic finding is that the normally veridical interpretation of moving patterns is attained through image segmentation cues which are unrelated to motion *per se* (see Stoner & Albright 1993, for a review). For example, we recently reported (Tommasi & Vallortigara 1999) that a translating oriented grating viewed through a circular aperture with an occluding area in the middle appears to be moving alternately in an oblique or a vertical direction, depending on the foreground/background assignment on the central occluding area; the effect occurs even when the central area is simply removed from the display, thus giving rise to a "subjective" occluder. These results show that surface segmentation mechanisms play a crucial part in the interpretation of motion signals.

These and related findings (see Bressan et al. 1993; Stoner et al. 1990; Trueswell & Hayhoe 1993; Vallortigara & Bressan 1991) challenge notions of modularity in which it is assumed that the processing of specific scene properties, such as motion, can be studied in isolation from other visual processes. The ability to detect object motion may well depend on object colour, the perceived direction of motion of part of a visual scene may well depend on its foreground/background assignment, the slanting in depth of a stereo surface may well depend on its being partly occluded. However, this should not be taken to indicate that we must resurrect a role for cognition in the processing of scene properties. It is important that these phenomena not be confused with evidence for cognitive penetrability. Quite the contrary, the aforementioned results suggest, for example, that occlusion rules, which have sometimes been considered examples of high-level vision, are implemented at an early stage of cortical visual processing (Shimojo 1989).

Psychophysical and physiological data have supported the idea of strong segregation of visual information into a series of non-interactive parallel pathways, each responsible for analysis of information concerning colour, form, and motion (Livingstone & Hubel 1987). However, despite many advantages, information segregation conveys a specific problem, namely, of a pervasive ambiguity within each specific module. It is typically impossible for a module to come up with a single ("unique") solution to a problem when its input is limited to a very restricted stimulus domain. Introducing prior "natural constraints" within a domain provides a solution to this problem. Examples are of course the "rigidity assumption" in the perception of structure-form motion, or the "smoothness constraint" in stereoscopic perception.

Natural constraints are not the only possible solution to ambiguity resulting from analysis restricted to a limited domain of information. Another strategy for constraining a solution in one domain would be to recruit information from *other* domains (see also Trueswell & Hayhoe 1993). The constraints on a particular module would in this case be provided by information coming from different modules. This strategy is reasonable because in natural scenes multiple sources of information are usually available to determine a perceptual solution, and their combined use can limit the risks of having perceptual solutions based on general but locally incorrect assumptions about the properties of the world. Thus, information seemingly unrelated to the particular domain which is carrying on the main computation can indeed be employed by the visual system. This is not high-level information related to beliefs and past experience, but simply information arising from other nearly decomposable visual modules.

ACKNOWLEDGMENT
The author is also in the Dipartimento di Psicologia, Università di Trieste, Italy.

An ecological approach to cognitive (im)penetrability

Rob Withagen and Claire F. Michaels
Faculty of Human Movements Sciences, Vrije Universiteit, 1081 BT, Amsterdam, The Netherlands. **r.withagen@fbw.vu.nl**
cfmichaels@fbw.vu.nl

Abstract: We offer an ecological (Gibsonian) alternative to cognitive (im)penetrability. Whereas Pylyshyn explains cognitive (im)penetrability by focusing solely on computations carried out by the nervous system, according to the ecological approach the perceiver as a knowing agent influences the entire animal-environmental system: in the determination of what constitutes the environment (affordances), what constitutes information, what information is detected and, thus, what is perceived.

According to Pylyshyn, a system is cognitively penetrable if the function it computes is sensitive, in a semantically coherent way, to the perceiver's goals and beliefs. Penetrability is illustrated by Bruner's demonstration that hunger affects the likelihood of perceiving food, or poverty effects the perceived size of coins. Impenetrability, in contrast, is illustrated by perceptual illusions such as the Müller-Lyer figure, the Ponzo illusion, and Ames' room, wherein experience with these illusions does not affect their strength. Pylyshyn explains impenetrability by proposing that a substantial part of the visual system, the early vision system, which provides a structured representation of the 3-D surfaces of objects, is encapsulated from cognition. Cognition *can* intervene, he argues, in determining the nature of perception at two loci: prior to the operation of the early vision system and after the early vision system, at an evaluation, selection and inference stage, which accesses long-term memory.

Let us begin with one phenomenon that Pylyshyn associates with impenetrability, illusions. One simple premise of ecological psychology advances us a long way in understanding illusions: Perception does not involve computations carried out on retinal data; it is the detection of information that specifies significant aspects of the perceiver's environment. Two corollaries of this are that spatio-temporal patterns in the ambient light *specify* environmental properties and that the visual system consists of *smart perceptual devices* (Runeson 1977) that pick up this information. Illusions embody perceptual information to which we are attuned. As to the Müller-Lyer illusion, for example, Gibson (1966) distinguished between information about the length of a line and the length of a line itself. The length of a line as measured by a ruler is, he claimed, not information for human perceivers. Telling people about the illusion cannot result in detecting a sensory pattern to which one's visual system is not sensitive.

How does the ecological approach deal with phenomena that Pylyshyn associates with cognitive penetrations? To address this, we add a few more points to our above sketch of the ecological approach. What a person sees depends on the information detected and this is in part under the perceiver's control. In short, attention can be said to control detection (Michaels & Carello 1981). Patterns in light reflected off an apple, say, might specify the size, distance, variety, and ripeness (edibility) of the apple. Which information is picked up will depend on psychological states, such as hunger – they affect the perceiver's attentiveness to particular information. Hunger, we would say, directs attention to information about edibility.

One might argue that cognitive penetration is Pylyshyn's way of getting meaning into perception: goals, beliefs, and knowledge influence the semantics of the output of processing. In the ecological approach, the animal-environment system is already permeated with meaning. Given that the study of perception is the study of an animal knowing its own environment, a complete account of knowing has to include both the knower and the known (Shaw & McIntyre 1974). Gibson proposed that what an animal needs to know about its environment are the affordances, the action possibilities. The confluence of action and perception in affordances has important consequences for the distribution of meaning in the animal-environment system. First, the environment as an object of knowledge is not describable independent of the animal; objects and events must be measured in animal-referential metrics. Second, information specifies the affordances of these objects and events. Third, actions, both exploratory and performatory, help create information. Thus, inasmuch as both the environment and information that specifies it are in terms of the animal's actions, they are necessarily meaningful. In this framework, animals can detect meaning and need not alter or supplement computational processes to manufacture it.

Is early visual processing attention impenetrable?

Su-Ling Yeh[a] and I-Ping Chen[b]

[a]*Department of Psychology, National Taiwan University, Taipei 10764, Taiwan,* [b]*Department of Psychology, National Chung-Cheng University, Chia-Yi 62117, Taiwan.* **suling@ccms.ntu.edu.tw**
psyipc@ccunix.ccu.edu.tw

Abstract: Pylyshyn's effort in establishing the cognitive impenetrability of early vision is welcome. However, his view about the role of attention in early vision seems to be oversimplified. The allocation of focal attention manifests its effect among multiple stages in the early vision system, it is not just confined to the input and the output levels.

Pylyshyn has provided a largely coherent account of early vision as a functionally self-contained component, with its content and processing opaque to other cognitive faculties. The only two sites where this encapsulated module might interact with other cognitive functions reside in its inputs and outputs. According to this scheme, many contextual effects on visual perception could be explained away by the viewers' strategically allocated focal attention at the input level. The clockwork of early vision can thus be geared up and down by attention, but its general machinery and operations cannot be modified by other top-down processes originating outside the visual system. Although most of Pylyshyn's arguments about the self-contained attribute of early vision are sound and convincing, his treatment of the role of attention is oversimplified and calls for a closer inspection. Pre-perceptual filtering by directing attention towards basic properties such as location has been one of the most prevalent mechanisms posited by the early selectionists. Our own studies also tended to support this view. For example, we have shown that a location-specific adaptation, the figural aftereffect, is modulated by attention (Yeh et al. 1996). The optimal apparent position shift of the test figure induced by the previously viewed adapting figure occurs when the adapting figure is attended. The effect can be attributed to attentional modulation on the strengths of the location labels of the adapting stimuli, a pre-perceptual attention allocation view consistent with Pylyshyn's. It is equally possible, however, that it is the interactions between the location coding units across the temporal frames, not the strength of the stimuli *per se,* that are under attentional modulation. In the latter case, attention operates *after* location coding, not *before.*

We raise the issue here to argue for attentional modulation at multiple stages in the early vision system, not only at the input/output level. Over the past decades, increased evidence from psychophysics, neurophysiology, brain imaging, and clinical observations has shown that attention is not as unitary as had been thought; rather, it manifests its effects at multiple stages and multiple sites. Although defining the cognitive architecture on a purely functional basis is useful, one cannot set aside our current understandings of the neurophysiology and anatomy of the brain completely. The physical architecture of the visual brain is componentially hierarchical. This sets a strong constraint on the temporal placement of the functional stages for a biologically plausible cognitive model. If "early vision" is truly post-attentional as Pylyshyn argues, it should be placed no earlier than V4 and MT in the "what" and "where" pathways, respectively, because no evidence for attentional modulation before these two sites has been found so far (Moran & Desimone 1985; Treue & Maunsell 1996). Hence, to place attention prior to early vision would inevitably lead to the conclusion that the encapsulated "early vision" module was located somewhere after V1. Most important to note, along these two pathways there are multiple sites in which attentional modulation has been found, such as MST, posterior parietal lobe, and TE. This would pose serious problems for the exact site of "early vision" if it is strictly post-attentional and attention impenetrable.

Phenomenally, cognitive impenetrability resembles the inaccessible nature of the inferred "preattentive stage." Owing to the viewpoint and illumination dependent characteristics of V1 cells, and their lack of direct connections to higher stages such as prefrontal cortex, information from V1 has been postulated as neither directly useful nor accessible for our conscious experiences (e.g., Crick & Koch 1995). In this sense, processing in area V1 can be considered preattentive, consistent with the finding that attention sets the spatial resolution limits *after* what is supposed to be the visual resolution of V1 cells (He et al. 1996). Following this, should we then draw the flowchart as preattentive stage ⇒ attentive stage ⇒ early vision ⇒ cognition? This is quite a late processing stage for early vision. There seems to be only one alternative left for reconciling all these conflicts: early visual processing and attentional modulation are intimately interwoven. That is, instead of being sandwiched by attention at its input and output, early vision is attention penetrable at various stages.

Even if we limit our discussion to purely functional accounts, attention is by no means a single, unidirectional mechanism operating solely prior to early visual processing. Consider Treisman's feature integration theory as an example. As evidence accumulated, Treisman (1993) added to her original location-based model three other ways in which attention might operate: (1) the feature-based inhibition, (2) the object-based selection, and (3) the late selectionist's view of the sole selection stage: the control of actions. Feature-based inhibition (1) can be viewed as the top-down interaction within the visual system, and (3) works on the output level in Pylyshyn's account. This leaves (2) unspecified. The object-based selection is a good candidate for bridging cognition and early vision to resolve the inherent ambiguity in the object binding problem. It is unlikely that this job is done after the completion of early visual processing.

Penetrating the impenetrable?

Karen Yu

Department of Psychology, University of the South, Sewanee, TN 37383-1000. **www.sewanee.edu/collegecatalog/collegefaculty/yuk.html**
kyu@sewanee.edu

Abstract: A distinction is made between structural and semantic knowledge, focusing on the possible influences of the latter. To the extent that early vision may be influenced by object-identity, it would seem necessary to evoke compiled transducers to explain such an influence. Compiled transducers may furnish a way in which vision can be and is penetrated.

If vision is indeed cognitively impenetrable, should we be surprised? Perhaps not. Our perceptual systems are our gateway to the world, facilitating our interactions in and with the world by building a reliable representation of it. Vision cannot fulfill this role if it is influenced by all sorts of knowledge. Thus, it is critical that vision be independent of knowledge at some level, precisely because of the kind of information we depend on vision to provide. Indeed, informational encapsulation of perceptual processes such as vision may help guarantee that perception provides a reasonably accurate representation of the external world as it *is,* rather than as we would like or expect it to be.

Nevertheless, it is likely that by embodying reliable relationships in the world, perceptual processes can gain greater efficiency with minimal cost. As Pylyshyn argues, early vision can be impenetrable while still embodying such relationships in the form of natural constraints. But just which relationships does early vision incorporate? This is an empirical question, yet it would seem that the most stable, reliable relationships would be the most likely to be built into the system. In other words, what one might refer to as generic structural knowledge – knowledge about basic physical properties of the world (e.g., the relationship between physical size, distance, and retinal image size; the fact that opaque objects occlude one another). Certainly there is evidence to support the influence of such factors on vision (e.g., Shimojo & Nakayama 1990a; 1990b; Sigman & Rock 1974) – influences of the sort that are consistent with an impenetrable system in the sense that their influence seems to be exerted in an automatic, unmodifiable, and often unconscious fashion.

What about generic semantic knowledge? Here I refer to knowledge about certain classes of objects – knowledge that to some extent depends on recognizing and assigning meaning. Here, too, one might argue that fairly reliable relationships exist between object identity and other visually-derived characteristics. Indeed, these relationships contribute to object identification in the first place. Might it benefit early visual processing to incorporate such semantic constraints as well? Doing so might increase efficiency (although not without the distinct possibility of introducing errors). In the case of color, bananas tend to be yellow and carrots tend to be orange; stop signs tend to be red and yield signs tend to be yellow. Does an object's identity influence one's perception of its color? In the case of motion, rockets typically move vertically relative to the horizon, whereas people typically move horizontally relative to the horizon; cars typically move more quickly than bicycles; and the parts of one's body move only in certain constrained ways relative to one another. Does an object's identity influence one's perception of its motion? Various studies suggest it might (e.g., Shiffrar & Freyd 1990; Yu 1994).

And then there are some relationships that are more difficult to categorize as structural or semantic. For example, objects with a "front" tend to move in the direction that they face. Evidence suggests that the visual system makes use of this principle (e.g., McBeath et al. 1992; Yu 1994).

Based on Pylyshyn's analysis, generic structural knowledge, to the extent that it influences visual processing, is likely incorporated in early vision itself rather than exerting its influence from more general long-term memory. But what about generic semantic knowledge? Could this too be embodied in early vision? This seems unlikely given that semantic knowledge depends on object identity, which is assumed not to be extracted in early vision. It also seems unlikely to the extent that many of these semantic categories were not relevant in an evolutionary sense (e.g., cars and stop signs); although other categories were arguably quite important evolutionarily (e.g., biological motion). In the case of the former, Pylyshyn seems to offer a means by which cognitively impenetrable early vision might in fact be penetrated with time and repetition (sect. 6.4, para. 5). If indeed these relationships between semantic identity and other visual attributes are reliable enough (just what this means remains to be determined), then perhaps they would become compiled into the visual system (sect. 6.4, para. 5). This possible influence of object identity on the perception of basic visual characteristics such as color, motion, and the like also emphasizes a conceptualization of vision as a process of continued refinement through feedforward and feedback. These basic visual characteristics contribute to object identification, and yet object identity may influence the perception of the same basic visual characteristics.

Thus the notion of compiled transducers (sect. 6.4, para. 5) seems to furnish a mechanism by which outside knowledge can penetrate early vision, given enough time and repetition. However, the implication also seems to be that once this knowledge has penetrated and been compiled into early vision, it too becomes impenetrable, at least until another compiled transducer comes along. If this is indeed the case, it seems we have arrived at a slightly different meaning of impenetrability.

Author's Response

Vision and cognition: How do they connect?

Zenon Pylyshyn

Rutgers Center for Cognitive Science, Rutgers University, New Brunswick, NJ 08903. **ruccs.rutgers.edu/faculty/pylyshyn.html**
zenon@ruccs.rutgers.edu

Abstract: The target article claimed that although *visual apprehension* involves all of general cognition, a significant component of vision (referred to as *early vision*) works independently of cognition and yet is able to provide a surprisingly high level interpretation of visual inputs, roughly up to identifying general shape-classes. The commentators were largely sympathetic, but frequently disagreed on how to draw the boundary, on exactly what early vision delivers, on the role that attention plays, and on how to interpret the neurophysiological data showing top-down effects. A significant number simply asserted that they were not willing to accept any distinction between vision and cognition, and a surprising number even felt that we could never tell for sure, so why bother? Among the topics covered was the relation of cognition and consciousness, the relation of early vision to other modules such as face recognition and language, and the role of natural constraints.

R1. Introduction

The unusually large number of commentaries provoked by my target article cover a correspondingly wide range of topics and in most cases offer useful refinements, additions, and corrections to my thesis. I was pleased with this outcome. Bruner's thesis has been a very influential one (it impressed me greatly at the time and had a lot to do with my shift from physics to psychology). The idea fit well into the mid-century Zeitgeist. But it appears that times are changing: The majority of the commentators agreed in general with the independence thesis and offered welcome suggestions for its improvement. On the other hand, others characterized the distinction between early vision and cognition as anything from untenable to "old news" (**Egeth, Sanocki**) or impossible to verify – though at least one commentator felt I had not taken a strong enough stand (**Krueger**). As I read the commentaries, the main allegations seem to be that: (1) The distinction between cognition and vision (including early vision) is not worth making, or is unhealthy or impossible to decide, or is not important, or is really about such things as the difference between conscious and unconscious, or conceptual and nonconceptual, or subpersonal and subject levels; or that (2) the distinction is empirically invalid as stated because it fails to take into account evidence from neurophysiology or ecological psychology or certain psychophysical findings, which show that visual processing is awash with top-down effects; or that (3) the distinction is generally valid but is not drawn at the precise point it should be – in particular that the output of the early vision system may not be exactly as described in the target article. To those who saw merit in the enterprise but felt it went astray in places, I thank you for your suggestions and will try to respond to them and incorporate all those that seem to me to be relevant. To those who think I am wasting my time I will try to provide a brief response, though with little hope of dissuading them, because the differences in goals and assumptions are often very large.

R2. Distinctions and decidability

I will begin by commenting on some of the more general points. It is often (or perhaps always) the case that the need for certain distinctions is clearer than our ability to make them in a precise and operational manner. Chomsky (1980) pointed out that distinctions are driven by the clear cases and that it is the ensuing theory that dictates the sorting of unclear cases. In linguistics the distinction between competence and performance or between syntax and semantics (or phonology or pragmatics) is a case in point. No set of necessary and sufficient conditions is ever put forward at the outset – the distinctions become recognized as the theory develops to provide the basis for a better understanding of the nature of the distinction. The distinction between mass and momentum is another example from physics. Only the development of physical theory has legitimized the distinction in the end.

I have very little to say to those who would rather focus on borderline cases and on the problem of providing necessary and sufficient conditions for the distinction, or who believe that there is no point in trying to idealize because in the end everything is connected to everything else. I have even less hope that I can dissuade those who think the question of whether a particular theory of cognitive architecture (e.g., based on the distinction between early vision and cognition) is empirically undecidable. I find that entire line of reasoning to be both otiose and tiresome. Inductive inference, on which all scientific theories depend, is inherently undecidable. One cannot even be sure whether the earth is spherical or flat (as the International Flat Earth Research Society can demonstrate): All one needs to do to explain the data is to modify the theory in small ways and all the observations will be compatible with the earth being flat. This is a point I made in my argument with John Anderson. In his commentary, **Uttal** cites Anderson's undecidability argument, but not my response (Pylyshyn 1979), in which I showed that the argument has no grounds, at least in the case of the form of representation of mental images. Of course there are some oppositions (e.g., serial versus parallel processing) that can be shown to be undecidable relative to a particular class of data or a particular method (e.g., reaction time data, ERP, signal detection analysis), but that is not the issue that commentators raised in this case. It is also true that the distinction I am making – and every other claim about how the mind works – is undecidable in principle if what we want is certainty. But it is just as undecidable, no more and no less, as is the hidden structure of matter – yet one will never find physicists wasting their time arguing whether, for example, one could ever tell whether matter is atomic or continuous. Those who think the case for mental constructs is any different owe us an argument why that is the case: **Uttal** and **Dresp** simply assert that it is. Such claims as "the inner workings of the brain cannot be uniquely resolved by input-output methods" (Uttal) or that we cannot know the underlying process because "all we can possibly measure in psychophysics or experimental psychology is a behavioral correlate that we believe is a valid indicator of some underlying . . . process" (Dresp) have exactly the same status as the claim that the inner workings of the sun cannot be resolved because it is too hot and too far away to visit and all we have are long-distance measurements "that we believe are valid indicators of some underlying process." This pessimistic view (which is much more prevalent in psychology than in any other field of endeavor) simply underestimates scientists' cleverness in developing indirect methods based on converging evidence and theoretical systems (for more on this point as it pertains to cognitive science, see Chs. 4 and 5 of Pylyshyn, 1984). Of course we may not be in quite the same position vis-à-vis mental processes as physicists are vis-à-vis thermonuclear processes, but that is not an argument in principle. There is a world of difference between the wistful remark that psychology is hard and we have a long way to go and that we may be on the wrong track and so on, and the foundational claim that information process theories in psychology are in principle undecidable. The latter seems to me to have no more place in cognitive science than in any other empirical endeavor. The same can be said of most other arguments that proceed from some general metaphysical desideratum and go on to conclude that one empirical theory is better than another. For example, contrary to what **Dresp** seems to think, a belief in monism has absolutely no implication for whether there are principles at different levels of organization (or for whether one source of data should take precedence over another, or for whether one theoretical vocabulary is to be preferred over another). These are long-term empirical questions. One does not hear arguments in geology about whether the principles of erosion, glacier action, or volcanic processes should be eliminated in favor of a formulation in terms of Schrödinger's equation just because one believes in the unity of science. It is just accepted that different principles occur at different levels of abstractness and that many of them are general enough to define a distinct scientific domain (like, perhaps, geology or cognition).

No matter how hard it is to operationalize certain distinctions we need them to avoid being muddled about our claims. For example, **Moore** denigrates the vision-cognition distinction, then goes on to argue in favor of a "knowledge-based theory of perception" – a notion that has no meaning unless you can independently define what is meant by being "knowledge based." Moore says that even a hypothesize-and-test theory of vision would require what she calls an "entry-level representation" that is impenetrable, so that all the target article was doing was suggesting that the entry level be moved from being the retinal image to being the output of the early vision system. But that change represents a major difference in the nature of the visual architecture, and of the nature of mind. Considering that Moore gives more or less the same arguments and comes to essentially the same conclusion as I do about the "entry level," I am left wondering why a fuss is being made over terminology and why I am accused of question-begging at the end of her commentary (I am also puzzled about why she includes Rock's elegant and thoughtful work in the category of a "knowledge-based" theory of vision, because it presents some of the clearest evidence for the independence thesis – see the quote from his book that I reproduce in the target article). The point is that *everyone* (except maybe a Gibsonian) assumes a distinction such as the one between early vision knowledge-based processes because it is clear to everyone that the processes are *somehow* different. I merely tried to show how they are different and what the consequences were of making the distinction that way.

Some commentators missed the point that I tried hard to convey, namely, that visual apprehension – the apprehension of the world through the visual modality – is clearly dependent on every cognitive faculty we have, from early vision to motor control of the eyes and head, to problem solving and inductive inference. The challenge is to discover a way to factor that system into independent systems with restricted forms of interaction. The tripartite division of visual apprehension into attention–early-vision–cognition was a proposal for how the clearly knowledge-dependent visual apprehension process might decompose. It does no good to argue, as **Pani** does, that visual apprehension must depend on knowledge because there are examples in magazines where advertisers "make a model look healthy, happy, attractive, alluring," which means that these "visual properties . . . depend on knowledge and low level vision working in a single system." Of course they do; that single system is "visual cognition." The question is: How do the subsystems work together? And in particular, can we deliberately isolate a part of the system in which knowledge has no effect? My contention, for which I actually present empirical evidence and arguments, is that one *can* isolate such a stage: It is what I called early vision. Early vision could have turned out not to exist or to contain only sensors, but the evidence is that it contains a great deal more. That was the empirical finding. **Schyns,** like **Pani,** wants to conclude that cognition penetrates early vision by arguing that attentional filtering must be part of early vision itself. That depends on how one defines early vision. If early vision is a stage that is hypothesized to be independent of cognition, the empirical question is not whether it is penetrated by attention, but whether it exists (and is nonempty).

Taking a Gibsonian line, **Crassini et al.** and **Withagen & Michaels** are among those who feel that the perception-cognition distinction is pointless, but in their case it is because they do not believe that vision produces any representations at all. This puts us so far apart that there may not be a common language within which to disagree. The idea that representations require a representation-interpretation process, and so on ad infinitum (Crassini et al., and also **Gellatly**), is an argument that was once made by Wittgenstein against rules. But this abstract argument does not explain how it is that computers interpret both rules and representations without running into a regress. The answer in both cases is that "interpretation" has to be grounded on architectural properties, which themselves are not representations (unlike turtles, it is not "representations all the way down"). Of course debates such as those about imagery and vision involve a substantial degree of faith – as did the debates about color (eg., between Newton and Hooke) or about action at a distance. But that does not make it a debate about ideology. Sooner or later the facts will assert themselves (perhaps, as Thomas Kuhn once suggested, only after the respective proponents are dead) because this is really and truly an empirical issue. (By the way, there *may* be a debate about the nature of mental imagery, but I see no debate about whether perception or cognition requires *representations* – at least not outside behaviorist and Gibsonian circles). As for being interested in different questions, as Crassini et al. suggest, this may well be the case. It is also about what one is willing to take as givens in building a theory of vision and for which aspects of the animal-environment interaction one thinks a theory exists. For example, Withagen & Michaels are willing to take "direct attention to" as a primitive operation that is not amenable to an information-processing explanation (so that, for example, they can rest on such statements as that "hunger . . . directs attention to information about edibility"). As Fodor and I pointed out (Fodor & Pylyshyn 1981), without principled constraints on what can be attended or "picked up," this form of explanation becomes vacuous. On the other hand, I do agree with Withagen & Michaels that environments must be described relative to the organism and that meaning arises from the interaction of an organism and its environment. But surely we are also interested in the question of what sorts of mechanisms, compatible with physics and biology, are capable of realizing attention and of achieving the connection between world, mind, and action that Crassini et al. and Withagen & Michaels say is the purpose of perception. For this goal it does not help to say that the information is "in the external world, not inside heads" or that "perception is considered an achievement of action," and so on. We also want to know *how it works.*

Closely related, though from a different ideological perspective, **Noë & Thompson** offer similar assertions, as when they say that the subject of vision is not the "retina-early vision information-processing stream, but rather the whole environmentally situated animal, actively engaged in movement and exploration." All such ideologically motivated statements are part-truisms and part-slogans that need to be unpacked and developed into a research program leading to a causal theory (or at least some fragments of one), which has not happened outside the information processing tradition. In fact the idea that vision must be viewed as in some sense "situated" is one to which I am very sympathetic, but find to be a natural extension of the information-processing view (Pylyshyn, in press) where the goal is to discover the mechanisms that make this situatedness possible. The sorts of bland statements quoted above have few direct empirical consequences, in contrast to the question of whether vision is cognitively penetrable. That question really is about the fundamental nature of the mechanisms involved, and therefore about the nature of mind. It is about *how* organisms manage to recognize objects and plan and anticipate actions as part of being "situated in the world" – all of which can be shown to involve knowledge (they are cognitively penetrable by, among other things, what one reads in the paper or overhears on the street) as well as knowledge-independent operations provided by the mechanisms of early vision.

In a rather different vein, **Latimer** questions the value of any simple binary distinction, such as between perception and cognition, and cites Newell's critique of research aimed at testing certain binary oppositions. Newell's (1973) essay, which I consider extremely important and meriting the wide attention it has received, provides a critique of a particular piecemeal research strategy widely practiced in psychology, not of the existence of distinctions (Newell's SOAR approach embodies lots of distinctions). Newell correctly argued for a more top-down research strategy in which the development of larger-scope theories guides the narrower experimental research questions one asks. I agree with that, but in the meantime I see the strategy of running together processes that we have reason to believe are separable as a bad starting point (and insofar as SOAR does that with respect to an implicit assumption of a uniform learning mechanism, I have been critical of it: Pylyshyn 1991). Of course at first glance everything *does* appear to depend on everything else – the world is not designed for the convenience of the scientist. Nonetheless, we will understand complex processes only if we can discover their parts and the constraints within which the parts interact. Here I refer Latimer to an essay by Newell's longtime collaborator, Herb Simon, who provided a very nice argument why complex systems have to be organized (and understood) as "partially decomposable systems" (Simon 1969a).

Dawson & Piercey make a similar point with regard to the methodology for uncovering how complex systems work. They, like Newell, favor larger-scale theoretical systems cast in the form of computer programs. With reference to the question of the independence of vision and cognition they state that "most current theories in perception are not detailed enough to address this basic issue." The value of the approach they (and Newell) advocate is indisputable. But the fact remains that there will always be decision points in the construction of such theories where the designer will need to consult both the empirical data and the arguments made by those who think that vision can (or cannot) be influenced by one's beliefs and expectations and will need to confront just the sorts of issues I discuss in the target article.

Other commentators do not question the need for the distinction, but feel that there are better ways to make it. For example, **McFarland & Cacace** propose to do it in terms of modality specificity, following their study of modality-specific information processing disorders. Although there is an obvious sense in which vision and cognition differ – in that vision is specific to the visual modality and cognition is in some sense amodal (although the latter claim is itself problematic because the thought "the apple is large, round, and red" has a clear modality-specific content) – the notion of the visual modality is dependent on what one thinks constitutes the visual process. We cannot take the term *visual modality* to apply to any process or representation that originates at the eye because that includes all the knowledge we obtained by reading and watching, so we need a prior notion of a visual process. To do that we can use McFarland & Cacace's idea that "to the extent that some stage of information processing is driven mainly by sensory information, we may conclude that it is perceptual" – and that is exactly what I do in the target article. The bulk of the discussion in the target article tried to pin down what that means and to deal with various objections to it. So, for example, "driven mainly by sensory information" is cashed out in terms of cognitive impenetrability – a notion that was proposed in Pylyshyn (1980) and (1984) as a general means of distinguishing between the properties of cognitive architecture and those of what is represented. The particular way I have tried to distinguish vision and cognition and the criteria I have suggested stem from a general goal I set for distinguishing inference-based (semantically-dependent) processes from functions that we may treat as essentially "wired in" in the organism.

The methodological issues surrounding operationalizing the vision-cognition distinction drew additional comments. **Rhodes & Kalish** discuss their use of Signal Detection Theory (SDT) to distinguish visual effects from post-visual decision effects and suggest that although it *could* in principle be the case that their SDT measure fails to distinguish these two stages, for reasons pointed out by Norris (1995), it is very likely to provide a valid measure in the way they used it (because they chose nonwords to be minimally different from words in their study). Moreover, they suggest that if one cannot distinguish the two stages in this way, perhaps they cannot be distinguished at all. I do not wish to get involved in the debate between Rhodes & Kalish and Norris (who claims to have simulated the detection process and shown that the Rhodes et al. [1993] use of the d' sensitivity measure is contaminated by response bias effects) because my point was simply that SDT cannot serve as a methodological panacea for operationalizing the distinction. The general point about the weakness of SDT in this context is vigorously supported by **Krueger** (who provides additional reasons to shun its use to assess cognitive penetrability) and is confirmed and enriched by the helpful commentary of **Macmillan,** who argues that certain versions of SDT can be useful in assessing cognitive penetrability, particularly in conjunction with an explicit model of exactly how cognitive and sensory factors combine to determine performance. This is a conclusion I can support. SDT is certainly a tool that is helpful when used appropriately and with full knowledge of its assumptions and limitations.

Schirillo also discusses methodology in his commentary and suggests that my concentration on form led me to ignore work on the effect of knowledge on the perception of such simple dimensions as color. Color was in fact one of the earliest of features investigated in relation to cognitive penetrability. But in no case was the measurement of penetration of color judgment by knowledge separated from two confounds: memory and contextual selection. When subjects are asked to compare the color of an object (such as an apple) to the color of a standard Munsell color chip they must (a) focus on the color to the exclusion of other properties of the object, and (b) retain the color in memory as they shift their gaze from the object to the color chip. The one thing we *know* to be cognitively penetrable is memory. Memory is quick to assimilate to characteristically good exemplars (although I suppose someone might be able to factor a cognitively impenetrable stage in memory just as I have tried to do for vision). So it is not surprising that in such comparisons the selected color tends to be one that is characteristic of the object in question. The study by Delk and Fillenbaum (1965) cited by Schirillo does indeed contain much better controls for the memory involved in the actual comparisons. The task, however, is still to match cutouts of shapes associated with the color red (apple, heart, lips) against Munsell chips, so the influence of color recall for objects of the test shapes continues to be a factor (as the authors themselves recognize). If subjects understand the task as that of matching the color that objects of that shape typically have, then memory is the crucial variable, despite the other well-founded controls that were incorporated. True metameric color matching, where the subject simply serves as a null detector without having to attend selectively to color over other properties, and where memory is excluded, is insensitive to just about anything except such physical properties as the spectral properties of the colored object and the incident light, and perhaps the area and viewing distance of the stimulus.

Finally, **Sowden** and **Moore** both note that although I insist that the thesis is about the distinction between early vision and cognition, I sometimes lapse into referring to the relation between *vision* and cognition. They are right: I only introduce the qualifier "early" to avoid unnecessary quarrels over terminology. I still believe that "vision" is the correct usage. When Chomsky refers to the language faculty he means the capacity specific to language, and when we study the mathematical capacity we mean the part that is specific to mathematics. Of course the *exercise* of the language or mathematics faculty inevitably requires the use of knowledge of all kinds (e.g., knowledge of reading, speaking, writing, memory, understanding the appropriate context for the exercise of these skills, and so on). It is perfectly appropriate in a study of the process by which we visually apprehend the world that we focus on what is unique to vision – and that we call that *the visual system* (this, of course, carries with it the empirical hypothesis that there are *some* mechanisms specific to vision). That is what we do with other functional organs of the body (whose function can always be seen as blending in with those of other organs) and vision is no different in that respect.

R3. Top-down effects versus cognitive penetrability

A line of argument one often hears used against the independence thesis, especially from those who focus on the data of neuroanatomy, is that there are lots of top-down influences in the visual cortex (**Tsotsos,** especially the last two paragraphs, **Grunewald, Bullier**). I was at pains to point out (in sects. 1.1, 3.2, and 5) that there are plenty of reasons to expect what have been called top-down effects, wherein analysis that is further along the visual stream from the incoming information plays a role in determining how the stimulus is analyzed in earlier stages. I carefully distinguished this sort of effect from a cognitive effect, in which beliefs, goals, and expectations determine the content of the percept (what things look like). I acknowledged that there is plenty of evidence that processes that are neurophysiologically further up in the visual stream do have their influence on earlier centers. Because early vision, as I understand that term, encompasses a complex and multilevel system it is not surprising that there are interlevel influences within this system. My thesis was that *extravisual* influences, particularly cognitive ones, are extremely limited and are confined primarily to those that can be realized by modulations (mostly attenuation and enhancement of signal strength) from focal attention prior to the operation of early vision, and selection/decision operations, applied after the stage of early vision. Thus in relation to the cognitive penetrability claim, it does not help to point out that there are many fibers leading down from higher stages of vision to early stages such as V1 or MT.

A good example of how the criterion of cognitive penetration is misread by critics is to be found in the commentary by **Tsotsos,** who argues that evidence that an animal can attend to a certain stimulus location (as in the Moran & Desimone [1985] study) shows that "modulation must be logically connected with the organism's goals" and therefore that vision is cognitively penetrable. But this misses the point of the penetrability criterion: Of course *what one chooses to attend to* (and therefore the resulting attentional modulation) is determined by rational cognitive processes – that is the whole point of having an attention mechanism. Indeed the claim of the target article is that attention allocation is one of the two types of influence that account for all the observed cases of cognitive intervention in vision (the other being postperceptual selection and decision-making). My thesis (as Tsotsos correctly quotes it) is precisely that cognition only affects perception by determining where and to what (and perhaps the degree to which) attention is focused. It does *not* in any more direct way "alter the contents of perceptions in a way that is logically connected to the content of beliefs, expectations, values, and so on." After quoting this criterion for what constitutes cognitive penetration, Tsotsos refers to it as an "assumption" and charges that the assumption conveniently "dismisses the evidence that makes [my] thesis impossible." But nothing is being "assumed" in providing this criterion. My point was merely to clarify what I mean by "cognitive penetrability" (a concept that is spelled out at much greater length in Pylyshyn [1984]) so as to rule out cases of indirect influence (as when our cognitive state alters our perception in a non-content-related way by causing us to put on our glasses) or cases of intravisual top-down effects that are not cognitive. The important Moran and Desimone (1985) study that Tsotsos cites shows that an animal's behavior, as well as the action of single cells, can be trained to respond selectively to activity at a particular site within its receptive field, providing that the site in question is being attended. This particular example is a clear case of location-based focal attention that I discuss in section 6.4 and elsewhere (though, as I mention in the target article and below, location is not the only basis for focusing attention). Similarly, the intriguing study by Chelazzi et al. (1993) (also cited by Tsotsos) shows the effect of attentional modulation, only this time directed at the location of the to-be-foveated stimulus. Both the Moran and Desimone and the Chelazzi et al. findings raise interesting questions about the neurophysiological mechanism by which attention can be focused – especially to the targeted location of a future movement. But once again, from the perspective of the target article this is a perfectly straightforward case of attentional selection operating via a preperceptual gating process.

Incidentally, one must be careful when making assumptions about what constitutes a cognitive or a visual input. For example, there is no reason to assume that a delayed stimulus must operate through a cognitive encoding, as **Grunewald** appears to assume. The "decision to keep something in short-term memory" *may* be a cognitive one, as Grunewald states, but information may also be held in a noncognitive buffer. Visual inputs, though they are often thought of as static images, are not equivalent to instantaneous retinal patterns. A delayed match-to-sample task may be mediated by either a temporally-persistent visual input or, what seems even more likely, by a persistent locus of attentional focus, without involving a cognitive memory system in either case.

Although most cases of top-down attentional modulation arise from within early vision, there are some exceptions. In section 7.1, I discussed some special cases in which tactile, proprioceptive or vestibular input can have a modulating visual effect, and I suggested that the boundaries of early vision may for some purposes be taken to include aspects of other *spatial* senses. Thus the example that **Tsotsos** cites of influences from the tactile modality (Haenny et al. 1988), has no bearing on the basis thesis of cognitive penetrability, though it may require us to rethink the precise boundary of what can serve as inputs to the early vision system. **Grunewald** agrees with my suggestion that the encapsulated early vision system has many of the characteristics of a "spatial system." He points out that the convergence of auditory and visual signals need not occur at a central level, but may be as early as V4 and IT for tactile and auditory convergence with vision, respectively. However, Grunewald also makes an important point that the association between auditory and visual information need not be spatial *per se,* but can be based on identity. I did not go into this point in the target article, but it is known that attention often appears to be directed to what are referred to in the literature as *objects* (e.g., Bayliss & Driver 1993) rather than to unoccupied locations in the visual field. The idea that cross-modality integration may be mediated by the identification of the same individuals in the two modalities is an intriguing one and is compatible with there being a spatial system in which individuals are located, as in the Kahneman and Treisman (1992) idea of *object files.* This idea is developed in Pylyshyn (1998; in press), as well as in a book manuscript in progress.

Bullier goes into more detail regarding the question of what constitutes early vision and the sorts of modulation that occur in visual cortex and elsewhere. His brief commentary cites the most interesting and challenging of the neuroscience data, although his opening statement that most of the evidence I present is from "experimental psychology" rather than neuroscience betrays a value judgment that is very far from justified. For example, when he raises the question of what constitutes early vision he clearly wants an answer in neuroscience terms. Such an answer may not be forthcoming for many reasons: The state of neuroscience may not be up to finding the biological substrate (which we all believe exists) corresponding to the functional taxonomy most relevant to understanding the information processing task of vision. Because of this it is problematic to find evidence for the penetration of early vision by looking for modulating activity within cortical areas assumed to correspond to early vision. As I suggested above, early vision could be widely distributed (presumably, though not inevitably, within cortex). Putting this problem aside, however, one can still accept many of Bullier's claims. There is not only neuroscience evidence but also psychophysical evidence that visual activity can be modulated by such stimulus properties as the direction of movement, shape, and color of the stimulus as well as by premovement motor preparation (which Bullier calls "significance of the stimulus for subsequent behavior"). None of these present any challenge to the impenetrability thesis because we know that attention can be focused on the basis of such physical features. Binocular rivalry has been cited as a challenge because it has been alleged that the selection among rivalrous percepts is (a) based on semantic properties of the images (e.g., plausibility), and (b) effective very early (prior to stereo fusion). But for such cases to represent a challenge one would need independent evidence that the selection was not taking place after stereo fusion and after *both* rivalrous percepts had been constructed by early vision. Such a possibility is not addressed by the single-cell work cited by Bullier because, as he himself recognizes, the neuronal firing patterns in inferotemporal cortex, which change when the percept flips, may be modulated either from below or from above. In the former case it could be governed by focal attention and in the latter case it is compatible with the selection between percepts being made postperceptually by a cognitive selection mechanism.

Bullier's central claim is that one ought to look at the temporal dimension to get a clearer picture of cognitive penetration. There one finds evidence that some neurons in V1 are activated later than some neurons in frontal, parietal or inferotemporal cortex. Moreover, certain neurons in primate prefrontal cortex can be modulated at very short latencies by a signal that informs the monkey whether or not the receptive field of that neuron will be the target of an eye movement. Bullier takes this to show "the capacity of the decision system to influence *very* early vision." This is indeed an interesting phenomenon, but it need not constitute a *direct* influence of "the decision system" as opposed to one that is mediated by focal attention. All it shows is that there is anticipatory activity in the relevant cells that immediately precedes an eye movement to that area (a phenomenon also reported by Duhamel et al. 1992). But we know that before an eye movement takes place attention is shifted to the target location, so it is reasonable to suppose that what we are seeing in these studies is the effect of the attentional shift that precedes the eye movement.

Finally, there is a belief in some quarters (e.g., **Rhodes & Kalish**) that neuroimaging data showing the activation of visual areas such as V1 by mental imagery demonstrate that cognition penetrates vision. Let us put aside for another occasion the serious problems of interpreting this sort of activation as being the basis of mental imagery. There remains a big gap between showing that there is activity in a certain area during both vision and mental imagery and showing that the activity represents the influence of cognition on perceptual content. In view of the behavioral evidence I have cited in favor of the view that cognition does *not* affect perceptual content, jumping to such a conclusion on the basis of this neuroimaging data in the present state of understanding, as many people do, is unwarranted, to say the least.

R4. Where is the boundary of early vision?

R4.1. Does consciousness mark the boundary between vision and cognition? My main thesis was that there is a process I call early vision that is not only impervious to cognitive influence (outside of that mediated by focal attention and postperceptual selection) but also encompasses a significant part of the process that is proprietary to vision – that is, not shared with other processes such as reasoning. This means that it reaches fairly "high" toward the cognitive interface. But as to what exactly it encompasses, this remains an empirical question for which we have only provisional suggestions. Whatever the boundary of early vision, I tried to make it clear that, unlike **Cavanagh, Bermúdez,** or **Gellatly,** I do not believe it has anything to do with what is consciously experienced. Lots of consciously inaccessible properties of the world are visually detected and are effective in influencing behavior (Nisbett & Wilson 1977), and lots of what seem like conscious properties of the perceived world are illusory and ineffectual. For example, the world as perceived is not the richly detailed panorama that our phenomenal experience suggests it to be (as the change-blindness work shows).

The concept of consciousness recurs in a number of commentaries. For example, **Nöe & Thompson** discuss my claim about how misleading subjective experience can be in suggesting how vision works. They appear to agree that the purpose of visual experience is "to bear witness to what goes on in the world, not to what goes on in the head" (though they later take this back and say that the purpose is, rather, the guidance of action). But they do not think that the content of conscious perceptual experience "foists on perceivers a naïve model of the subpersonal organization of the visual system." I would suggest that as a matter of sociology of science this is exactly what has happened. Noë & Thompson should try reading the vast literature on vision (for example, in introductory texts), and the even more egregious literature on mental imagery, where the dominant view is that both perception and imagery consist in constructing pictures-in-the-head, which are viewed in what Dennett (1991) calls the "Cartesian theater" – because that's what it feels like (see the excellent critique of this view in O'Regan 1992; for that matter, see my critique of the mental imagery research in Pylyshyn 1973; 1981). Noë & Thompson themselves draw conclusions based on certain beliefs about conscious content, which lead them to say odd things, for example, that "if the output of early vision were conscious, it would be cognitively penetrable" and therefore that I am committed to the view that all outputs of early vision are unconscious. The only conclusion about consciousness I am committed to is that we can make nothing of it for the purposes of building a scientific theory because we have no idea what it is. To equate the distinction between conscious and unconscious with the distinction between personal and subpersonal is to invite such nonsensical conclusions as "the output of early vision is inaccessible to the subject" (Noë & Thompson), which by parity of reasoning would lead one to believe that *nothing* computed by the brain is accessible to the subject. My view is that some things are and some things are not and we have no idea what makes the difference.

Part of **Cavanagh**'s commentary is also off the mark because he equates what he calls "cognitive-style processes" with processes of which we are conscious, or that we can verbalize. (He claims that "what we know" must be understood in terms of what "we can tell another person," which is certainly not the way the term is used in the study of psycholinguistics, implicit memory, animal and infant cognition, or artificial intelligence). Cavanagh agrees with me on what is really the main point of the independence thesis, which is that vision and cognition have access to different data bases, but accuses me of "cognocentrism" in claiming that the processes involved in vision and cognition are different. There can scarcely be any doubt that most of what goes on in early vision is different from what goes on in reasoning. Reasoning, after all, is not concerned with computing the inverse mapping from distal to proximal stimulation or with constructing a conceptual representation from an optical pattern. But one can still ask whether some of the processes are the same. I claim that they are different at least in the respect that vision *cannot* (because of the nature of the visual architecture) use information that is both freely available to cognition and relevant to the perceptual task at hand (such as judging which of two lines is longer or whether two lines are collinear). (Incidentally, because of such architectural constraints I claimed that vision does not carry out "inferences." Here I use the term in a way that is confined to what Stich [1978] has called "inferentially promiscuous" or unconstrained interactions among beliefs. The term could clearly be used more broadly [as in the commentary by **Singh & Hoffman,** as well as in Fodor & Pylyshyn 1981], but nothing substantial or relevant to the independence thesis hinges on this choice of terminology.) The systematic restriction on what knowledge can be used makes visual processes appear "irrational" in many circumstances (I am not the first to remark on this apparent irrationality of vision – in this I follow the example of both Kanisza and Rock). This is not the same as people failing to use something they know when they are solving problems or making decisions (as when they tolerate incompatible beliefs, perhaps – as Cavanagh says – "by choice"), because such failures are invariably not architectural or even principled (e.g., unlike the visual system, people generally acknowledge errors of omission when reminded, and if they do not, it is usually for practical cost/benefit reasons).

Gellatly similarly makes much of the conscious-unconscious distinction and says we have to be clear about whether we are talking about the conscious experience or the information processing operations. I would have thought it was obvious that I am concerned only with the latter, which may or may not be conscious (this is shown by what Gellatly calls my "mixed bag" of examples, which are all clear examples of cognitively penetrable inferential processes, most of which are unconscious). Gellatly appears to agree with me in most of his diverse points. For example, he agrees that vision is an organ much like the heart (rather than like the entire brain). He also agrees that even in clear cases of inference people cannot always state their premises. But then he asks, why should they be able to do it in vision? Exactly so, but the issue has nothing to do with what observers can report; it has to do with what sorts of information can alter what they see and in what way this alteration is achieved.

R4.2. Some border issues. A more interesting set of questions concerns where the early vision system ends and cognition begins. I was admittedly speculating when I suggested that it included the three-dimensional layout of visible surfaces, perhaps articulated into individual objects. It seemed to me that there was evidence for that view, beginning with Marr (1982) and notably including the work of Nakayama et al. (1995) that I cited. Many of the commentators provided additional suggestions. Whereas many felt I had admitted too much into early vision, others (e.g., **Cohen & Kubovy, Hollingworth & Henderson, Krueger**) suggested that I had not gone far enough. Still others provided useful suggestions for refining this boundary (e.g., to include some but not all depth, orientation, and aggregation information concerning visible surfaces) or suggestions based on the approximate nature of quantitative spatial representations (**Aloimonos & Fermüller**) or the availability of alternate surface percepts once their existence is pointed out (**Papathomas**) or the piecemeal way spatial information is assimilated (**Peterson**). I will try to comment on some of these suggestions later, because they represent useful additions and modifications to my own speculations.

Several other writers commented on my proposal that early vision may deliver a representation of the surface layout. **Edelman** argues that the proposal that surface representation is obligatory is undermined by the fact that his model of recognition and categorization manages to work "without surface representations as such." However, other models do assume the availability of what Marr (1982) has called the 2.5-D sketch and the large amount of evidence cited by Nakayama et al. (1995) makes such a proposal highly plausible. Edelman later softens his position, admitting that "the issue of explicit representation of surfaces is still undecided" – a conclusion I can support. As I said earlier, the exact nature of the output of early vision is not firmly established, but it seems clear that it is at a much higher level than early accounts had assumed, and a very long way from the traditional "sense data" view. And by the way, the claim that early vision outputs such high level constructions as a surface representation does not entail that one should not also be able to see dots and edges (as Edelman seems to believe). An encapsulated and opaque early vision system that computes surfaces or even object-shapes (see sect. 4.3) can also compute edges and dots and even shading. It can do so either separately, when attention is focused on such properties, or it might compute them independent of or even after surfaces are derived, as has been suggested is the case for some illusory contours (Nakayama et al. 1995). The latter explanation is also supported by evidence that it takes longer to perceive many primitive constituent features than to perceive objects of which they are composed (the so-called object-superiority effect that parallels the word-superiority effect, wherein recognition of words is faster than recognition of their constituent letters). In any case the vast majority of image properties to which the visual system can be shown to be sensitive (and which are used to compute higher level perceptual constructs) are themselves not normally available to cognition. For example, the visual system is sensitive to derivatives of luminance and to statistical properties of textures under conditions where people do not "see" such properties directly (i.e., where such properties are not explicitly output by the visual system). Also some image properties, examples of which are mentioned in the target article, may only be available to the motor system. And finally, even if some intermediate stages in visual analysis were available to cognition, it would not be damaging to the particular form of the impenetrability thesis I was advocating in the target article, where the main claim went in the other direction – it concerned the insensitivity of the visual process to cognitive states.

R4.3. Does early vision compute a canonical shape-description? There is one interesting possibility for what early vision constructs that may be emerging from a number of commentators who discussed this issue. It is the idea that early vision may compute what might be called a shape-class or shape-category, which functions somewhere between being a description of visible surfaces in depth and the recognition of a known object. It is an idea that I was groping for in my section 7.2 and note 14. Suppose that the architecture of the early visual system assigns visual objects (segregated regions of the visual field) to equivalence classes based on what we might roughly call their (qualitative) appearance. Suppose one of these classes is assigned the data structure (or label) Q374. (We give it a neutral atomic name to emphasize that the name does not provide a direct means to access knowledge about members of this class, the way a word in our language would.) What this label *could* allow some decision stage to do, however, is to determine very simply whether two such objects are tokens of the same type. So a person could easily decide that two such visual objects correspond to the same kind of thing (e.g., that they are both human faces or cars or members of what Rosch et al. [1976] would call the same *basic object* or *natural category*). Note that identifying members of such general categories need not involve a memory lookup, the way you would need to consult memory to identify a particular individual token that you know about, such as "my friend Mary Jones" or "my 1986 Honda Accord." So long as the category remains shape-based it could be a consequence of the way the visual system is wired up (or was rewired over time with experience), because, as I noted in the target article, any many-one mapping induces a set of equivalence classes. Of course to make the judgment that the object before you at the moment is one you have seen before, some memory is required. It does not require a cognitive memory or a knowledge base, however; only a record that you have seen, say, Q374 before.

Armed with this idea, let us now look at the interesting suggestions in the commentaries by **Hollingworth & Henderson, Papathomas, Peterson** and **Rosenfeld. Hollingworth & Henderson's** proposal is that a "presemantic" match between the description generated in the course of early vision and memory may take place within early vision itself, allowing access to information about the object type. They base this proposal on the finding that when we control for the effect that context might have in focusing attention on certain locations, recognizing an object is not aided by the semantic congruence between the object and its scene context – it is just as easy to recognize a toaster or a chicken, whether they are shown as being in a kitchen or in a farmyard. Because semantic congruity does not help recognition, recognition itself cannot be mediated by semantic factors or by such semantically-sensitive processes as reasoning. However, this result is compatible (as Hollingworth & Henderson recognize) with the idea that the early visual system simply outputs what I have called a shape category (which I assume is very close to their notion of "presemantic" recognition). On the other hand, as I have argued, such a presemantic categorization need not involve matching anything to memory. It could simply be the structured description corresponding to an equivalence class induced by early vision. Of course recognizing this class (e.g., Q374) *as* a toaster (with all that this implies about its functional and conventional properties) *does* require making contact with a knowledge base, which requires not only lookup but perhaps various inferences, as well (e.g., that it is usually found on the kitchen counter), and therefore this sort of recognition lies outside the province of early vision. I agree, therefore, with Hollingworth & Henderson's proposal, but do not see it as requiring memory lookup except when it is associated with a judgment such as that the object is something that has been seen before or that it has a certain name. Once an appropriate canonical shape-description is constructed it may be a relatively simple matter to match it to a knowledge base of names. In fact what the Hollingworth & Henderson experiments show is that the step between achieving the shape-description and locating its name is, at least in the case of simple objects, fast and direct and not enhanced by semantic congruence or degraded by semantic incongruence of the context.

Rosenfeld's comments are also relevant to this idea. He suggests that in object recognition, as opposed to what he characterizes as more general visual tasks (such as apprehending an unknown scene or the control of motor actions) the content of "object memory" may have to be consulted. But as Rosenfeld recognizes, the problem of looking up images in a large database of object models in model-based computer vision systems remains computationally intractable. The secret for dealing with this problem comes from an insight to which Rosenfeld has himself contributed significantly with his notion of a generalized (or general-purpose) model (Zucker et al. 1975). A generalized model is very much like what I have called a shape-class. It greatly reduces the computational load of indexing because the computation of such a shape-class by early vision does precisely what Rosenfeld says needs to be done, namely, "rapidly extract from our visual images features that have sufficiently rich descriptions to be useful for indexing into long-term memory, so that only a few of the many possible object classes become candidates for verification." **Sanocki** makes the same point when he argues that the use of lookup tables in early vision would lead to intractable complexity. My claim in the target article is that efficiency is achieved precisely when the use of canonical shape descriptions reduces the set of lookup candidates (in Rosenfeld's sense) to very few objects and that therefore the decision among them can be confined to the stage after early vision has completed its work.

As I understand **Rosenfeld**'s proposal it differs from mine only in that (1) he, like **Hollingworth & Henderson** and **Sanocki** think of the process as requiring a match against models stored in memory, whereas I find that way of talking unnecessary for reasons given above, and (2) he views the process of accessing object memory as involving "hypothesizing" that an object of a particular class may be present in the image. I do not disagree with (2) but see it as occurring in a post-early-vision stage, inasmuch as the empirical data I examine suggest that it does not, as Rosenfeld surmises, "involve cognitive penetration into the early vision system itself" or that "finding initial and confirming features may be expectation-dependent." At least with regard to human vision, these are empirical questions to which the answer appears to be that they do not operate exactly the way Rosenfeld's "broadcast" model suggests (though rather close to it).

R4.4. Does cognition select from among possible interpretations provided by early vision? The idea that early vision delivers not the unique percept but a (small) *set* of canonical three-dimensional shape descriptors among which cognition chooses (as has been convincingly demonstrated with distinct senses of words in the case of lexical ambiguity [Swinney 1979]) also helps explain why we sometimes appear to be able to select voluntarily among ambiguous perceptual interpretations, as in **Papathomas**'s examples. Although Papathomas's example is intriguing, I do not see why post-perceptual selection cannot serve as the mediator in that case as it does in other similar cases I cite in the target article. In the typical case of this sort, early vision may deliver one or two possible interpretations of an ambiguous figure and the cognitive system then selects among them. In the case of ambiguous visual figures (such as a Necker cube) the mechanism for enhancing and maintaining this selection in most cases appears to be clear: The observer directs focal attention to the local cue that favors one or the other interpretation. This mechanism also appears to be the appropriate means of searching for and maintaining the desired perspective once one is located in the output of early vision. I might add that although I know of no data in vision that exactly parallel the Swinney finding with lexical ambiguity, there is evidence that different aspects of a figure are integrated over time (e.g., "amodal figure-completion" takes place over time – Sekuler & Palmer 1992), so it could well be that possible analyses are available for brief periods before considerations of global consistency eliminate some of them. (This would be parallel to the Swinney finding that different readings of an ambiguous word are available for a short time before the context eliminates the semantically anomalous one.)

The idea that early vision provides alternatives and that cognition can select among them also explains **Peterson**'s finding that under certain conditions one can intentionally "hold" one particular interpretation once it is achieved. An obvious mechanism for implementing this "hold" is focal attention. Indeed, Peterson's own examples show that attentional focusing does just that: The percept that emerges in the figures that she and Hochberg studied in their fascinating work (Peterson & Hochberg 1983) depends primarily on the local cue at the locus of attention, and not on the whole figure. (I am puzzled about why Peterson concludes that "intention" penetrates vision, because I frequently cite her elegant work as strong evidence that focal attention is the mechanism by which selection of a particular percept is achieved. Indeed, Peterson & Hochberg [1983] themselves appear to agree with this view when they conclude from their study that "the interactions obtained indicate that where the subject is instructed to look helped determine what is reported as being perceived, even under instructions to hold one or the other organization" [p. 187].)

Thus the idea that early vision delivers a small set of alternative shape-based perceptual options from which a selection is made, often by the mediation of focal attention, is consistent with most of the commentators' proposals. I might also add that the idea of computing canonical shape-classes is far from being a new idea. Computer vision research devotes considerable attention to developing systems for canonical shape descriptions, such as generalised cylinders, superquadrics, deformable models, and geon-based descriptions, all of which are computed from the visual information without reference to a general knowledge base. Within the philosophical theory of concepts, there has been a discussion of the idea that there may be a "preconceptual" form of perceptual representation (see the commentary by **Bermúdez,** or Peacocke [1992, Ch. 3]), or even appearance-concepts that are more complex than the usual sorts of sensory concepts (red, moving, and other such "sensations") as suggested by Fodor (1998, p. 135). There is also evidence from neuropsychological studies (involving both clinical cases and neuroimaging studies) for a dissociation between "perceptual categorization" and "object recognition" (Warrington 1982; Warrington & Taylor 1978).

Aloimonos & Fermüller's excellent addition to my analysis of early vision is another example where a plausible way to view what is going on is in terms of post-perceptual selection and integration. These commentators argue that the spatial information provided by early vision is metrically approximate and indeed fails to meet precise Euclidean axioms. Since, however, we are able to deal with space in a rather precise manner, the output of early vision needs to be augmented by cognitive factors relating to certain special tasks that we undertake using the spatial information. It is true that we are accurate when dealing with space in some tasks (e.g., reaching, locomoting, and recognizing such objects as faces by their shape) and poor in other tasks (giving directions). Aloimonos & Fermüller suggest that the visual system generates several descriptions of the spatiotemporal environment, rather than one general one and "as they make sense in the context of a class of tasks, they cannot be cognitively impenetrable." I do not see the reasoning here. I have already argued repeatedly that an appropriate role for early vision is to provide alternatives among which cognition selects, so there is nothing here that violates cognitive impenetrability. But, more important, the observation that early vision may generate several different forms of output that can be used in different tasks is exactly what one would expect if the visual system contained submodules of the sort already alluded to above and in the target article. For example, Gallistel (1990) has shown that there is a geometrical module in the rat and other animals (recently extended to apply to humans; Hermer & Spelke 1996) and that certain skills in some animals require extremely accurate measurements of distance and time – measurements accessible to navigation but often not to other tasks. These sorts of submodules of vision and visuomotor control are therefore common and very much in the spirit of independence thesis because none of the submodules appear to be cognitively penetrable, even though selecting among their conflicting outputs may be.

R5. The nature and role of attention

Most commentators had some remark to make about focal attention, which is understandable, given the general role it plays in interfacing early vision with cognition. But many have also pointed out that in my target article I require this mechanism to do a great deal despite the fact that there is much that we do not understand about it (e.g., **Pani** asks: "What is the relationship between the many phenomena of spatial organization . . . and visual attention?"). This is true, but there is also much that we do understand about it. For example, one relevant thing we understand is that perception of ambiguous patterns can be affected in predictable ways depending on where attention is focused (**Peterson**) and that, in turn, can explain the selection phenomena I discussed in section R4.4 above.

Sowden suggests that even though clearly mediated by focal attention, what he calls early visual learning (EVL) represents "a form of long-term, indirect cognitive penetration of early vision." Once again, my only disagreement is with terminology. I have said that visual perception can be altered over time by shaping the early vision system in certain ways (usually a task-specific shaping induced by practice). Visual expertise can also be acquired by learning how to allocate attention. Between these two effects we can achieve a variety of sorts of perceptual learning. If Sowden wishes to refer to either of these forms of learning as cognitive penetration, he is free to do so. But then he will need a different term to refer to what I have been calling cognitive penetration. The point is that there is a distinction that needs to be made between the sort of effects we might get when we find out something about what we are looking at (e.g., that in that particular situation we are more likely to see an X than a Y) and the indirect sort of effect that is mediated by learning where or to what we must attend. It is not the terminology that matters; it is the distinctions we make. In this case the question is whether visual perception can be altered in roughly the way we alter our decisions as a function of being told something we did not know before (together with our utilities). The New Look and most psychologists assume it can; hence the target article.

The simplest view of visual attention (the first-order view) is that it allows us to select information from a limited spatial domain – that it is spatially focused. But this, as several people have pointed out, is too simple a view. In fact there has been interest recently in object-based attention, wherein attention is focused on individuals regardless of where they are at any particular time (Baylis & Driver 1993). Indeed, in my laboratory we have shown that it is possible to pick out and keep track of up to four or five objects moving independently in a display and to do it without keeping track of their distinct locations (Pylyshyn & Storm 1988). As I pointed out in the target article, there are a number of other priorities of a stimulus that can be used as the basis for selective attention, though they must remain a highly constrained basis set to avoid circularity (see Fodor & Pylyshyn 1981).

Yeh & Chen suggest that attention may operate at multiple stages and sites within the visual system. Citing neurophysiological data, they argue that if attention operates prior to early vision, that poses a problem for where it is located, because attentional modulation has been found in MST, PPL, and TE. As I said in the target article, it is difficult to locate early vision and attentional selection neuroanatomically because the correspondence between a functional and an anatomical taxonomy is unknown. Early vision consists of many submodules (e.g., computing motion, color, stereo, etc.) and may therefore be anatomically distributed over a variety of brain loci, with attentional modulation similarly occurring at many different sites, as Yeh & Chen suggest. In other words, it could be that attention is indeed distributed among cortical areas without violating the principle that attention takes place prior to early vision. But we must not be concerned by the observation that what seem like complex functions are subject to attention, as in so-called object-based attention (Baylis & Driver 1993). Despite their complex appearance these functions may in fact be computed very early. Our own work suggests that object-based attention is an extremely early mechanism that operates after a small number of primitive visual objects are segregated, individuated, and indexed (by a mechanism known as a FINST Visual Index) but prior to any of their properties (including their locations) being encoded. Indeed, we have hypothesized that multiple object individuation and indexing are preattentive processes that not only occur prior to early vision, but even prior to the allocation of focal attention (Pylyshyn 1989; 1998).

Egeth presents some additional supportive evidence for the view that focal attention operates early in vision. He cites some of his own work in the 1960s showing that a "mental set" induced by some cues works better for purposes of early perceptual selection than the set induced by other cues – in particular that picture cues are better than cues based on meaning or category membership in enabling early selection. Egeth concludes that what makes pictures better cues in such experiments is that they provide information about where to look when the stimulus is briefly presented and that this sort of location information plays a special role in selection.

Cohen & Kubovy take issue with an incidental remark I made that one of the functions of attention is to allow the encoding of conjunctions of properties, as in the conjunction search experiments of Treisman (1988). Cohen & Kubovy provide an alternative explanation of the conjunction search results that assumes that the stimuli used in conjunction search are "perverse" in that there are two ways to segregate the stimulus boundaries that lead to different item parsings and it is this ambiguity, and not feature integration, that requires focal attention to resolve. Thus their view is that attention "is called in only when normal preattentive processing is in trouble." I will express no views about this interesting proposal, except to remark that it is consistent with the idea, expressed above, that one of the earliest mechanisms prior to early vision (which we call a FINST index) has the function of individuating and indexing objects without encoding their properties.

Schyns also points to attention as the source of what he takes to be cognitive penetration. He asks whether "attending to the visual cues that solve a categorization task can change the actual perception of the stimulus." Clearly there is a sense in which it can: The perception of a stimulus can be changed by many things – looking away, squinting, and attending to one aspect rather than another. For example, in **Peterson**'s figures whether or not the figure reverses is determined by where we focus our gaze. And because we can focus on a particular spatial frequency band we can determine how we see a set of letters arranged to form a large T; it may appear to be a T or the letters that make up the T, and we are "set" for one or the other by an attentional mechanism (Pomerantz 1983). But these are clearly not the sorts of attention sets that constitute bona fide cases of cognitive penetration (recall we are concerned with whether the content of our beliefs affects the content of our perceptions in a logically explicable way – whether we see what we expect or want to see). What about the cases that Schyns cites? They are all cases in which "different categorization tasks can flexibly use the cues associated with a different spatial scale" or in which people "only perceive the information associated with the spatial scale that is the basis of their categorizations." I discussed such selective use of features that are "diagnostic of a categorization" in section 6.3. So long as the features by which the selection is made can legitimately be the basis of attentional filtering (e.g., spatial frequency, but not true versus false sentences or genuine versus fake Rembrandts), focal attention can provide the mechanism for this sort of cognitive influence, as proposed in the target article. Thus the Schyns examples are a clear case of this sort of selection. And claiming that such filtering can be viewed as part of early vision itself misses the point of the factoring of vision into an early vision stage, as I explained in paragraph 4 of section R2 of this response. As for Schyns's claim that "allocating attention" "is always an *explanans* and never an *explanandum,*" I say that one person's *explanans* is another's *explanandum.* If I explain how perceptual learning occurs by showing that it is done by learning what to attend to, then attention is an *explanandum.* If I want to explain how attention works, then it is an *explanans.* The only worry that I should have is that it not be circular, as it is in such (Gibsonian style) explanations as those in which perception of, say, someone's face, is said to consist merely of attending to (and "picking up") the invariant properties that specify it. If *what you can attend to* is appropriately constrained (and empirically supported), then referring to what you attend to can be a perfectly legitimate explanation of why you classify a stimulus one way rather than another.

Incidentally, Schyns's empirical findings are very close to those discussed by **Kolinsky & Morais,** though they draw a quite different moral from them. Kolinsky & Morais discuss the sort of perceptual learning that occurs with explicit schooling. Sometimes, perceptual discrimination can be shown to require analyses over units (e.g., phonemes) that are themselves initially unavailable as outputs of a perceptual system, yet which can, with training (e.g., after learning a written alphabet), become objects of perception. Kolinsky & Morais give other examples from their work showing that with instruction and practice, certain fused dimensions can be dissociated and, presumably, given an explicit code. In such cases the new code can be mapped onto some overt task (such as spelling) and it can also result in the generalization of certain visual discrimination skills, as when learning a system of writing generalizes to detecting subtle embedded parts of visual patterns, and learning a writing system that requires distinguishing mirror image letters generalizes to an unrelated mirror-image discrimination task. Although the phenomena are closely related to those that lead Schyns to claim that perceptual learning involves the creation of "new" features, neither I nor Kolinsky & Morais would put it that way. Rather, it seems to me that their results show a case of learning to selectively attend to relevant perceptual dimensions. Of course what constitutes an attendable (and therefore learnable) dimension is tightly constrained by the architecture of our early vision system and the principles by which it may be (noncognitively) modified over a period of time.

Rhodes & Kalish also remark on my use of the mechanism of attention and raise the concern that attention may hide a homunculus because they see it as having too much computational power. On the contrary, attention is an extremely primitive mechanism that provides selection by only the most elementary properties, such as location and a few other primitively transduced properties. As I was careful to point out, if the properties we allow to be the bases of attentional selection are not severely restricted, attention can be used as a way to trivialize the problem of perception (as we claimed happens in Gibson's theory – Fodor & Pylyshyn 1981). And a mechanism that can recognize a property specifiable as a template or a table-lookup cannot be said to have any computational power itself, contrary to Rhodes & Kalish's claim. Of course if there are no restrictions on how it can be composed, a template-matcher can be used to implement any computational mechanism we like, including a Turing machine, but only in the trivial sense that a simple on-off switch, or even a wire, can – all we need to do is put together an (arbitrarily large) set of such mechanisms in the right configuration.

R6. Comments on the role of natural constraints

Several commentators who found the general theme of the target article congenial nevertheless had some concerns about my appeal to the notion of natural constraints (first proposed by David Marr). **Dannemiller & Epstein** claim that natural constraints, such as those I mentioned in section 5.1, cannot account for a number of complex visual phenomena involving motion and shape. They cite the example of the perception of curvature of nonrigid shapes and the ability of certain cues (such as the presence of a visible aperture) to override the rigidity constraint. There are in fact quite a large number of examples where the rigidity constraint can be overriddden, as I mentioned in my note 12. Clearly, individual constraints can be overridden when they occur in combination with conflicting cues (as **Vallortigara** also pointed out) and equally clearly we have a great deal to learn about how conflicts can be resolved and how constraints may combine. Constraint-based explanations cover only a small part of the landscape of how one can solve the inverse-mapping problem. In fact I cite them only because they represent an example of how certain properties of the physical world appear to have been compiled into the machinery of early vision. In this respect I agree to some extent with the comments of **Yu** who suggests that "the most stable, reliable relationships" are the most likely to be built into the visual architecture because these allow perceptual processes to gain efficiency by capitalizing on redundancies in the perceptual patterns (as Barlow also noted – see my note 11). But notice that *all* the lawful physical regularities of the world (as well as the biological and social regularities) qualify as "reliable, stable relationships," yet most of them are *not* built in (as I pointed out in the final two paragraphs of sect. 5.2). What is built in tends to be only the optico-spatial regularities of the world. The evidence I have reviewed suggests that what Yu refers to as generic semantic knowledge does not in general function the way optico-spatial regularities do. In particular, ones such as in her example – the tendency of apparent motion to be seen in the direction toward which an object appears to be facing – are quite plausibly mediated by attentional orientation, given that we tend to move our attention (and sometimes our eyes – see Kowler 1989) in the direction in which we expect objects to move.

Note that the invocation of natural constraints is quite different from the appeal to contextual effects. **Sanocki** wonders why we can accept that natural constraints are a legitimate form of explanation, even though they can occasionally lead to failure of veridicality, but we do not accept that there are context effects because of evidence such as that of **Hollingworth & Henderson.** He suggests that context effects should be accorded the same status as natural constraints because in highly valid contextual environments they, too, can have an effect, even though they fail in other situations. Putting aside the questionable empirical claim that highly valid contexts work differently from other context effects, the situation is quite different in these two cases. A natural constraint is a hypothesis about the structural regularities that are claimed to constrain the possible inverse mappings computable by early vision (much as Universal Grammar is claimed to constrain the set of humanly achievable languages). Natural constraints always hold (although they can be overridden by other constraints and they sometimes lead to invalid interpretations) because they are part of the architecture and are not tailored to particular situations. By contrast, context effects are not structural – they depend on the semantic coherence of particular object-context relationships. Because "semantic coherence" is a logical relationship it involves inference from general knowledge (e.g., that a toaster belongs in a kitchen and not on a roadway). So if a context effect fails to influence perception, it requires an explanation for why relevant knowledge was not brought to bear.

Singh & Hoffman accept the importance of natural constraints but prefer to view them as involving a restricted type of inference. As I remarked earlier, I have no problem with this terminological policy so long as one recognizes that the inferences involved in vision are structurally prohibited from using certain kinds of knowledge of the particular stimulus being perceived (see also my response to **Cavanagh**). Describing the computations carried out by early vision in Bayesian terms is perfectly reasonable and has indeed led to some useful insights. The crucial point that is relevant to the present thesis is that the Bayesian formulations, though they can be viewed as probabilistic inferences, are cast over optico-spatial properties and do not take into account such things as the probability that what one sees is an X, given that one knows from general considerations that Xs are likely to be at this place at this time. In other words if we are modeling a part of a theory of early vision, we do not base our Bayesian computation on such things as the probability that what we see is a mouse scurrying about, given that we were told that there are mice in the house. Although there is obvious utility in such a method, it remains part of a post-early-vision recognition system, which may also use a Bayesian inference system, though this time taking into account everything we know.

Singh & Hoffman also propose that we view the notion of "veridicality" of perception in terms of utility rather than resemblance and point out the evolutionary value of taking that approach. This seems perfectly reasonable to me because I do not endorse a notion of veridicality that is related to "resemblance" (indeed I have no idea what that means – resemblance, like "depiction" is a problematic term that involves a relation to a perceiver, as in "A resembles B for observer O"). Veridicality means *correspondence* and the correspondence between an organism's representation R and the represented world W can either be defined in terms of a semantic theory or in terms of a weaker notion of equivalence, such as "R is equivalent to W with respect to the observable outcome of operations that O can perform on W," which, if I read them correctly, is what Singh & Hoffman propose. Although this sort of pragmatic theory of what amounts to visual semantics has serious problems in the long term, none of them bear on the present thesis.

R7. Subsystems of vision, other modules, and other modalities

There is a difference between what I have called the Early Vision system and various submodules within this system. **Vallortigara** is quite right that with such early vision subprocesses as those responsible for computing stereopsis and occlusion, as well as those that deal with color, motion, shape, three-dimensional shape-from shading, and so on, a major problem is how the operation of these individual and relatively independent subprocesses come to a consensus about the interpretation of the visual information – especially when two subprocesses produce incompatible interpretations. In such cases, as Vallortigara correctly argues, the various submodules must communicate with one another. They may do so by direct communication or by providing multiple outputs that are then adjudicated by a central system. He argues, quite plausibly in my view, that the convergence is likely to occur through a process by which one submodule provides constraints on another, without involving the cognitive system. This view is of course quite compatible with the independence of vision thesis, not only because, as Vallortigara notes, the cross-talk is of a limited nature (and so the systems continue to be "partially independent") but also because these are *within vision* interactions that do not require the intervention of general knowledge about the scene.

The impenetrability of the early vision system is obviously compatible with there being other impenetrable systems. Among those that have frequently been proposed are the language system and the face recognition system (others are the "theory of mind system" [Leslie 1994] and the geometry system [Gallistel 1990]). Clearly, many of these systems interact (or may be nested). Take the example discussed by **Bowers:** the recognition of written words. Surely this process involves both the vision and the language systems, and possibly even other systems. Moreover, it is possible that certain phenomena associated with recognizing words from written sentences reveal properties of a modular system, yet the module in question may be the language module (Fodor 1983), rather than early vision. The question of exactly which computations are performed by early vision to allow lexical lookup are not known (at least not by me). It is my understanding (and apparently Bowers's as well) that there are several streams of encoding: one is in terms of graphemes and another is in terms of phonetic features (Shulman et al. 1978). Because there are dual or even multiple streams leading to word recognition I do not see why Bowers concludes that the independence thesis is mistaken because "the perceptual categories for words are not structured on the basis of geometrical similarity and must hence lie outside early vision." I do not see how any of the evidence cited by Bowers bears on the question of where the locus of category assignment occurs. It could occur in another modular system – the language system – where the lexicon may well be stored (see my note 8). If the evidence were to suggest that there is a modality-specific lexicon, that, too, would be compatible with my thesis (because there is no reason why a graphemic lexicon could not be located in early vision). These options are all open within the thesis I was advocating. Nothing that Bowers has said is inconsistent with the view that phoneme restoration and the word superiority effect are a result of top-down effects within the language system rather than within the early vision system.

Another input system that may be modular and that overlaps with vision is the system for face recognition. **Bruce et al.** offer some important observations relevant to the independence thesis, based on their work with face perception. The perception of properties of conspecifics is one of the most important forms of perception there can be from an evolutionary perspective. Consequently, most species have developed perceptual systems of exquisite precision, not only for detecting members of their species, but also for encoding their significant attributes (such as their emotional states). Some of the properties of this system are shared with vision in general (e.g., one such property mentioned by Bruce et al. is a preference for convex over concave interpretations of surface curvature) and others are unique to face perception (e.g., the automatic orienting of attention in the direction in which the perceived eyes are pointing). Because of the importance of such perceptual functions it is not surprising that much of it has gotten wired into the early vision system (or perhaps constitutes one of the subsystems of vision), that it manifests a variety of special patterns of neurological breakdowns (as in prosopagnosia), and that it exhibits many data-driven properties (such as gaze-direction showing the same exogenous attention-orienting properties as sudden onsets). Perhaps a face is one of the "shape-categories" referred to earlier. Yet, as Bruce et al. rightly point out, it would be an oversimplification to view the early vision system as simply passing on a shape description to a categorical cognitive selection system. The perception of face patterns as faces may be automatic, but in all likelihood so, too, is the perception of the face as being that of "mother" or of the face as speaking certain words or looking in a certain direction, or as showing a degree of anger or arousal. Of course not all features of faces will fall under this sort of automatic recognition (e.g., we are poor at recognizing certain types of emotion) nor is the information always broadcast to the complete cognitive system. It may, for example, be made available to an orienting reflex or to a modular language system or to some other task-dependent subsystem (as I claimed was the case in visuomotor control). Clearly, the situation with both face perception and language perception is special in a variety of ways, most of which are still poorly understood. Whatever the eventual findings, however, it seems clear that much of both language and face perception will turn out to be modular in just the ways that the target article claims, though perhaps with a more subtle and diverse interface to cognition.

Gentaz & Rossetti discuss the discontinuity thesis in relation to haptic perception. They suggest that if haptics were found to be impenetrable, it would lend support to the view that sensory systems in general are impenetrable. They cite the oblique effect in haptic perception, an effect that closely parallels a similarly-named effect in vision (where horizontal and vertical distances are judged more accurately than oblique ones). They then cite evidence for the malleability of the oblique effect by gravitational cues and show that the effect is altered when the arm is supported, concluding that "these observations showed that the occurrence of the haptic oblique effect was influenced by the encoding conditions of manual exploration" and therefore that "all haptic processes that generate the oblique effect are cognitively penetrable." However, I do not see the relevance of these observations to my thesis for a number of reasons. In the first place, "the encoding conditions of manual exploration" are no more an example of a cognitive effect than are the illumination conditions of vision. In the second place, I did not claim that all modalities are impenetrable and that the existence of one that is not would therefore have little bearing on my thesis. In the third place, I doubt very much that one should view haptics as a single perceptual modality (or what Fodor (1983) calls an "input system"). Haptic perception involves a variety of distinct sensors (tactile, kinesthetic, proprioceptive, and perhaps others, as well). Also, as Gentaz & Rossetti themselves point out, it involves a much greater voluntary exploration component than does vision, because we *can* see pretty well even if we are restricted in the exploratory possibilities. Indeed, given these special properties of haptics it would have been surprising if judging the length and orientation of rods had proved to be insensitive to the addition of arm supports. If evidence more general than this had supported the cognitive impenetrability of haptic perception (by cognitive inputs, not weights tied to the arm), then it might have shown that the impenetrable systems transcended sense modalities (as I suggested might be the case in sect. 7.1). In fact, I am not so sure that this is not the case. The facts cited in the Gentaz & Rossetti commentary deal with only one aspect of haptic exploration – that concerned with estimating length orientation. There is considerable evidence that some components of haptic perception (e.g., proprioception and kinaesthesis) interact with vision in interesting ways and that there may be some reason to think that a spatial system may be at work in both cases. For example, one observation that has impressed me is that smooth pursuit eye tracking can occur to both visually perceived motion and to the voluntary motion of one's arm in the dark (Glenny & Haywood 1979), but not to other stimuli, including voluntarily initiated imagined movement (Kowler 1990). But it is too early to stake much on the significance of such evidence.

R8. Conclusions

In my target article I proposed a bold hypothesis (to use Bruner's [1957] own phrase, coined in connection with the opposite hypothesis): that a major portion of vision, called the early vision system, does its job without the intervention of knowledge, beliefs or expectations, even when using that knowledge would prevent it from making errors. Knowledge and expectations of course affect what we see something *as,* but this effect happens either in the attentional selection stage prior to the operation of early vision, or in the perceptual selection or decision stage after the operation of early vision. The target article was long because there are many ways to take issue with this simple claim and because, like any empirical distinction, it has lots of borderline cases. One can focus on these borderline cases and conclude that the thesis is false, or one can see if there might be second-order interaction effects that could account for their apparent borderline status. That is how it is done in all sciences – except in psychology where the strategy of "accounting for variance" often sends people searching for a fuzzy or holistic theory (in which "everything depends on everything else") as the first resort as soon as they see a borderline case.

What I have offered is an empirical hypothesis, not a conceptual analysis. Consequently, it may be wrong either in detail or in its principal thesis. Only time and a lot of hard work in laboratories and theoreticians' offices will settle the matter.

References

Letters "a" and "r" appearing before authors' initials refer to target article and response, respectively.

Abernethy, B. (1991) Visual search strategies and decision-making in sport. Special issue: Information processing and decision-making in sport. *International Journal of Sport Psychology* 22:189–210. [aZP]

Abernethy, B., Neil, R. J. & Koning, P. (1994) Visual-perceptual and cognitive differences between expert, intermediate and novice snooker players. *Applied Cognitive Psychology* 8:185–211. [aZP]

Adelson, E. & Movshon, J. A. (1982) Phenomenal coherence of moving visual patterns. *Nature* 300:523–25. [GV]

Adrián, J. A., Alegria, J. & Morais, J. (1995) Metaphonological abilities of Spanish illiterate adults. *International Journal of Psychology* 30:329–53. [RK]

Aglioti, S., DeSouza, J. F. X. & Goodale, M. A. (1995) Size-contrast illusions deceive the eye but not hand. *Current Biology* 5:679–85. [aZP]

Ahissar, M. & Hochstein, S. (1993) Attentional control of early perceptual learning. *Proceedings of the National Academy of Sciences, USA* 90:5718–22. [PTS]

(1995) How early is early vision? Evidence from perceptual learning. In: *Early vision and beyond,* ed. T. V. Papathomas, C. Chubb, A. Gorea & E. Kowler. MIT Press (A Bradford Book). [aZP]

Aloimonos, Y. (1995) Qualitative vision. *International Journal of Computer Vision* 14:115–17. [YA]

Aloimonos, Y., ed. (1993a) *Active perception. Advances in computer vision.* Erlbaum. [YA]

(1993b) *Visual navigation: From biological systems to unmanned ground vehicles. Advances in computer vision.* Erlbaum. [YA]

Aloimonos, Y., Fermüller, C. & Rosenfeld, A. (1995) Seeing and understanding: Representing the visual world. *ACM Computing Surveys* 27:307–309. [YA]

Anderson, J. R. (1978) Arguments concerning representations for mental imagery. *Psychological Review* 85:249–77. [WRU]

Appelle, S. (1972) Perception and discrimination as a function of stimulus orientation: The "oblique effect" in man and animals. *Psychological Bulletin* 78:266–78. [EG]

Appelle, S. & Countryman, M. (1986) Eliminating the haptic oblique effect: Influence of scanning incongruity and prior knowledge of the standards. *Perception* 15:365–69. [EG]

Arnheim, R. (1986) The two faces of Gestalt psychology. *American Psychologist* 41:820–24. [LEK]

Ashby, F. G., Prinzmetal, W., Ivry, R. & Maddox, W. T. (1996) A formal theory of feature binding in object perception. *Psychological Review* 103:165–92. [aZP]

Attneave, F. (1954) Some informational aspects of visual perception. *Psychological Review* 61:183–93. [LEK]

(1974) Apparent movement and the what-where connection. *Psychologia* 17:108–20. [MRWD]

Avrahami, J. (1998) Presence and absence of top and right advantage as an indication of the attentional strategies underlying object selection. *Perception* 27:431–38. [CL]

Baird, J. C. (1997) *Sensation and judgement.* Erlbaum. [BD]

Ballard, D. H. (1991) Animate vision. *Artificial Intelligence* 48:57–86. [AN]

Barlow, H. B. (1997) The knowledge used in vision and where it comes from. *Philosophical Transactions of the Royal Society of London: B Biological Science* 352:1141–47. [aZP, PTS]

Baylis, G. C. & Driver, J. (1993) Visual attention and objects: Evidence for hierarchical coding of location. *Journal of Experimental Psychology: Human Perception and Performance* 19:451–70. [rZP]

Becker, C. A. (1976) Allocation of attention during visual word recognition. *Journal of Experimental Psychology: Human Perception and Performance* 2:556–66. [TS]

Bennett, B., Hoffman, D. & Prakash, K. (1989) *Observer mechanics: A formal theory of perception.* Academic Press. [aZP]

Bennett, B., Hoffman, D. D., Prakash, C. & Richman, S. (1996) Observer theory, Bayes theory, and psychophysics. In: *Perception as Bayesian inference,* ed. D. C. Knill & W. A. Richards. Cambridge University Press. [MS]

Bentin, S., Allison, T., Puce, A., Perez, E. & McCarthy, G. (1996) Electrophysiological studies of face perception in humans. *Journal of Cognitive Neuroscience* 8:551–65. [JB]

Bermúdez, J. L. (1995) Nonconceptual content: From perceptual experience to subpersonal computational states. *Mind and Language* 10:333–69. [JLB]

(1998a) *The paradox of self-consciousness.* MIT Press. [JLB]

(1998b) Philosophical psychopathology. *Mind and Language* 13:287–307. [JLB]

Besners, D., Coltheart, M. & Davelaar, E. (1984) Basic processes in reading: Computation of abstract letter identities. *Canadian Journal of Psychology* 38:126–34. [JSB]

Bichot, N. P, Schall, J. D. & Thompson, K. G. (1996) Visual feature selectivity in frontal eye fields induced by experience in mature macaques. *Nature* 381:697–99. [JB]

Biederman, I. (1981) On the semantics of a glance at a scene. In: *Perceptual organization,* ed. M. Kubovy & J. R. Pomerantz. Erlbaum. [TS]

(1987) Recognition-by-components: A theory of human image interpretation. *Psychological Review* 94:115–48. [AH, aZP, TS]

(1995) Visual object recognition. In: *Visual cognition,* ed. S. M. Kosslyn & D. N. Osherson. MIT Press. [AH]

Biederman, I., Mezzanotte, R. J. & Rabinowitz, J. C. (1982) Scene perception: Detecting and judging objects undergoing relational violations. *Cognitive Psychology* 14:143–77. [AH, aZP]

Biederman, I. & Shiffrar, M. S. (1987) Sexing day-old chicks: A case study and expert systems analysis of a difficult perceptual-learning task. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 13:640–45. [aZP, PTS]

Binford, T. O. (1982) Survey of model-based image analysis systems. *International Journal of Robotics Research* 1(1):18–64. [AR]

Blackmore, S. J., Brelstaff, G., Nelson, K. & Troscianko, T. (1995) Is the richness of our visual world an illusion? Transsaccadic memory for complex scenes. *Perception* 24:1075–81. [SE]

Blake, R. (1995) Psychoanatomical strategies for studying human visual perception. In: *Early vision and beyond,* ed. T. V. Papathomas, C. Chubb, A. Gorea & E. Kowler. MIT Press (A Bradford Book). [aZP]

Bonnel, A. M., Possami, C. A. & Schmitt, M. (1987) Early modulation of visual input: A study of attentional strategies. *The Quarterly Journal of Experimental Psychology* 39A:757–76. [aZP]

Boring, E. G. (1930) A new ambiguous figure. *American Journal of Psychology* 42:444. [aZP]

(1950) *A history of experimental psychology, 2nd edition.* Appelton-Century-Croft. [BC]

Bowers, J. S. (1996) Different perceptual codes support word and pseudoword priming: Was Morton right all along? *Journal of Experimental Psychology: Learning, Memory, and Cognition* 22:1336–53. [JSB]

Bowers, J. S. & Michita, Y. (1998) An investigation into the structure and acquisition of orthographic knowledge: Evidence from cross-script Kanji-Hiragana priming. *Psychonomic Bulletin and Review* 5:259–64. [JSB]

Bowers, J. S., Vigliocco, G. & Haan, R. (1998) Orthographic, phonological, and articulatory contributions to masked letter and word priming. *Journal of Experimental Psychology: Human Perception and Performance* (in press). [JSB]

Boyce, S. J., Pollatsek, A. & Rayner, K. (1989) Effect of background information on object identification. *Journal of Experimental Psychology: Human Perception and Performance* 15:556–66. [AH]

Brainard, D. H. & Freeman, W. T. (1997) Bayesian color constancy. *Journal of the Optical Society of America* A14:1393–411. [MS]

Braitenberg, V. (1984) *Vehicles: Explorations in synthetic psychology.* MIT Press. [MRWD]

Braunstein, M. L. & Andersen, G. J. (1984) Shape and depth perception from parallel projections of three-dimensional motion. *Journal of Experimental Psychology: Human Perception and Performance* 10:749–60. [JLD]

Braunstein, M. L., Hoffman, D. D. & Pollick, F. E. (1990) Discriminating rigid from nonrigid motion: Minimum points and views. *Perception and Psychophysics* 47:205–14. [JLD]

Bressan, P., Ganis, G. & Vallortigara, G. (1993) The role of depth stratification in the solution of the aperture problem. *Perception* 22:215–28. [GV]

Bridgeman, B. (1992) Conscious vs. unconscious processes: The case of vision. *Theory and Psychology* 2:73–88. [aZP]

(1995) Dissociation between visual processing modes. In: *The handbook of brain theory and neural networks,* ed. M. A. Arbib. MIT Press. [aZP]

Broadbent, D. E. (1967) Word-frequency effect and response bias. *Psychological Review* 74:1–15. [aZP]

Brodsky, T., Fermüller, C. & Aloimonos, Y. (1998) Shape from video: Beyond the epipolar constraint. In: *Proceedings of the DARPA Image Understanding Workshop.* [YA]

Broerse, J. & Crassini, B. (1981) Misinterpretation of imagery-induced McCollough effects: A reply to Finke. *Perception and Psychophysics* 30:96–98. [BC]

(1984) Investigations of perception and imagery using CAEs: The role of experimental design and psychophysical method. *Perception and Psychophysics* 35:155–64. [BC]

Broerse, J., Li, R. & Ashton, R. (1994) Ambiguous pictorial depth cues and perceptions of non-rigid motions in the three-loop figure. *Perception* 23:1049–62. [BC]

Brooks, R. A., Greiner, R. & Binford, T. O. (1979) The ACRONYM model-based vision system. In: *Proceedings of the Sixth International Joint Conference on Artificial Intelligence,* Tokyo, August. [AR]

Brown, C. M. (1984) Computer vision and natural constraints. *Science* 224:1299–1305. [aZP]

Bruce, V. (1991) *Face recognition.* Erlbaum. [aZP]

Bruce, V. & Young, A. (1998) *In the eye of the beholder: The science of face perception.* Oxford University Press. [VB]

Bruner, J. S. (1957) On perceptual readiness. *Psychological Review* 64:123–52. [JLB, AGe, RK, CMM, JRP, aZP, JAS]

Bruner, J. S. & Mintum, A. L. (1955) Perceptual identification and perceptual organization. *Journal of General Psychology* 53:21–28. [aZP]

Bruner, J. S. & Postman, L. (1949) On the perception of incongruity: A paradigm. *Journal of Personality* 18:206–23. [LEK, aZP]

Bruner, J. S., Postman, L. & Rodrigues, J. (1951) Expectations and the perception of color. *American Journal of Psychology* 64:216–27. [JAS]

Bruner, J. S. & Potter, M. C. (1964) Interference in visual recognition. *Science* 144:424–25. [aZP]

Bruyer, R. (1991) Covert face recognition in prosopagnosia: A review. *Brain and Cognition* 15:223–35. [AGe]

Bülthoff, H. H., Edelman, S. & Tarr, M. J. (1995) How are three-dimensional objects represented in the brain? *Cerebral Cortex* 5:247–60. [SE, AH]

Bülthoff, H. H. & Mallot, H. A. (1988) Interaction of depth modules: Stereo and shading. *Journal of the Optical Society of America* 5:1749–58. [SE]

Bülthoff, H. H. & Yuille, A. L. (1991) Bayesian models for seeing shapes and depth. *Comments on Theoretical Biology* 2:283–314. [MS]

Burkell, J. & Pylyshyn, Z. W. (1997) Searching through subsets: A test of the visual indexing hypothesis. *Spatial Vision* 11:225–58. [aZP]

Cacace, A. T. & McFarland, D. J. (1998) Central auditory processing disorder in school-aged children: A critical review. *Journal of Speech, Language and Hearing Research* 41:355–73. [DJM]

Campbell, R., Landis, T. & Regard, M. (1986) Face recognition and lip-reading: A neurological dissociation. *Brain* 109:509–21. [VB]

Carroll, L. (1895) What the tortoise said to Achilles. *Mind* 4:278–80. [AGe]

Castiello, U. & Umiltá, C. (1992) Orienting attention in volleyball players. *International Journal of Sport Psychology* 23:301–10. [aZP]

Cavanagh, P. (1988) Pathways in early vision. In: *Computational processes in human vision: An interdisciplinary perspective,* ed. Z. W. Pylyshyn. Ablex Publishing. [aZP]

Chakravarty, I. (1979) A generalized line and junction labeling scheme with applications to scene analysis. *IEEE Transactions, PAMI,* 202–205. [aZP]

Chapman, D. W. (1932) Relative effects of determinate and indeterminate *Aufgaben. American Journal of Psychology* 44:163–74. [HE]

Chase, W. G. & Simon, H. A. (1973) Perception in chess. *Cognitive Psychology* 5:55–81. [aZP]

Chelazzi, L., Miller, E. K., Duncan, J. & Desimone, R. (1993) A neural basis for visual search in inferior temporal cortex. *Nature* 363:345–47. [JB, rZP, JKT]

Cheng, K. (1986) A purely geometric module in the rat's spatial representation. *Cognition* 23:149–78. [aZP]

Cheong, L., Fermüller, C. & Aloimonos, Y. (1998) Effects of errors in the viewing geometry on shape estimation. *Computer Vision and Image Understanding* 71:356–72. [YA]

Chignell, M. H. & Krueger, L. E. (1984) Further evidence for priming in perceptual matching: Temporal, not spatial, separation enhances the fast-*same* effect. *Perception and Psychophysics* 36:257–65. [LEK]

Chomsky, N. (1980) Rules and representations. *Behavioral and Brain Sciences* 3(1):1–15. [rZP]

Churchland, P. M. (1988) Perceptual plasticity and theoretical neutrality: A reply to Jerry Fodor. *Philosophy of Science* 55:167–87. [aZP]

Clark, A. (1997) *Being there: Putting brain, body, and the world together again.* MIT Press/ A Bradford Book. [AN]

Clowes, M. B. (1971) On seeing things. *Artificial Intelligence* 2:79–116. [aZP]

Cohen, D. J. (1997) Visual detection and perceptual independence: Assessing color and form. *Perception and Psychophysics* 59:623–35. [DJC]

Cohen, J. D., Dunbar, K. & McClelland, J. L. (1990) On the control of automatic processes: A parallel distributed processing account of the Stroop effect. *Psychological Review* 97:332–61. [CL]

Coltheart, M. (1981) Disorders of reading and their implications for models of normal reading. *Visible Language* 3:245–86. [JSB]

Connine, C. M. & Clifton, C. (1987) Interactive use of lexical information in speech perception. *Journal of Experimental Psychology: Human Perception and Performance* 13:291–99. [aZP]

Corbin, H. (1942) The perception of grouping and apparent movement in visual depth. *Archives of Psychology, No. 273.* [MRWD]

Cowey, A. & Stoerig, P. (1991) The neurobiology of blindsight. *Trends in Neuroscience* 14:140–45. [AGe]

Crick, G. & Koch, C. (1995) Are we aware of neural activity in primary visual cortex? *Nature* 375:121–23. [S-LY]

Dai, H., Scharf, B. & Buss, S. (1991) Effective attenuation of signals in noise under focused attention. *Journal of the Acoustical Society of America* 89:2837–42. [aZP]

Danziger, E. & Pederson, E. (1999) Through the looking-glass: Literacy, writing systems and mirror-image discrimination. *Written Language and Literacy* 1(2):153–64. [RK]

Davies, I. R. L., Sowden, P. T., Hammond, S. & Ansell, J. (1994) Expertise at mammography: A perceptual or conceptual skill? In: *Medical imaging 1994: Image perception, Proceedings of SPIE* 2166:86–94, ed. H. L. Kundel. SPIE-The International Society for Optical Engineering. [PTS]

Davies, I. R. L., Sowden, P. T., Jerrett, D. T., Jerrett, T. & Corbett, G. G. (1998) A cross-cultural study of English and Setswana speakers on a colour triads task: A test of the Sapir-Whorf hypothesis. *British Journal of Psychology* 89:1–15. [PTS]

Dawson, M. R. W. (1989) Apparent motion and element connectedness. *Spatial Vision* 4:241–51. [MRWD]

(1991) The how and why of what went where in apparent motion: Modeling solutions to the motion correspondence process. *Psychological Review* 98:569–603. [MRWD]

(1998) *Understanding cognitive science.* Blackwell. [MRWD]

Dawson, M. R. W., Nevin-Meadows, N. & Wright, R. D. (1994) Polarity matching in the Ternus configuration. *Vision Research* 34:3347–59. [MRWD]

Dawson, M. & Pylyshyn, Z. W. (1988) Natural constraints in apparent motion. In: *Computational processes in human vision: An interdisciplinary perspective,* ed. Z. W. Pylyshyn. Ablex Publishing. [aZP]

Dawson, M. R. W. & Wright, R. D. (1989) The consistency of element transformations affects the visibility but not the direction of illusory motion. *Spatial Vision* 4:17–29. [MRWD]

(1994) Simultaneity in the Ternus configuration: Psychophysical data and a computer model. *Vision Research* 34:397–407. [MRWD]

Delk, J. L. & Fillenbaum, S. (1965) Differences in perceived color as a function of characteristic color. *American Journal of Psychology* 78:290–93. [rZP, JAS]

Dennett, D. C. (1991) *Consciousness explained.* Little, Brown. [SE, rZP]

Deregowski, J. B. (1989) Real space and represented space: Cross-cultural perspectives. *Behavioral and Brain Sciences* 12:51–119. [RK]

Desimone, R. (1996) Neural mechanisms for visual memory and their role in attention. *Proceedings of the National Academy of Science, USA* 93:13494–97. [aZP]

Desimone, R. & Duncan, J. (1995) Neural mechanisms of selective attention. *Annual Review of Neuroscience* 18:193–222. [JKT]

De Valois, R. L. & De Valois, K. K. (1990) *Spatial vision.* Oxford University Press. [PGS]

De Weerd, P., Desimone, R. & Ungerleider, L. G. (1996) Cue-dependent deficits in grating orientation discrimination after V4 lesions in macaques. *Visual Neuroscience* 13:529–38. [AGr]

Dobbins, A. C., Jeo, R. M., Fiser, J. & Allman, J. M. (1998) Distance modulation of activity in the visual cortex. *Science* 281:552–55. [WRU]

Downing, C. J. (1988) Expectancy and visual-spatial attention: Effects on perceptual quality. *Journal of Experimental Psychology: Human Perception and Performance* 14:188–202. [aZP]

Dretske, F. (1969) *Seeing and knowing.* University of Chicago Press. [MRWD]

(1981) *Knowledge and the flow of information.* MIT Press. [MRWD]

Driver, J., Davis, G., Kidd, P., Maxwell, E., Ricciardelli, P. & Baron-Cohen, S. (in press) Shared attention and the social brain: Gaze perception triggers automatic visuospatial orienting in adults. *Visual Cognition.* [VB]

Duhamel, J.-R., Colby, C. L. & Goldberg, M. E. (1992) The updating of the representation of visual space in parietal cortex by intended eye movements *Science* 255(5040):90–92. [rZP]

Duncker, K. (1939) The influence of past experience upon perceptual properties. *American Journal of Psychology* 52:255–65. [JAS]

Durlach, M. I. & Braida, L. D. (1969) Intensity perception. I. Preliminary theory of intensity resolution. *Journal of the Acoustical Society of America* 46:372–83. [NAM]

Edelman, S. (1998) Representation is representation of similarities. *Behavioral and Brain Sciences* 21(4):449–98. [TS]

Edelman, S. & Duvdevani-Bar, S. (1997) A model of visual recognition and categorization. *Philosophical Transactions of the Royal Society of London, B* 352(1358):1191–1202. [SE]

Edelman, S. & Poggio, T. (1992) Bringing the grandmother back into the picture: A memory-based view of pattern recognition. *International Journal of Pattern Recognition and Artificial Intelligence* 6:37–61. [AR]

Egeth, H. (1967) Selective attention. *Psychological Bulletin* 67:41–57. [HE]

Egeth, H. & Bevan, W. (1973) Attention. In: *Handbook of General Psychology,* ed. B. B. Wolman. Prentice-Hall. [HE]

Egeth, H. & Smith, E. E. (1967) Perceptual selectivity in a visual recognition task. *Journal of Experimental Psychology* 74:543–49. [HE]

Egeth, H. E., Virzi, R. A. & Garbart, H. (1984) Searching for conjunctively defined targets. *Journal of Experimental Psychology* 10:32–39. [aZP]

Ellman, J. L. & McClelland, J. L. (1988) Cognitive penetration of the mechanisms of perception: Compensation for coarticulation of lexically restored phonemes. *Journal of Memory and Language* 27:143–65. [aZP]

Epstein, W. (1977) *Stability and constancy in visual perception: Mechanisms and processes.* Wiley. [MRWD]

(1982) Percept-percept couplings. *Perception* 11:75–83. [aZP]

Essock, E. A. (1980) The oblique effect of stimulus identification considered with respect to two classes of oblique effects. *Perception* 9:37–46. [EG]

Evans, G. (1982) *The varieties of reference.* Clarendon Press. [JLB]

Fahle, M., Edelman, S. & Poggio, T. (1995) Fast perceptual learning in hyperactivity. *Vision Research* 35:3003–13. [aZP]

Farah, M. J. (1989) Semantic and perceptual priming: How similar are the underlying mechanisms? *Journal of Experimental Psychology: Human Perception and Performance* 15:188–94. [aZP, GR]

(1990) *Visual agnosia: Disorders of object recognition and what they tell us about normal vision.* MIT Press. [aZP]

Felleman, D. J. & Van Essen, D. C. (1991) Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex* 1:1–47. [aZP, JKT]

Felleman, D. J., Xiao, Y. & McClendon, E. (1997) Modular organization of occipital-temporal pathways: Cortical connections between visual area 4 and visual area 2 and posterior inferotemporal ventral area in macaque monkeys. *Journal of Neuroscience* 17:3185–3200. [aZP]

Fermüller, C. & Aloimonos, Y. (1995a) Direct perception of three-dimensional motion from patterns of visual motion. *Science* 270:1973–76. [YA]

(1995b) Qualitative egomotion. *International Journal of Computer Vision* 15:7–29. [YA]

(1995c) Representations for active vision. In: *Proceedings of the International Joint Conference on Artificial Intelligence,* August 1995. [YA]

(1996) Ordinal representations of visual space. In: *Proceedings of the ARPA Image Understanding Workshop,* February 1996. [YA]

(1997) On the geometry of visual correspondences. *International Journal of Computer Vision* 21:223–47. [YA]

(1998) Ambiguity in structure from motion: Sphere versus plane. *International Journal of Computer Vision* 28:137–54. [YA]

Fermüller, C., Cheong, L. & Aloimonos, Y. (1997a) Visual space distortion. *Biological Cybernetics* 77:323–37. [YA]

(1998a) 3-D motion and shape representations in visual servo control. *International Journal of Robotics Research* 17:4–18. [YA]

Fermüller, C., Pless, R. & Aloimonos, Y. (1997b) Families of stationary patterns producing illusory movement: Insights into the visual system. *Proceedings of the Royal Society of London B* 264:795–806. [YA]

(1998b) The Ouchi illusion as an artifact of biased flow estimation. Technical Report CAR-TR-890, Center for Automation Research, University of Maryland. [YA]

Ferrera, V. P., Rudolph, K. K. & Maunsell, J. H. R. (1994) Responses of neurons in the parietal and temporal visual pathways. *Journal of Neuroscience* 14:6171–85. [AGr]

Feyerabend, P. (1962) Explanation, reduction and empiricism. In: *Minnesota studies in philosophy of science, vol. 3,* ed. H. Feigl & G. Maxwell. University of Minnesota Press. [aZP]

Fodor, J. A. (1983) *The modularity of mind: An essay on faculty psychology.* MIT Press/A Bradford Book. [JSB, MRWD, SE, TVP, MAP, arZP, GR]

(1985) Précis of *The modularity of mind. Behavioral and Brain Sciences* 8:1–42. [LEK]

(1998) *Concepts: Where cognitive science went wrong.* Oxford University Press. [arZP]

Fodor, J. A. & Pylyshyn, Z. W. (1981) How direct is visual perception? Some reflections on Gibson's "ecological approach." *Cognition* 9:139–96. [VB, arZP]

Foley, J. M. (1967) Effects of voluntary eye movement and convergence on the binocular appreciation of depth. *Perception and Psychophysics* 11:423–27. [YA]

(1980) Binocular distance perception. *Psychological Review* 87:411–34. [YA]

Freeman, W. T. (1994) The generic viewpoint assumption in a framework for visual perception. *Nature* 368:542–45. [MS]

(1996) The generic viewpoint assumption in a Bayesian framework. In: *Perception as Bayesian inference,* ed. D. C. Knill & W. A. Richards. Cambridge University Press. [MS]

Frégnac, Y., Shulz, D., Thorpe, S. & Bienenstock, E. (1988) A cellular analogue of visual cortical plasticity. *Nature* 333:367–70. [PTS]

Freuder, E. C. (1986) Knowledge-mediated perception. In: *Pattern recognition by humans and machines: Visual perception,* ed. H. C. Nusbaum & E. C. Schwab. Academic Press Inc. [aZP]

Friedman, A. (1979) Framing pictures: The role of knowledge in automatized encoding and memory for gist. *Journal of Experimental Psychology: General* 108:316–55. [AH]

Friedman-Hill, S. Wolfe, J. M. (1995) Second-order parallel processing: Visual search for odd items in a subset. *Journal of Experimental Psychology: Human Perception and Performance* 21:531–51. [aZP]

Frisby, J. P. & Clatworthy, J. L. (1975) Learning to see complex random-dot stereograms. *Perception* 4:173–78. [aZP]

Gale, A. G. & Findlay, J. M. (1983) Eye movement patterns in viewing ambiguous figures. In: *Eye movements and psychological functions: International views,* ed. R. Groner, C. Menz, D. F. Fisher & R. A. Monty. Erlbaum. [aZP]

Gellatly, A. (1988) Human inference. In: *Human and machine problem solving,* ed. K. K. Gilhooly. Plenum. [AGe]

(1992) The misleading concept of cognitive competences. *Theory and Psychology* 2:363–90. [AGe]

(1995) Colourful Whorfian ideas: Linguistic and cultural influences on the perception and cognition of colour and on the investigation of them. *Mind and Language* 10:199–225. [AGe]

Gallistel, C. R. (1990) *The organization of learning.* MIT Press (A Bradford Book). [arZP]

Gentaz, E. (in press) Existe-il un "effet de l'oblique" dans la perception tactile des orientations? [Is there an oblique effect in the tactual perception of orientation?] *L'Année Psychologique.* [EG]

Gentaz, E. & Hatwell, Y. (1995) The haptic "oblique effect" in children's and adults' perception of orientation. *Perception* 24:631–46. [EG]

(1996) Role of gravitational cues in the haptic perception of orientation. *Perception and Psychophysics* 58:1278–92. [EG]

(1998) The haptic oblique effect in the perception of rod orientation by blind adults. *Perception and Psychophysics* 60:157–67. [EG]

Gibson, B. S. & Peterson, M. A. (1994) Does orientation-independent object recognition precede orientation-dependent recognition? Evidence from a cueing paradigm. *Journal of Experimental Psychology: Human Perception and Performance* 20:299–316. [MAP]

Gibson, E. J. (1991) *An odyssey in learning and perception.* MIT Press. [aZP]

Gibson, J. J. (1966) *The senses considered as perceptual systems.* Houghton Mifflin. [RW]

(1979) *The ecological approach to visual perception.* Houghton Mifflin. [JLB, BC, JRP, aZP]

Gibson, J. R. & Maunsell, J. H. R. (1997) Sensory modality specificity of neural activity related to memory in visual cortex. *Journal of Neurophysiology* 78:1263–75. [AGr]

Gilbert, C. D. & Wiesel, T. N. (1992) Receptive field dynamics in adult primary visual cortex. *Nature* 356:150–52. [PTS]

Gilchrist, A. (1977) Perceived lightness depends on perceived spatial arrangement. *Science* 195:185–87. [aZP]

Girgus, J. J., Rock, I. & Egatz, R. (1977) The effect of knowledge of reversibility on the reversibility of ambiguous figures. *Perception and Psychophysics* 22:550–56. [aZP]

Glenny, G. & Heywood, S. (1979) Hanz Gertz revisited: The different effects of invisibility and darkness on pursuit eye movements. *Perception* 8(1):31–36. [rZP]

Gnadt, J. W. & Andersen, R. A. (1988) Memory related motor planning activity in posterior parietal cortex of macaque. *Experimental Brain Research* 70:216–20. [AGr]

Gnadt, J. W., Bracewell, R. M. & Anderson, R. A. (1991) Sensorimotor transformation during eye movements to remembered visual targets. *Vision Research* 31:693–715. [aZP]

Gogel, W. C. (1973) The organization of perceived space. *Psychologische Forschung* 37:195–221. [aZP]

(1976) An indirect measure of perceived distance from oculomotor cues. *Perception and Psychophysics* 20:419–29. [MAP]

Gogel, W. C. & Tietz, J. D. (1992) Absence of compensation and reasoning-like processes in the perception of orientation in depth. *Perception and Psychophysics* 51:309–18. [MRWD]

Goldstone, R. L. (1994) Influences of categorization on perceptual discrimination. *Journal of Experimental Psychology: General* 123:178–200. [aZP]

(1995) The effect of categorization on color perception. *Psychological Science* 6:298–304. [aZP]

Goodale, M. A. (1983) Vision as a sensorimotor system. In: *Behavioral approaches to brain research,* ed. T. E. Robinson. Oxford University Press. [aZP]

(1988) Modularity in visuomotor control: From input to output. In: *Computational processes in human vision: An interdisciplinary perspective,* ed. Z. W. Pylyshyn. Ablex Publishing. [aZP]

(1998) Where does vision end and action begin? *Current Biology* 8:R489-R491. [AGr]

Goodale, M. A., Jacobson, J. S. & Keillor, J. M. (1994) Differences in the visual control of pantomimed and natural grasping movements. *Neuropsychologia* 32:1159–78. [aZP]

Goodale, M. A., Pelisson, D. & Prablanc, C. (1986) Large adjustments in visually guided reaching do not depend on vision of the hand or perception of target displacement *Nature* 320:748–50. [aZP]

Goren, C. C., Sarty, M. & Wu, R. W. K. (1975) Visual following and patter discrimination of face-like stimuli by new-born infants. *Pediatrics* 56:544–49. [VB]

Gottlieb, J. P., Kusunoki, M. & Goldberg, M. E. (1998) The representation of visual salience in monkey parietal cortex. *Nature* 391:481–84. [AGr]

Gottschaldt, K. (1926) Uber den Einfluß der Erfahrung auf die Wahmehmung von Figuren: 1. *Psychologische Forschung* 8:261–318. [RK]

Green, B. F. & Anderson, L. K. (1956) Color coding in a visual search task. *Journal of Experimental Psychology* 51:19–24. [aZP]

Greenfield, P. M. & Bruner, J. S. (1966) Culture and cognitive growth. *International Journal of Psychology* 1:89–107. [RK]

Greenfield, P. M., deWinstanley, P., Kilpatrick, H. & Kaye, D. (1994) Action video games and informal education: Effects on strategies for dividing visual attention. Special issue: Effects of interactive entertainment technologies on development. *Journal of Applied Developmental Psychology* 15:105–23. [aZP]

Gregory, R. L. (1968) Perceptual illusions and brain models. *Proceedings of the Royal Society of London B* 171:279–96. [TVP]

(1970) *The intelligent eye.* McGraw-Hill. [CMM]

(1973) The confounded eye. In: *Illusion in nature and art,* ed. R. L. Gregory & E. H. Gombrich. Duckworth. [VB]

(1980) Perceptions as hypotheses. *Philosophical Transactions of the Royal Society of London B* 290:181–97. [TVP]

(1997) Knowledge in perception and illusion. *Philosophical Transactions of the Royal Society of London B* 352:1121–28. [TVP]

Grimson, W. E. L. (1990) The combinatorics of object recognition in cluttered environment using constrained search. *Artificial Intelligence* 44:121–66. [aZP]

Grosof, D. H., Shapley, R. M. & Hawken, M. J. (1993) Macaque V1 neurons can signal "illusory" contours. *Nature* 365:550–52. [TVP]

Haber, R. N. (1966) Nature of the effect of set on perception. *Psychological Review* 73:335–51. [HE, aZP]

Haenny, P. E., Maunsell, J. H. R. & Schiller, P. H. (1988) State-dependent activity in monkey visual cortex. II. Retinal and extraretinal factors in V. *Experimental Brain Research* 69:245–59. [JB, rZP, JKT]

Haenny, P. & Schiller, P. (1988) State-dependent activity in monkey visual cortex. I. Single cell activity in V1 and V4 on visual tasks. *Experimental Brain Research* 69:225–44. [JB, aZP]

Hamlyn, D. W. (1957) *The psychology of perception.* Routledge & Kegan Paul. [MRWD]

Hammett, S. T. & Snowden, R. J. (1995) The effect of contrast adaptation on briefly presented stimuli. *Vision Research* 35:1721–25. [PGS]

Hanson, N. R. (1958) *Patterns of discovery.* Cambridge University Press. [aZP]

Hatwell, Y. (1978) Form perception and related issues in blind humans. In: *Handbook of sensory physiology, VII: Perception,* ed. R. Held, H. W. Leibowitz & H. L. Teuber. Springer Verlag. [EG]

(1986) *Toucher l'espace: La main et la perception tactile de l'espace* [*Touching space: The hand and the haptic perception of space*]. Presses Universitaires de Lille. [EG]

(1990) Spatial perception by eye and hand: Comparison and intermodal integration. In: *Development of eye-hand coordination across life-span,* ed. C. Bard, M. Fleury & L. Hay. South Carolina University Press. [EG]

(1994) Transferts intermodaux et intégration intermodale [Intermodal transfer and intermodal integration]. In: *Traité de psychologie expérimentale* [*Handbook of experimental psychology*], ed. M. Richelle, J. Requin & M. Robert. Presses Universitaires de France. [EG]

He, S., Cavanagh, P. & Intriligator, J. (1996) Attentional resolution and the locus of visual awareness. *Nature* 383:334–37. [S-LY]

Heeley, D. & Buchanan-Smith, H. (1992) Orientation acuity estimated with simultaneous and successive procedures. *Spatial Vision* 6:1–10. [EG]

Heeley, D., Buchanan-Smith, H., Cromwell, J. & Wright, J. (1997) The oblique effect in orientation acuity. *Vision Research* 32:235–42. [EG]

Heil, J. (1983) *Perception and cognition.* University of California Press. [MRWD]

Hering, E. (1964) *Outlines of a theory of light sense,* trans. L. M. Hurvich & D. Jameson. Harvard University Press. [JAS]

Hermer, L. & Spelke, E. (1996) Modularity and development: The case of spatial reorientation. *Cognition* 61(3):195–232. [rZP]

Hernandez-Péon, R., Scherrer, R. H. & Jouvet, M. (1956) Modification of electrical activity in the cochlear nucleus during "attention" in unanesthetized cats. *Science* 123:331–32. [aZP]

Hill, H. & Bruce, V. (1993) Independent effects of lighting, orientation and stereopsis on the hollow-face illusion. *Perception* 22:887–97. [VB]

(1994) A comparison between the hollow face and hollow "potato" illusions. *Perception* 23:1335–37. [VB]

Hinde, R. A. (1974) *Biological bases of human social behaviour.* McGraw-Hill. [BC]

Hochberg, J. (1956) Perception: Toward a recovery of a definition. *Psychological Review* 63:400–405. [MAP]

(1974) Higher order stimuli and inter-response coupling in the perception of the visual world. In: *Perception: Essays in honor of James J. Gibson,* ed. R. B. Macleod & H. L. Pick. Cornell University Press. [MAP]

(1978) *Perception.* (2nd edition). Prentice Hall. [CMM]

Hochberg, J. & Peterson, M. A. (1987) Piecemeal organization and cognitive components in object perception: Perceptually coupled responses to moving objects. *Journal of Experimental Psychology: General* 116:370–80. [MAP]

Hoffman, D. D. (1998) *Visual intelligence: How we create what we see.* Norton. [MS]

Hoffman, D. D. & Singh, M. (1997) Salience of visual parts. *Cognition* 63:29–78. [MS]

Hollingworth, A. & Henderson, J. M. (1998) Does consistent scene context facilitate object perception? *Journal of Experimental Psychology: General* 127:398–415.. [AH, aZP, TS]

(in press) Object identification is isolated from scene semantic constraint: Evidence from object type and token discrimination. *Acta Psychologica: Special Issue on Object Perception and Memory.* [AH]

Holmes, G. (1918) Disturbances in visual orientation. *British Journal of Ophthalmology* 2:449–506. [JLB, aZP]

Howard, I. P. (1982) *Human visual orientation.* Wiley. [EG, aZP]

Hubel, D. H. & Wiesel, T. N. (1962) Receptive fields, binocular interaction, and functional architecture in the cat's visual cortex. *Journal of Physiology* 160:106–54. [aZP]

Humphreys, G. W. & Riddoch, M. J. (1987) *To see but not to see: A case study of visual agnosia.* Erlbaum. [aZP]

Hupé, J. M., James, A. C., Payne, B. R., Lomber, S. G., Girard, P. & Bullier, J. (1998) Cortical feedback improves discrimination between figure and background by V1, V2 and V3 neurons. *Nature* 394:784–87. [JB]

Irwin, D. E. (1993) Perceiving an integrated visual world. In: *Attention and performance XIV,* ed. D. E. Meyer & S. Kornblum. MIT Press. [aZP]

Ittelson, W. H. (1952) *The Ames demonstrations in perception.* Princeton University Press. [RK]

Ittelson, W. H. & Ames, A. J. (1968) *The Ames demonstrations in perception.* Hafner. [aZP]

Jacoby, L. L. & Brooks, L. R. (1984) Nonanalytic cognition: Memory, perception, and concept learning. In: *The psychology of learning and motivation, vol. 18,* ed. G. H. Bower. Academic Press. [TS]

Jeffreys, D. A. (1996) Evoked potential studies of face and object processing. *Visual Cognition* 3:1–38. [JB]

Johanssen, G. (1950) *Configurations in event perception.* Houghton Mifflin. [aZP]

Johnson, M. H., Dziurawiec, S., Ellis, H. & Morton, J. (1991) Newborns' preferential tracking of face-like stimuli and its subsequent decline. *Cognition* 40:1–19. [VB]

Johnson-Laird, P. N. (1983) *Mental models.* Cambridge University Press. [AGe]

Johnston, A., Hill, H. & Carman, N. (1992) Recognising faces: Effects of lighting direction, inversion, and brightness reversal. *Perception* 21:365–75. [VB]

Johnston, E. B. (1991) Systematic distortions of shape from stereopsis. *Vision Research* 31:1351–60. [YA]

Johnston, W. A. & Hawley, K. J. (1994) Perceptual inhibition of expected inputs: The key that opens closed minds. *Psychonomic Bulletin and Review* 1:56–72. [TVP]

Jolicoeur, P. (1988) Mental rotation and the identification of disoriented objects. *Canadian Journal of Psychology* 42:461–78. [MAP]

Jones, E. & Bruner, J. (1954) Expectancy in apparent visual movement. *British Journal of Psychology* 45:157–65. [MRWD]

Julesz, B. (1971) *Foundations of Cyclopean perception.* University of Chicago Press. [aZP]

(1984) Toward an axiomatic theory of preattentive vision. In: *Dynamic aspects of neocortical function,* ed. G. Edeman, M. Cowen & E. Gall. Wiley. [CMM]

(1990) Early vision is bottom up, except for focal attention. *Cold Spring Harbor Symposium on Quantitative Biology* 55:973–78. [aZP]

Julesz, B. & Papathomas, T. V. (1984) On spatial-frequency channels and attention. *Perception and Psychophysics* 36:398–99. [aZP]

Kahneman, D. & Treisman, A. (1992) The reviewing of object files: Object-specific integration of information. *Cognitive Psychology* 24(2):175–219. [arZP]

Kanisza, B. (1985) Seeing and thinking. *Acta Psychologica* 59:23–33. [CMM]

Kanizsa, G. (1969) Perception, past experience and the impossible experiment. *Acta Psychologica* 31:66–96. [aZP]

(1976) Subjective contours. *Scientific American* 234:48–52. [TVP]

(1985) Seeing and thinking. *Acta Psychologica* 59:23–33. [aZP]

Kanizsa, G. & Gerbino, W. (1982) Amodal completion: Seeing or thinking? In: *Organization and representation in perception,* ed. B. Beck. Erlbaum. [aZP]

Kanwisher, N., McDermott, J. & Chunn, M. M. (1997) The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience* 17:4302–11. [aZP]

Karni, A. & Bertini, G. (1997) Learning perceptual skills: Behavioral probes into adult cortical plasticity. *Current Opinions in Neurobiology* 7:530–35. [PTS]

Karni, A. & Sagi, D. (1995) A memory system in the adult visual cortex. In: *Maturational Windows and adult cortical plasticity,* ed. B. Julesz & I. Kovacs. Addison-Wesley. [aZP]

Katz, S. (1983) R. L. Gregory and others: The wrong picture of the picture theory of perception. *Perception* 12:269–79. [BC]

Kawabata, N. (1986) Attention and depth perception. *Perception* 15:563–72. [aZP]

Kelly, M. D. (1971) Edge detection by computer using planning. In: *Machine intelligence 6,* ed. B. Meltzer & D. Michie. University of Edinburgh Press. [aZP]

Knill, D. C. & Richards, W. A., eds. (1996) *Perception as Bayesian inference.* Cambridge University Press. [MS]

Koenderink, J. J., Kappers, A. & van Doorn, A. (1992) Local operations: The embodiment of geometry. In: *Artificial and biological vision systems,* ed. G. A. Orban & H. H. Nagel. Springer-Verlag. [YA]

Koffka, K. (1935) *Principles of Gestalt psychology.* Harcourt. [BC]

Kolers, P. (1972) *Aspects of motion perception.* Pergamon Press. [MRWD]

Kolinsky, R. (1998) Spoken word recognition: A stage-processing approach to language differences. *The European Journal of Cognitive Psychology* 10:1–40. [RK]

Kolinsky, R., Morais, J. & Brito-Mendes, C. (1990) Embeddedness effects on part verification in children and unschooled adults. *Psychologica Belgica* 30:49–64. [RK]

Kolinsky, R., Morais, J., Content, A. & Cary, L. (1987) Finding parts within figures: A developmental study. *Perception* 16:399–407. [RK]

Kolinsky, R., Morais, J. & Verhaeghe, A. (1994) Visual separability: A study on unschooled adults. *Perception* 23:471–86. [RK]

Kosslyn, S. M. (1994) *Image and brain.* MIT Press. [aZP]

(1995) Mental imagery. In: *An invitation to cognitive science: Vol. 2. Visual cognition, 2nd edition,* ed. S. M. Kosslyn & D. N. Osherson. MIT Press. [GR]

Kosslyn, S. M. & Sussman, A. L. (1995) Roles of imagery in perception: Or, there is no such thing as immaculate perception. In: *The cognitive neurosciences,* ed. M. S. Gazzaniga. MIT Press. [EG]

Kosslyn, S. M., Thompson, W. L., Kim, I. J. & Alpert, N. M. (1995) Topographic representations of mental images in primary visual cortex. *Nature* 378:496–98. [BC]

Kowler, E. (1989) Cognitive expectations, not habits, control anticipatory smooth oculomotor pursuit. *Vision Research* 29:1049–57. [rZP]

(1990) The role of visual and cognitive processes in the control of eye movement. In: *Eye movements and their role in visual and cognitive processes,* ed. E. Kowler. Elsevier. [rZP]

Krueger, L. E. (1984) The category effect in visual search depends on physical rather than conceptual differences. *Perception and Psychophysics* 35:558–64. [LEK]

(1989) Cognitive impenetrability of perception (Commentary on D. W. Massaro, "Speech perception by ear and eye: A paradigm for psychological inquiry"). *Behavioral and Brain Sciences* 12:769–70. [LEK]

Kreuger, L. E. & Tsay, C.-Y. (1990) Analyzing vision at the complexity level: Misplaced complexity? (Commentary on J. K. Tsotsos, "Analyzing vision at the compexity level"). *Behavioral and Brain Sciences* 13:449–50. [LEK]

Kubovy, M., Cohen, D. J. & Hollier, J. (in press) Feature integration that routinely occurs without focal attention. *Psychonomic Bulletin and Review.* [DJC]

Kuhn, T. (1972) *The structure of scientific revolutions.* University of Chicago Press. [aZP]

Külpe, O. (1904) Versuche über Abstraktion. *Bericht über den Ie Kongresz für Experimentale Psychologie.* [HE]

Kutas, M., McCarthy, G. & Donchin, E. (1977) Augmenting mental chronometry: The P300 as a measure of stimulus evaluation time. *Science* 197:792–95. [aZP]

Lamme, V. A. F. (1995) The neurophysiology of figure ground segregation in primary visual cortex. *Journal of Neuroscience* 15:1605–15. [JB]

Lamme, V. A. F., Zipser, K. & Spekreijse, H. (1998) Figure-ground activity in primary visual cortex is suppressed by anesthesia. *Proceedings of the National Academy of Sciences USA* 95:3263–68. [JB]

Langton, S. R. H. & Bruce, V. (in press) Reflexive visual orienting in response to the social attention of others. *Visual Cognition.* [VB]

Latimer, C. R. (1988) Eye movement data: Cumulative fixation time and cluster analysis. *Behavior Research Methods, Instruments and Computers* 20:437–70. [CL]

Latimer, C. R. & Stevens, C. J. (1993) Eye movement measures and connectionist models of form perception. In: *Cognitive approaches to human perception,* ed. S. Ballesteros. Erlbaum. [CL]

Lawrence, D. H. & Coles, G. R. (1954) Accuracy of recognition with alternatives before and after the stimulus. *Journal of Experimental Psychology* 47:208–14. [HE]

Lee, D. N. & Lishman, J. R. (1975) Visual proprioceptive control of stance. *Journal of Human Movement Studies* 1:87–95. [aZP]

Leopold, D. A. & Logothetis, N. K. (1996) Activity changes in early visual cortex reflect monkeys' percepts during binocular rivalry. *Nature* 379:549–53. [JB]

Leslie, A. M. (1988) The necessity of illusion: Perception and thought in infancy. In: *Thought without language,* ed. L. Weiskrantz. Oxford Science Publications. [aZP]

(1994) Pretending and believing: Issues in the theory of ToMM. *Cognition* 50(1–2):211–38. [rZP]

Liberman, I. Y., Shankweiler, D. P., Fischer, F. W. & Carter, B. (1974) Explicit syllable and phoneme segmentation in the young child. *Journal of Experimental Child Psychology* 18:201–12. [RK]

Lindsay, P. H. & Norman, D. A. (1977) *Human information processing: An introduction to psychology.* Academic Press. [aZP]

Lindsay, P. H., Taylor, M. M. & Forbes, S. M. (1968) Attention and multidimensional discrimination. *Perception and Psychophysics* 4:113–17. [NAM]

Livingstone, M. S. & Hubel, D. H. (1987) Psychophysical evidence for separate channels for the perception of form, color, movement, and depth. *The Journal of Neuroscience* 7:3416–68. [aZP, GV]

Longstreth, L. E., El-Zahhar, N. & Alcorn, M. B. (1985) Exceptions to Hick's Law: Explorations with a response duration measure. *Journal of Experimental Psychology:* General 114:417–34. [aZP]

Lowe, D. (1987) Three-dimensional object recognition from single two-dimensional images. *Artificial Intelligence* 31:355–59. [aZP]

Lumer, E. D., Friston, K. J. & Rees, G. (1998) Neural correlates of perceptual rivalry in the human brain. *Science* 280:1930–34. [JB]

Lupker, S. J. & Massaro, D. W. (1979) Selective perception without confounding contributions of decision and memory. *Perception and Psychophysics* 25:60–69. [aZP]

Lynch, J. C. (1980) The functional organization of posterior parietal cortex. *Behavioral and Brain Sciences* 3:485–534. [aZP]

Mace, W. M. (1977) James Gibson's strategy for perceiving: Ask not what's in your head, but what your head is inside of. In: *Perceiving, acting and knowing,* ed. R. Shaw & J. Bransford. Erlbaum. [MRWD]

Macmillan, N. A. & Creelman, C. D. (1991) *Detection theory: A user's guide.* Cambridge University Press. [LEK]

Macmillan, N. A., Goldberg, R. F. & Braida, L. D. (1988) Resolution for speech sounds: Basic sensitivity and context memory on vowel and consonant continua. *Journal of the Acoustical Society of America* 84:1262–80. [NAM]

Magnuson, S. (1970) Reversibility of perspective in normal and stabilized viewing. *Scandinavian Journal of Psychology* 11:153–56. [aZP]

Maloney, L. T. & Wandell, B. A. (1990) Color constancy: A method for recovering surface spectral reflectance. In: *Image understanding 1989,* ed. S. Ullman & W. Richards. Ablex Publishing. [aZP]

Marr, D. (1982) *Vision: A computational investigation into the human representation and processing of visual information.* W. H. Freeman. [YA, MRWD, SE, LEK, AN, JRP, aZP, AR, PTS]

Marr, D. & Nishihara, H. K. (1978) Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society of London, B* 200:269–94. [AH]

Marr, D. & Poggio, T. (1979) A theory of human stereo vision. *Proceedings of the Royal Society of London* 204:301–28. [aZP]

Massaro, D. W. (1988) *Experimental psychology: An information processing approach.* Harcourt Brace Jovanovich. [aZP]

Masson, M. E. J. & Borowsky, R. (1998) More than meets the eye: Context effects in word identification. *Memory and Cognition* (in press). [GR]

Maunsell, J. H. R., Sclar, G., Nealey, T. A. & Depriest, D. D. (1991) Extraretinal representations in area V4 in the macaque monkey. *Visual Neuroscience* 7:561–73. [AGr]

McBeath, M. K., Morikawa, K. & Kaiser, M. K. (1992) Perceptual bias for forward-facing motion. *Psychological Science* 3:362–67. [KY]

McCarthy, G. & Donchin, E. (1981) A metric for thought: A comparison of P300 latency and reaction time. *Science* 211:77–80. [aZP]

McClelland, J. L. (1976) Preliminary letter identification in the perception of words and nonwords. *Journal of Experimental Psychology: Human Perception and Performance* 3:80–91. [JSB]

(1991) Stochastic interactive processes and the effect of context on perception. *Cognitive Psychology* 23:1–44. [aZP]

McClelland, J. L. & Elman, J. L. (1986) The TRACE model of speech perception. *Cognitive Psychology* 18:1–86. [aZP]

McClelland, J. L. & Rumelhart, D. E. (1981) An interactive model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review* 80:375–407. [TS]

McFarland, D. J. & Cacace, A. T. (1995) Modality specificity as a criterion for diagnosing central auditory processing disorders. *American Journal of Audiology* 4:36–48. [DJM]

(1997) Modality specificity of auditory and visual pattern recognition: Implications for the assessment of central auditory processing disorders. *Audiology* 36:249–60. [DJM]

McGurk, H. & MacDonald, J. (1976) Hearing lips and seeing voices. *Nature* 264:746–48. [VB]

McIntosh, A. R., Grady, C. L., Ungerleider, L. G., Haxby, J. V., Rapoport, S. I. & Horwitz, B. (1994) Network analysis of cortical visual pathways mapped with PET. *Journal of Neuroscience* 14:655–66. [AGr]

McLeod, P., Driver, J. & Crisp, J. (1988) Visual search for a conjunction of movement and form is parallel. *Nature* 332:154–55. [aZP]

McLeod, P., Driver, J., Dienes, Z. & Crisp, J. (1991) Filtering by movement in visual search. *Journal of Experimental Psychology: Human Perception and Performance* 17:55–64. [aZP]

Mel, B. (1997) SEEMORE: Combining color, shape, and texture histogramming in a neurally-inspired approach to visual object recognition. *Neural Computation* 9:777–804. [SE]

Merigan, W. H., Nealey, T. A. & Maunsell, J. H. R. (1992) Qualitatively different effects of lesions of cortical areas V1 and V2 in macaques. *Perception* 21:55–56. [aZP]

Michaels, C. F. & Carello, C. (1981) *Direct perception.* Prentice-Hall. [MRWD, RW]

Millar, S. (1994) *Understanding and representing space: Theory and evidence from studies with blind and sighted space.* Clarendon Press. [EG]

Miller, G. A. (1962) Decision units in the perception of speech. *Institute of Radio Engineers: Transactions on Information Theory* 8:81–83. [aZP]

Miller, G. A., Bruner, J. S. & Postman, L. (1954) Familiarity of letter sequences and tachistoscopic identification. *Journal of General Psychology* 50:129–39. [aZP]

Milner, A. D. & Goodale, M. A. (1995) *The visual brain in action.* Oxford University Press. [aZP]

Mollon, J. D. & Danilova, M. V. (1996) Three remarks on perceptual learning. *Spatial Vision* 10:51–58. [PTS]

Moore, C. M. & Egeth, H. (1998) Perception without attention: Evidence of grouping under conditions of inattention. *Journal of Experimental Psychology: Human Perception and Performance* 23:339–52. [HE]

Morais, J., Bertelson, P., Cary, L. & Alegria, J. (1986) Literacy training and speech segmentation. *Cognition* 24:45–64. [RK]

Morais, J., Cary, L., Algeria, J. & Bertelson, P. (1979) Does awareness of speech as a sequence of phones arise spontaneously? *Cognition* 7:323–31. [RK]

Morais, J. & Kolinsky, R. (1994) Perception and awareness in phonological processing: The case of the phoneme. *Cognition* 50:287–97. [RK]

(1995) The consequences of phonemic awareness. In: *Speech and reading: Comparative approaches,* ed. B. de Gelder & J. Morais. Erlbaum. [RK]

Moran, J. & Desimone, R. (1985) Selective attention gates visual processing in the extrastriate cortex. *Science* 229:782–84. [arZP, JKT, S-LY]

Moray, N. (1970) *Attention: Selective processes in vision and hearing.* Academic Press. [LEK]

Morton, J. (1969) Interaction of information in word recognition. *Psychological Review* 76:165–78. [aZP]

Motter, B. C. (1993) Focal attention produces spatially selective processing in visual cortical areas V1, V2 and V4 in the presence of competing stimuli. *Journal of Neurophysiology* 70:909–919. [PTS]

(1994a) Neural correlates of attentive selection for color or luminance in extrastriate area V4. *Journal of Neuroscience* 14:2178–89. [JB]

(1994b) Neural correlates of feature selective memory and pop-out in extrastriate area V4. *Journal of Neuroscience* 14:2190–99. [JB]

Mountcastle, V., Motter, B., Steinmetz, M. & Sestokas, A. (1987) Common and differential effects of attentive fixation on the excitability of parietal and prestriate (V4) cortical visual neurons in the macaque monkey. *Journal of Neuroscience* 7:2239–55. [aZP]

Muller, H. J. & Findlay, J. M. (1987) Sensitivity and criterion effects in the spatial cuing of visual attention. *Perception and Psychophysics* 42:383–99. [aZP]

Nakayama, K., He, Z. J. & Shimojo, S. (1995) Visual surface representation: A critical link between lower-level and higher-level vision. In: *Visual cognition: An invitation to cognitive science, vol. 2,* ed. S. M. Kosslyn & D. N. Osherson. MIT Press. [CMM, JRP, arZP]

Nakayama, K. & Shimojo, S. (1992) Experiencing and perceiving visual surfaces. *Science* 257:1357–63. [SE]

Nakayama, K. & Silverman, G. H. (1986) Serial and parallel processing of visual feature conjunctions. *Nature* 320:264–65. [aZP]

Naumann, E., Huber, C., Maier, S., Plihal, W., Wustmans, A., Diedrich, O. & Bartussek, D. (1992) The scalp topography of P300 in the visual and auditory modalities: A comparison of three normalization methods and the control of statistical type II error. *Electroencephalography and Clinical Neurophysiology* 83:254–64. [DJM]

Navon, D. (1977) Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology* 9:353–83. [LEK]

Neisser, U. (1967) *Cognitive psychology.* Appleton-Century-Crofts. [LEK]

Newell, A. (1973) You can't play 20 Questions with nature and win: Projective comments on the papers of this symposium. In: *Visual information processing,* ed. W. G. Chase. Academic Press. [CL, rZP]

(1980) HARPY, production systems, and human cognition. In: *Perception and production of fluent speech,* R. A. Cole. Erlbaum. [aZP]

(1990) *Unified theories of cognition.* Harvard University Press. [aZP]

Nisbett, R. E. & Wilson, T. D. (1977) Telling more than we can know: Verbal reports on mental processes. *Psychological Review* 84(3):231–59. [rZP]

Noë, A. (submitted) Action in perception. [AN]

Noë, A., Pessoa, L. & Thompson, E. (in press) Beyond the grand illusion: What change blindness really teaches us about vision. *Vision Cognition.* [AN]

Norris, D. (1987) Strategic control of sentence context effects in naming task. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology* 39A:253–75. [TS]

(1995) Signal detection theory and modularity: On being sensitive to the power of bias models of semantic priming. *Journal of Experimental Psychology: Human Perception and Performance* 21:935–39. [arZP, GR]

Nougier, V., Ripoll, H. & Stein, J.-F. (1989) Orienting of attention with highly skilled athletes. *International Journal of Sport Psychology* 20:205–23. [aZP]

Nowak, L. G. & Bullier, J. (1997) The timing of information transfer in the visual system. In: *Extrastriate visual cortex in primates, vol. 12,* ed. J. H. Kaas, K. L. Rockland & A. L. Peters. Plenum Press. [JB]

Ogle, K. N. (1964) *Researches in binocular vision.* Hafner. [YA]

Oliva, A. & Schyns, P. G. (1997) Coarse blobs, or fine scale edges? Evidence that information diagnosticity changes the perception of complex visual stimuli. *Cognitive Psychology* 34:72–107. [PGS]

O'Regan, J. K. (1992) Solving the "real" mysteries of visual perception: The world as an outside memory. *Canadian Journal of Psychology* 46:461–88. [SE, arZP]

O'Toole, A. J. & Kersten, D. J. (1992) Learning to see random-dot stereograms. *Perception* 21:227–43. [PTS]

Paap, K. R., Newsome, S. L., McDonald, J. E. & Schvaneveldt, R. W. (1982) An activation-verification model for letter and word recognition: The word-superiority effect. *Psychological Review* 89:573–94. [TS]

Pachella, R. G. (1974) The interpretation of reaction time in information processing research. In: *Human information processing: Tutorials in performance and cognition,* ed. B. H. Kantowitz. Erlbaum. [WRU]

(1975) The effect of set on the tachistoscopic recognition of pictures. In: *Attention and Performance V,* ed. P. M. A. Rabbitt & S. Dornic. Academic Press. [HE]

Palmer, S. E. (1975) The effects of contextual scenes on the identification of objects. *Memory and Cognition* 3:519–26. [AH]

(1977) Hierarchical structure in perceptual representation. *Cognitive Psychology* 9:441–74. [RK]

Pani, J. R. (1997) Descriptions of orientation in physical reasoning. *Current Trends in Psychological Science* 6:121–26. [JRP]

Pastore, N. (1949) Need as a determinant of perception. *The Journal of Psychology* 28:457–75. [MAP]

Peacocke, C. (1992) *A study of concepts.* MIT Press. [JLB, rZP]

Perrett, D. I., Harries, M. H., Benson, P. J., Chitty, A. J. & Mistlin, A. J. (1990) Retrieval of structure from rigid and biological motion: An analysis of the visual responses of neurones in the macaque temporal cortex. In: *AI and the eye,* ed. A. Blake & T. Troscianko. Wiley. [aZP]

Perrett, D. I., Mistlin, A. J. & Chitty, A. J. (1987) Visual neurons responsive to faces. *Trends in Neuroscience* 10:358–64. [aZP]

Pessoa, L., Thompson, E. & Noë, A. (1998) Finding out about filling in: A guide to perceptual completion for visual science and the philosophy of perception. *Behavioral and Brain Sciences* 21(6):723–80. [SE, aZP]

Petersik, J. T. (1989) The two-process distinction in apparent motion. *Psychological Bulletin* 106:107–27. [MRWD]

Peterson, M. A. (1986) Illusory concomitant motion in ambiguous stereograms: Evidence for nonsensory components in perceptual organization. *Journal of Experimental Psychology: Human Perception and Performance* 12:50–60. [MAP]

(1994a) Shape recognition can and does occur before figure-ground organization. *Current Directions in Psychological Science* 3:105–111. [MAP]

(1994b) The proper placement of uniform connectedness. *Psychonomic Bulletin and Review* 1:509–14. [MAP]

(1999a) What's in a stage name? *Journal of Experimental Psychology: Human Perception and Performance* 25(1) 276–86. [MAP]

(1999b) High-level vision. *MIT Encyclopedia of Cognitive Science.* [MAP]

Peterson, M. A., De Gelder, B., Rapcsak, S. V., Gerhardstein, P. & Bachoud-Lévi, A. C. (under review) A double dissociation between implicit and explicit object recognition processes in patients with brain damage. [MAP]

Peterson, M. A., Gerhardstein, P., Mennemeier, M. & Rapcsak, S. V. (1998) Object-centered attentional biases and object recognition contributions to scene segmentation in right hemisphere- and left hemisphere-damaged patients. *Psychobiology* 26(4):357–80. [MAP]

Peterson, M. A. & Gibson, B. S. (1991) Directing spatial attention within an object: Altering the functional equivalence of shape descriptions. *Journal of Experimental Psychology: Human Perception and Performance* 17:170–82. [MAP, aZP]

(1993) Shape recognition contributions to figure-ground organization in three-dimensional display. *Cognitive Psychology* 25:383–429. [MAP]

(1994a) Must figure-ground organization precede object recognition? An assumption in peril. *Psychological Science* 5:253–59. [JRP, MAP, TS]

(1994b) Object recognition contributions to figure-ground organization: Operations on outlines and subjective contours. *Perception and Psychophysics* 56:551–64. [MAP]

Peterson, M. A., Harvey, E. H. & Weidenbacher, H. L. (1991) Shape recognition inputs to figure-ground organization: Which route counts? *Journal of Experimental Psychology: Human Perception and Performance* 17:1075–89. [MAP]

Peterson, M. A. & Hochberg, J. (1983) Opposed-set measurement procedure: A quantitative analysis of the role of local cues and intention in form perception. *Journal of Experimental Psychology: Human Perception and Performance* 9:183–93. [MAP, rZP]

Peterson, M. A., Nadel, L., Bloom, P. & Garrett, M. F. (1996) Space and language. In: *Language and space,* ed. P. Bloom, M. A. Peterson, L. Nadel & M. F. Garrett. MIT Press. [MAP]

Pinker, S. (1984) *Visual cognition.* MIT Press. [GR]

(1997) *How the mind works.* Penguin Press. [GR]

Pitcher, G. (1971) *A theory of perception.* Princeton University Press. [1971]

Poggio, T., Torre, V. & Koch, C. (1990) Computational vision and regularization theory. In: *Image understanding 1989,* ed. S. Ullman & W. Richard. Ablex Publishing. [aZP]

Polk, T. A. & Farah, M. J. (1997) A simple common contexts explanation for the development of abstract letter identities. *Neural Computation* 9:1277–89. [JSB]

Pomerantz, J. R. (1983) Global and local precedence: Selective attention in form and vision. *Journal of Experimental Psychology: General* 112(4):516–40. [rZP]

Potter, M. C. (1975) Meaning in visual search. *Science* 187:965–66. [aZP]

Prinzmetal, W., Amiri, H., Allen, K. & Edwards, T. (1998) Phenomenology of attention: 1. Color, location, orientation, and spatial frequency. *Journal of Experimental Psychology: Human Perception and Performance* 24:261–82. [HE]

Proctor, R. W. (1981) A unified theory for matching-task phenomena. *Psychological Review* 88:291–326. [LEK]

Proctor, R. W. & Rao, K. V. (1983) Evidence that the same-different disparity in letter matching is not attributable to response bias. *Perception and Psychophysics* 34:72–76. [LEK]

Proteau, L. (1992) On the specificity of learning and the role of visual information for movement control. In: *Vision and motor control. Advances in psychology no. 85,* ed. L. Proteau & D. Elliot. North-Holland. [aZP]

Pylyshyn, Z. W. (1973) What the mind's eye tells the mind's brain: A critique of mental imagery. *Psychological Bulletin* 80:1–24. [BC, JRP, rZP]

(1978) Imagery and artificial intelligence. In: *Perception and cognition: Issues in the foundations of psychology,* ed. C. W. Savage. University of Minnesota Press. [aZP]

(1979) Validating computational models: A critique of Anderson's indeterminacy of representation claim. *Psychological Review* 86(4):383–94. [rZP]

(1980) Computation and cognition: Issues in the foundations of cognitive science. *Behavioral and Brain Sciences* 3(1):111–32. [BC, MRWD, rZP]

(1981) The imagery debate: Analogue media versus tacit knowledge. *Psychological Review* 88:16–45. [BC, arZP]

(1984) *Computation and cognition: Toward a foundation for cognitive science.* MIT Press. [MRWD, TVP, arZP]

(1989) The role of location indexes in spatial perception: A sketch of the FINST spatial-index model. *Cognition* 32:65–97. [arZP]

(1991) The role of cognitive architectures in theories of cognition. In: *Architectures for intelligence,* ed. K. VanLehn. Erlbaum. [rZP]

(1994) Some primitive mechanisms of spatial attention. *Cognition* 50:363–84. [aZP]

(1998) Visual indexes in spatial vision and imagery. In: *Visual attention,* ed. R. D. Wright. Oxford University Press. [rZP]

(in press) Connecting vision and the word: Tracking the missing link. In: *The foundations of cognitive science,* ed. A. Marques, J. Branquinho & J. Saagua. Oxford University Press. [rZP]

Pylyshyn, Z. W. & Storm, R. W. (1988) Tracking of multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vision* 3:1–19. [arZP]

Ramachandran, V. S. (1976) Learning-like phenomena in stereopsis. *Nature* 262:382–84. [aZP]

(1990) Visual perception in people and machines. In: *AI and the eye,* ed. A. Blake & T. Troscianko. Wiley. [aZP]

Ramachandran, V. S. & Anstis, S. M. (1985) The perception of apparent motion. *Scientific American* 254:102–109. [MRWD]

Ramachandran, V. S. & Braddick, O. (1973) Orientation specific learning in stereopsis. *Perception* 2:371–76. [aZP]

Rayner, K., McConkie, G. W. & Zola, D. (1980) Integrating information across eye movements. *Cognitive Psychology* 12:206–26. [JSB]

Reddy, D. R. (1975) *Speech recognition.* Academic Press. [aZP]

Reicher, G. M. (1969) Perceptual recognition as a function of meaningfulness of stimulus material. *Journal of Experimental Psychology* 81:275–80. [AH]

Reingold, E. M. & Jolicoeur, P. (1993) Perceptual versus postperceptual mediation of visual context effects: Evidence from the letter-superiority effect. *Perception and Psychophysics* 53:166–78. [GR]

Reynolds, R. I. (1981) Perception of an illusory contour as a function of processing time. *Perception* 10:107–15. [aZP]

(1985) The role of object-hypotheses in the organization of fragmented figures. *Perception* 14:49–52. [aZP]

Rhodes, G., Parkin, A. J. & Tremewan, T. (1993) Semantic priming and sensitivity in lexical decision. *Journal of Experimental Psychology: Human Perception and Performance* 19:154–65. [arZP, GR]

Rhodes, G. & Tremewan, T. (1993) The Simon then Garfunkel effect: Semantic priming, sensitivity, and the modularity of face recognition. *Cognitive Psychology* 25:147–87. [aZP, GR]

Richards, C. L., Mitchell, D. B. & Resnick, J. S. (1979) Mental travel: Some reservations. *Journal of Experimental Psychology: Human Perception and Performance* 5:13–18. [BC]

Richards, W. (1988) *Natural computation.* MIT Press (A Bradford Book). [aZP]

Richards, W. A., Jepson, A. & Feldman, J. (1996) Priors, preferences, and categorical percepts. In: *Perception as Bayesian inference,* ed. D. C. Knill & W. A. Richards. Cambridge University Press. [MS]

Riseman, E. M. & Hanson, A. R. (1987) A methodology for the development of general knowledge-based vision systems. In: *Vision, brain, and cooperative computation,* ed. M. A. Arbib & A. R. Hanson. MIT Press (A Bradford Book). [aZP]

Roberts, L. G. (1965) Machine perception of three-dimensional solids. In: *Optical and electro-optical information processing*, ed. J. P. Tippett. MIT Press. [aZP]

Rock, I. (1983) *The logic of perception.* MIT Press, a Bradford Book. [JLD, CMM, aZP]

(1984) *Perception.* Scientific American Books. Freeman. [GV]

(1997) *Indirect perception.* MIT Press. [JLD, CMM, aZP]

Rock, I. & Anson, R. (1979) Illusory contours as the solution to a problem. *Perception* 8:655–81. [aZP]

Rock, I. & Smith, D. (1981) Alternative solutions to kinetic stimulus transformations. *Journal of Experimental Psychology: Human Perception and Performance* 7:19–29. [JLD]

Rosch, E. H., Mervis, C. B., Gray, W. D., Johnson, D. M. & Boyes-Braem, P. (1976) Basic objects in natural categories. *Cognitive Psychology* 8:382–439. [arZP]

Rosenblatt, F. (1959) Two theorems of statistical separability in the perceptron. In: *Symposium on mechanization of thought processes.* Her Majesty's Stationery Office, London. [aZP]

Rosenfeld, A. (1987) Recognizing unexpected objects: A proposed approach. *International Journal of Pattern Recognition and Artificial Intelligence* 1:71–84. [AR]

(1989) Computer vision: A source of models for biological visual processes? *IEEE Transactions on Biomedical Engineering* 36:93–96. [AR]

(1990) Pyramid algorithms for finding global structures in images. *Information Sciences* 50:23–34. [AR]

Rosenfeld, A., Hummel, R. A. & Zucker, S. W. (1976) Scene labeling by relaxation operators. *IEEE Transactions on Systems, Man, and Cybernetics,* SMC-6:420–33. [aZP]

Rossetti, Y. (1998) Short-lived motor representation of space in brain-damaged and healthy subjects. *Consciousness and Cognition* 7:520–80. [EG]

(1999) In search of immaculate perception: Evidence from motor representation of space. In: *Toward a science of consciousness,* ed. S. Hameroff, A. Kaszniak & A. Scott. MIT Press. [EG]

Rossetti, Y. & Procyk, E. (1997) What memory is for action: The gap between percepts and concepts. *Behavioral and Brain Sciences* 20:34–36. [EG]

Rumelhart, D. E. (1977) *Human information processing.* Wiley. [aZP]

Runeson, S. (1977) On the possibility of "smart" perceptual mechanisms. *Scandinavian Journal of Psychology* 18:172–79. [RW]

Saarinen, J. & Levi, D. (1995) Orientation anisotropy in vernier acuity. *Vision Research* 35:2449–61. [EG]

Samuel, A. G. (1981) Phonemic restoration: Insights from a new methodology. *Journal of Experimental Psychology: General* 110:474–94. [LEK, aZP]

(1996) Does lexical information influence the perceptual restoration of phonemes? *Journal of Experimental Psychology: General* 125:28–51. [aZP]

Sanocki, T. (1993) Time course of object identification: Evidence for a global-to-local contingency. *Journal of Experimental Psychology: Human Perception and Performance* 19:878–98. [TS]

(1997) Structural contingencies and shifts of object-based attention during object recognition. *Journal of Experimental Psychology: Human Perception and Performance* 23:780–807. [TS]

Sanocki, T., Bowyer, K., Heath, M. & Sarkar, S. (1998) Are edges sufficient for object recognition? *Journal of Experimental Psychology: Human Perception and Performance* 24:340–49. [TS]

Sanocki, T. & Epstein, W. (1997) Priming spatial layout of scenes. *Psychological Science* 8:374–78. [TS]

Sanocki, T. & Oden, G. C. (1984) Contextual validity and the effects of low constraint sentence contexts on lexical decisions. *Quarterly Journal of Experimental Psychology* 36A:145–56. [TS]

Sarter, M., Berntson, G. G. & Cacioppo, J. T. (1996) Brain imaging and cognitive neuroscience: Toward strong inference in attributing function to structure. *American Psychologist* 51:13–21. [BC]

Saye, A. & Frisby, J. P. (1975) The role of monocularly conspicuous features in facilitating stereopsis from random-dot stereograms. *Perception* 4:159–71. [aZP]

Schacter, D. L. (1992) Priming and multiple memory systems: Perceptual mechanisms of implicit memory. *Journal of Cognitive Neuroscience* 4:244–56. [AH]

Schall, J. D., Hanes, D. P., Thompson, K. G. & King, D. J. (1995) Saccade target selection in frontal eye field of macaque. I. Visual and premovement activation. *Journal of Neuroscience* 15:6905–18. [JB]

Schiller, P. H. & Lee, K. (1991) The role of the primate extrastriate area V4 in vision. *Science* 251:1251–53. [AGr]

Schmolesky, M. T., Wang, Y., Hanes, D. P., Thompson, K. G., Leutgeb, S., Schall, J. D. & Leventhal, A. G. (1998) Signal timing across the macaque visual system. *Journal of Neurophysiology* 79:3272–78. [JB]

Schweinberger, S. R. (1996) How Gorbachev primed Yeltsin: Analyses of associative priming in person recognition by means of reaction times and event-related brain potentials. *Journal of Experimental Psychology: Learning, Memory and Cognition* 22:1383–407. [GR]

Schyns, P. G. (1998) Diagnostic recognition: Task constraints, object information and their interactions. *Cognition* 67:147–79. [PGS]

Schyns, P. G., Goldstone, R. L. & Thibaut, J.-P. (1998) The development of features in object concepts. *Behavioral and Brain Sciences* 21:1–54. [MRWD, aZP, PTS]

Schyns, P. G. & Oliva, A. (1994) From blobs to boundary edges: Evidence for time and scale dependent scene recognition. *Psychological Science* 5:195–200. [PGS]

(1998) Dr. Jekyll and Mr. Smile: When categorization flexibly modifies the perception of faces in rapid visual presentations. *Cognition* (in press). [PGS]

Schyns, P. G. & Rodet, L. (1997) Categorization creates functional features. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 23:681–96. [PTS]

Scliar-Cabral, L., Morais, J., Nepomuceno, L. & Kolinsky, R. (1997) The awareness of phonemes: So close - so far away. *International Journal of Psycholinguistics* 13:211–40. [RK]

Seidenberg, M. S. (1987) Sublexical structures in visual word recognition: Access units or orthographic redundancy? In: *Attention and performance XII: The psychology of reading,* ed. M. Coltheart. Erlbaum. [JSB]

Sekuler, A. B. & Palmer, S. E. (1992) Visual completion of partly occluded objects: A microgenetic analysis. *Journal of Experimental Psychology: General* 121:95–111. [aZP]

Sekuler, R., Anstis, S., Braddick, O. J., Brandt, T., Movshon, J. A. & Orban, G. (1990) The perception of motion. In: *Visual perception: The neurophysiological foundations,* ed. L. Spillman & J. S. Werner. Academic Press. [MRWD]

Sekuler, R. & Blake, R. (1994) *Perception.* McGraw-Hill. [aZP]

Selfridge, O. (1959) Pandemonium: A paradigm for learning. *Symposium on Mechanization of Thought Processes: National Physical Laboratory Symposium.* Her Majesty's Stationery Office, London. [aZP]

Shaw, M. L. (1984) Division of attention among spatial locations: A fundamental difference between detection of letters and detection of luminance increments. In: *Attention and performance, X,* ed. H. Bouma & D. G. Bouwhuis. Erlbaum. [aZP]

Shaw, R. E. & McIntyre, M. (1974) The algoristic foundations to cognitive psychology. In: *Cognition and the symbolic processes,* ed. W. B. Weimer & D. S. Palermo. Erlbaum. [RW]

Sheinberg, D. L. & Logothetis, N. K. (1997) The role of temporal cortical areas in perceptual organization. *Proceedings of the National Academy Science USA* 94:3408–13. [JB]

Shiffrar, M. & Freyd, J. J. (1990) Apparent motion of the human body. *Psychological Science* 1:257–64. [KY]

Shih, S.-I. & Sperling, G. (1996) Is there feature-based attentional selection in visual search? *Journal of Experimental Psychology: Human Perception and Performance* 22:758–79. [HE]

Shimojo, S. & Nakayama, K. (1990a) Amodal representation of occluded surface: Role of invisible stimuli in apparent motion correspondence. *Perception* 19:285–99. [KY]

(1990b) Real world occlusion constraints and binocular rivalry. *Vision Research* 30:69–80. [KY]

Shimojo, S., Silverman, G. H. & Nakayama, K. (1989) Occlusion and the solution to the aperture problem for motion. *Vision Research* 29:619–26. [GV]

Shirai, Y. (1975) Analyzing intensity arrays using knowledge about scenes. In: *Psychology of computer vision,* ed. P. H. Winston. MIT Press. [aZP]

Shiu, L. & Pashler, H. (1992) Improvement in line orientation discrimination is retinally local but dependent on cognitive set. *Perception and Psychophysics* 52:582–88. [PTS]

Shulman, G. L. & Wilson, J. (1987) Spatial frequency and selective attention to local and global information. *Perception* 16:89–101. [aZP]

Shulman, H. G., Hornak, R. & Sanders, E. (1978) The effects of graphemic, phonetic, and semantic relationships on access to lexical structure. *Memory and Cognition* 6(2):115–23. [rZP]

Sigman, E. & Rock, I. (1974) Stroboscopic movement based on perceptual intelligence. *Perception* 3:9–28. [aZP, KY]

Sillito, A. M., Jones, H. E., Gerstein, G. L. & West, D. C. (1994) Feature-linked synchronization of thalamic relay cell firing induced by feedback from the visual cortex. *Nature* 369:479–82. [aZP]

Simon, H. A. (1969a) *The sciences of the artificial.* MIT Press. [rZP]

(1969b) The architecture of complexity. In: *The sciences of the artificial.* MIT Press. [GV]

Simons, D. J. (1996) In sight, out of mind: When object representations fail. *Psychological Science* 7:301–305. [SE]

Singh, M. & Hoffman, D. D. (1998) Part boundaries alter the perception of transparency. *Psychological Science* 9:370–78. [MS]

Snodgrass, J. G. & Feenan, K. (1990) Priming effects in picture fragment completion: Support for the perceptual closure hypothesis. *Journal of Experimental Psychology: General* 119:276–96. [aZP]

Snyder, L., Batista, A. & Andersen, R. A. (1997) Coding of intention in the posterior parietal cortex. *Nature* 386:167–70. [AGr]

Soloman, R. L. & Postman, L. (1952) Frequency of usage as a determinant of recognition thresholds for words. *Journal of Experimental Psychology* 43:195–201. [aZP]

Sowden, P. T., Davies, I. R. L. & Roling, P. (submitted) Perceptual learning of the detection of features in X-ray images: A functional role for the improvements in adults' visual sensitivity? *Journal of Experimental Psychology: Human Perception and Performance.* [PTS]

Sowden, P. T., Davies, I. R. L., Rose, D. & Kaye, M. (1996) Perceptual learning of stereoacuity. *Perception* 25:1043–52. [PTS]

Spelke, E. S. (1990) Object perception. In: *Readings in philosophy and cognitive science,* ed. A. I. Goldman. MIT Press. [JLB]

Stadler, M. & Kruse, P. (1995) The function of meaning in cognitive order formation. In: *Ambiguity of mind and nature: Multistable cognitive phenomena,* ed. P. Kruse & M. Stadler. Springer-Verlag. [TVP]

Stark, L. & Ellis, S. R. (1981) Scanpaths revisited: Cognitive models direct active looking. In: *Eye movements: Cognition and visual perception,* ed. D. Fisher, R. Monty & J. Senders. Erlbaum. [aZP]

Starkes, J., Allard, F., Lindley, S. & O'Reilly, K. (1994) Abilities and skill in basketball. Special issue: Expert-novice differences in sport. *International Journal of Sport Psychology* 25:249–65. [aZP]

Sternberg, S. (1969) The discovery of processing stages: Extensions of Donders' method. *Acta Psychologica* 30:276–315. [aZP]

(1998) Discovering mental processing stages: The method of additive factors. In: *Methods, models, and conceptual issues,* ed. D. Scarborough & S. Sternberg. MIT Press (A Bradford Book). [aZP]

Stich, S. (1978) Beliefs and subdoxastic states. *Philosophy of Science* 45:499–518. [rZP]

Stoerig, P. (1996) Varieties of vision: From blind responses to conscious recognition. *Trends in Neuroscience* 19:401–406. [AGe]

Stone, G. O., Vanhow, M. & Van Orden, G. C. (1997) Perception is a two-way street: Feedforward and feedback phonology in visual word recognition. *Journal of Memory and Language* 36:337–59. [JSB]

Stoner, G. R. & Albright, T. D. (1993) Image segmentation cues in motion processing: Implications for modularity in vision. *Journal of Cognitive Neuroscience* 5:129–49. [GV]

Stoner, G. R., Albright, T. D. & Ramachandran, V. S. (1990) Transparency and coherence in human motion perception. *Nature* 344:153–55. [GV]

Street, R. F. (1931) *A gestalt completion test: A study of cross section of intellect.* Bureau of Publications, Teachers College, Columbia University. [aZP]

Streri, A. (1993) *Seeing, reaching, touching. The relations between vision and touch in infancy.* Harvester Wheatsheaf. [EG]

Stroop, J. R. (1935) Studies of interference in serial verbal reactions. *Journal of Experimental Psychology* 18:643–62. [CL]

Swets, J. A. (1998) Separating discrimination and decision in detection, recognition, and matters of life and death. In: *Methods, models, and conceptual issues,* ed. D. Scarborough & S. Sternberg. MIT Press (A Bradford Book). [aZP]

Swinney, D. A. (1979) Lexical access during sentence comprehension: (Re)consideration of context effects. *Journal of Verbal Learning and Verbal Behavior* 18(6):645–59. [rZP]

Tanner, W. P. & Swets, J. A. (1954) A decision-making theory of human detection. *Psychological Review* 61:401–409. [aZP]

Thompson, E. (1995) *Colour vision: A study in cognitive science and the philosophy of perception.* Routledge.Press. [AN]

Thompson, E., Palacios, A. & Varela, F. J. (1992) Ways of coloring: Comparative color vision as a case study for cognitive science. *Behavioral and Brain Sciences* 15:1–74. [AN]

Thorpe, S., Fize, D. & Marlot, C. (1996) Speed of processing in the human visual system. *Nature* 381:520–22. [JB]

Tittle, J. S., Todd, J. T., Perotti, V. J. & Norman, J. F. (1995) Systematic distortion of perceived three-dimensional structure from motion and binocular stereopsis. *Journal of Experimental Psychology: Human Perception and Performance* 21:663–78. [YA]

Todd, J. (1984) The perception of three-dimensional structure from rigid and nonrigid motion. *Perception and Psychophysics* 36:97–103. [JLD]

Tommasi, L. & Vallortigara, G. (1999) Figure-ground segregation modulates perceived direction of ambiguous moving gratings and plaids. *Vision Research* 39:777–87. [GV]

Treisman, A. (1985) Preattentive processing. *Computer Vision Graphics Image Processing* 31:156–77. [CMM]

(1988) Features and objects: The fourteenth Bartlett memorial lecture. *The Quarterly Journal of Experimental Psychology* 40A:201–37. [DJC, LEK, arZP]

(1993) The perception of features and objects. In: *Attention: Selection, awareness and control: A tribute to Donald Broadbent,* ed. A. Baddeley & L. Weiskrantz. Clarendon Press. [S-LY]

Treisman, A. & Gormican, S. (1988) Feature analysis in early vision: Evidence from search asymmetries. *Psychological Review* 95:15–48. [LEK]

Treue, S., Andersen, R. A., Ando, H. & Hildreth, E. C. (1995) Structure-from-motion: Perceptual evidence for surface interpolation. *Vision Research* 35:139–48. [SE]

Treue, S. & Maunsell, J. H. R. (1996) Attentional modulation of visual motion processing in cortical areas MT and MST. *Nature* 382:539–41. [JB, JKT, S-LY]

Trueswell, J. C. & Hayhoe, M. M. (1993) Surface segmentation mechanisms and motion perception. *Vision Research* 33:313–28. [GV]

Tsotsos, J. K. (1990) A complexity level analysis of vision. *Behavioral and Brain Sciences* 13:423–55. [JKT]

Tsotsos, J. K., Culhane, S., Wai, W., Lai, Y., Davis, N. & Nuflo, F. (1995) Modeling visual attention via selective tuning. *Artificial Intelligence* 78(1–2):507–47. [JKT]

Turvey, M. T., Shaw, R. E., Reed, E. S. & Mace, W. M. (1981) Ecological laws of perceiving and acting: In reply to Fodor and Pylyshyn (1981). *Cognition* 9:237–304. [MRWD]

Tweedy, J. R., Lapinski, R. H. & Schvaneveldt, R. W. (1977) Semantic-context effects on word recognition: Influence of varying the proportion of items presented in an appropriate context. *Memory and Cognition* 5:84–89. [TS]

Tye, M. (1991) *The imagery debate.* MIT Press. [BC]

Ullman, S. (1976) On visual detection of light sources. *Biological Cybernetics* 21:205–12. [aZP]

(1979) *The interpretation of visual motion.* MIT Press. [JLD, MRWD, aZP]

(1980) Against direct perception. *Behavioral and Brain Sciences* 3:373–415. [MRWD]

(1989) Aligning pictorial descriptions: An approach to object recognition. *Cognition* 32:193–254. [TS]

Ullman, S. & Richards, W. (1990) *Image understanding.* Ablex Publishers. [aZP]

Ungerleider, L. G. & Mishkin, M. (1982) Two cortical visual systems. In: *Analysis of visual behavior,* ed. J. Ingle, M. A. Goodale & R. J. W. Mansfield. MIT Press. [aZP]

Uttal, W. R. (1978) *The psychobiology of mind.* Erlbaum. [WRU]

(1998) *Toward a new behaviorism: The case against perceptual reductionism.* Erlbaum. [WRU]

Uttley, A. M. (1959) Conditional probability computing in the nervous system. *Symposium on Mechanization of Thought Processes.* Her Majesty's Stationery Office, London. [aZP]

Vallortigara, G. & Bressan, P. (1991) Occlusion and the perception of coherent motion. *Vision Research* 31:1967–78. [GV]

(1994) Occlusion, transparency, and steriopsis: A new explanation for stereo capture. *Vision Research* 21:2891–96. [GV]

Van Der Heijden, A. H. C. (1992) *Selective attention in vision.* Routledge. [AGe]

Van Essen, D. C. & Anderson, C. H. (1990) Information processing strategies and pathways in the primate retina and visual cortex. In: *Introduction to neural and electronic networks,* ed. S. F. Zornetzer, J. L. Davis & C. Lau. Academic Press. [aZP]

Varela, F. J., Thompson, E. & Rosch, E. (1991) *The embodied mind: Cognitive science and human experience.* MIT Press. [AN]

Verhaeghe, A. & Kolinsky, R. (1992) Discriminaçõ entre figuras orientadas em esphelho em funçõ do modo de apresentaçõ em adultos escolarisados e adultos iletrados. *Proceedings of the I Jornadas de Estudo dos Processos Cognitivos da Sociedade Portuguesa de Psicologia,* Astoria. [RK]

Verleger, R. (1988) Event-related potentials and cognition: A critique of the context updating hypothesis and an alternative interpretation of P3. *Behavioral and Brain Sciences* 11:343–427. [aZP]

Vogels, R. & Orban, G. A. (1985) The effect of practice on the oblique effect in line orientation judgements. *Vision Research* 25:1679–97. [EG]

Von der Heydt, R. & Peterhans, E. (1989) Mechanisms of contour perception in monkey visual cortex. I. Lines of pattern discontinuity. *Journal of Neuroscience* 9:1731–48. [TVP]

Von der Heydt, R., Peterhans, E. & Baumgartner, G. (1984) Illusory contours and cortical neuron responses. *Science* 224:1260–62. [TVP]

Wade, N. J., ed. (1983) *Brewster and Wheatstone on vision.* Academic Press. [VB]

Wagemans, J. (1988) Modules in vision: A case study of interdisciplinary in cognitive science. *Acta Psychologica* 67:59–93. [MRWD]

Wallach, H. (1949) Some considerations concerning the relation between perception and cognition. *Journal of Personality* 18:6–13. [aZP]

Wallach, H. & O'Connell, D. N. (1953) The kinetic depth effect. *Journal of Experimental Psychology* 45:205–17. [aZP]

Warrington, E. K. (1982) Neurological studies of object recognition. *Philosophical Transactions of the Royal Society of London B: Biological Science* 298(1089):15–33. [rZP]

Warrington, E. K. & Taylor, A. M. (1978) Two categorical stages of object perception. *Perception* 7(6):695–705. [rZP]

Weiskrantz, L. (1995) Blindsight: Not an island unto itself. *Current Directions in Psychological Science* 4:146–51. [aZP]

Weiskrantz, L., Warrington, E., Sanders, M. D. & Marshall, J. (1974) Visual capacity in the hemianopic field following restricted occipital ablation. *Brain* 97:709–29. [aZP]

Wickens, C. D., Kramer, A. F. & Donchin, E. (1984) The event-related potential as an index of processing demands of a complex target acquisition task. *Annals of the New York Academy of Sciences* 425:295–99. [aZP]

Wilson, J. A. & Robinson, J. O. (1986) The impossibly-twisted Pulfrich pendulum. *Perception* 15:503–504. [aZP]

Winston, P. H. (1974) *New progress in artificial intelligence*. MIT Artificial Intelligence Laboratory. [aZP]

Wong, E. & Mack, A. (1981) Saccadic programming and perceived location. *Acta Psychologia* 48:123–31. [aZP]

Woods, W. A. (1978) Theory formation and control in a speech understanding system with extrapolations towards vision. In: *Computer vision systems: Papers from the workshop on computer vision systems.* Academic Press. [aZP]

Woodworth, R. S. (1938) *Experimental psychology.* Holt. [LEK]

Wright, R. D. & Dawson, M. R. W. (1994) To what extent do beliefs affect apparent motion? *Philosophical Psychology* 7:471–91. [MRWD]

Wright, R. D., Dawson, M. R. W. & Pylyshyn, Z. W. (1987) Spatio-temporal parameters and the three-dimensionality of apparent motion: Evidence for two types of processing. *Spatial Vision* 2:263–72. [MRWD]

Yeh, S. L., Chen, I. P., De Valois, K. K. & De Valois, R. L. (1996) Figural aftereffects and spatial attention. *Journal of Experimental Psychology: Human Perception and Performance* 22:446–60. [S-LY]

Young, A. W., Newcombe, F., de Haan, E. H. F., Small, M. & Hay, D. C. (1993) Face perception after brain injury: Selective impairments affecting identity and expression. *Brain* 116:941–59. [VB]

Yu, K. (1994) Does semantic knowledge influence motion perception? Unpublished doctoral dissertation, Vanderbilt University, Nashville, TN. [KY]

Zanforlin, M. (1982) Figure organization and binocular interaction. In: *Organization and representation in perception,* ed. J. Beck. Erlbaum. [GV]

Zeki, S. M. (1993) *A vision of the brain.* Blackwell. [YA]

Zucker, S. W., Rosenfelf, A. & David, L. S. (1975) *General purpose models: Expectations about the unexpected.* Paper presented at the Fourth International Joint Conference on Artificial Intelligence, Tbilisi, Georgia (former USSR). [arZP]