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

Address correspondence to:

Zenon Pylyshyn
Rutgers Center for Cognitive Science
Rutgers University
Psychology Addition, Busch Campus,
New Brunswick, NJ 08903

Email: zenon@ruccs.rutgers.edu

IS VISION CONTINUOUS WITH COGNITION? THE CASE FOR COGNITIVE IMPENETRABILITY OF VISUAL PERCEPTION*

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 paper 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, including the use of signal detection theory and event-related potentials to assess cognitive penetration of vision.

A distinction is made among several stages in visual processing. These include, 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 that have been presented in both computer vision and psychology showing that vision is "intelligent" and involves elements of "problem solving". It is suggested that the cases of apparently intelligent interpretation that are sometimes cited in support of this claim do not show cognitive penetration, but rather that 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 identification of the stimulus.

* 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 each draft and providing useful additions. This work was initially supported by a grant from the Alfred P. Sloan Foundation and recently by the Rutgers Center for Cognitive Science.

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. And 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 that we receive at our eyes, together with our hard-wired visual system, or is it also to a large extent because those are the things we expect to see or the things we 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 19th century. At one extreme is the view that perception consists essentially of building larger and larger structures from elementary retinal or sensory features. Another view accepts this hierarchical picture but allows centripetal or top-down influences within a circumscribed part of vision. Then there is the unconscious-inference view initially proposed by von Helmholtz and rehabilitated in modern times by Bruner's 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), 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 positions on the nature 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 of this paper, 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 by 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

[1] This thesis is closely related to what Fodor (1983) has called the "modularity of mind" view and this paper owes much to Fodor's ideas. Because there are several independent notions conflated in the general usage of the term "module", we shall eschew the use of this term to designate cognitively impenetrable systems in this paper.

[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 primary visual cortex. Our usage is guided by an attempt to distinguish a functionally distinct system, regardless of its neuroanatomy. By 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.

utilities on the content of visual perception[4] and a cognitively mediated directing of the visual system (through focal attention) towards 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 and show that it is much more complex than the output of sensors and that it is likely 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 for 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 it in a later review paper (Bruner, 1957), the "*Value and needs…"* 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." Perception, according to Bruner, is characterized by two essential properties: it is categorical and it is inferential. Thus perception might 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]

There were literally thousands of experiments performed from the 1950s through the 1970's showing 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, Bruner 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.

---

[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 to refer to 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, in fact, play a central role in the present argument for the distinction between perception and cognition.

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, Bruner & Postman, 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 non-word 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 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 higher than the time 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 which investigated the hypothesize-and-test view more directly. One way this was done was by manipulating the "availability" of perceptual hypotheses. For example, Bruner & Minturn (1955) manipulated the readiness of the hypothesis that stimuli were numbers vs 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 the 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 within philosophy of science. Thomas gathered a cult following with his view of scientific revolutions, in which theory change was seen as guided more by social considerations than by new data, since theoretical constructs in different theories were seen as essentially incommensurable. Philosophers of science like Hanson (1958), Feyerabend (1962) and Kuhn (1972) argued that there was no such thing as objective observation since 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 believable 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 paper

Both the experimental and 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 do 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 within 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 cases where the top-down effect is a within-vision effect — i.e., visual interpretations computed by early vision affect other visual interpretations, separated either by space or time. The sort of influence that concerns us in this paper is one that originates outside the visual system and that appears to affect 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). 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, i.e., 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 the 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's 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* (see, however, note 2). Many processes other than those of early vision, however, enter into the construction of a 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 3D 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, Thompson & Noë, in press). What this means is that the interpretation of parts of a stimulus may depend upon 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 Sections 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 lots of other things.)

Note that the 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

[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, 1986)). 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, implemented as a subset of long-term memory.

perspective lines are added to form the Ponzo illusion, and so on. It's not just 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 really know they 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 upon how knowledgeable you are, what other sources of information you have, what your utilities are (what's 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 different vocabulary 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's why in computer vision a major part of early vision is concerned with what is called scene labeling or label-propagation (Rosenfeld, Hummel & Zucker, 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 the display. 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 (1997/1973) 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 Section 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 which is phenomenally quite striking). 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 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 & 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 and 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.

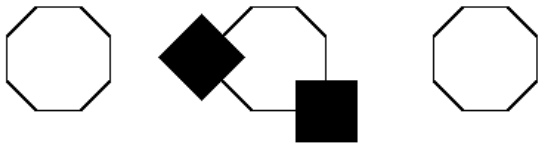

Figure 1: Kanizsa amodal completion figure. The completion preferred by the visual system is not the simplest figure despite the flanking examples. After Kanizsa (1985).

3. There is a great deal of evidence from neuroscience which 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 Section 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, while 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 paper 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 has 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 at least 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, since 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 & 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)—i.e., their analysis of images is guided by some stored model. Although model-based systems may not use general knowledge and draw inferences, they fall in 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, Rosenfeld & Davis, 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 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 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 since 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, Scherrer, & Jouvet, 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, Maunsell & Schiller, 1988; Moran & Desimone, 1985; Mountcastle, Motter, Steinmetz & Sestokas, 1987; van Essen & Anderson, 1990; Sillito, Jones, Gerstein, West, 1994). Indeed there is recent evidence that attention can have long term effects (Desimone, 1996; xx) 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 — i.e., they constitute an early attentional selection of relevant properties (typically location, but see Section 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). 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 & 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. While 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, Mistlin & Chitty, 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 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, Harries, Benson, Chitty & Mistlin (1990) describe cells that fire to the visual event of an experimenter "leaving the room" — and not to comparable movements that are not directed towards the door. Such cells clearly encode a complex class of events (perhaps involving the relational property "towards 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 (p195), "The particular significance of long-term disappearance of an experiment …

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 which 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 skills. 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 & Riddoch (1987). After suffering a stroke which resulted in bilateral damage to his occipital lope, 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 which can serve as the basis for recognition. As Humphreys & 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.

==========///==========

In the preceding pages 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 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. This section is necessarily more technical and may be skipped if the reader is less interested in methodological issues.

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 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 Hollingsworth 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 which 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) (Tanner & Swets, 1954; Swets, 1998), 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 Section 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 is 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 was 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 — i.e, 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 1950's and 1960's 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 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 set of stimuli, a generalization often called Hick'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, El-Zahhar & Alcorn (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 state 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 "phoneme 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* to 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'* which he attributed to prosodic differences between "replace" and "added" stimuli) but he did find significant 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 & Tremewan (1993) and Rhodes, Parkin & Tremewan (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, Mezzanotte & Rabinowitz (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 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

[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 essay this issue will not be raised further.

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 consist 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 & 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, McCarthy, & Donchin (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 & 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 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, Kramer & Donchin, 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 (readers not interested in the use of signal detection theory may wish to skip this section): 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 conjuncts 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 proto-hypotheses 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. Since in principle these 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 between Farah (1989), Rhodes, Parkin & Tremewan (1993), and Norris (1995) shows. Norris (1995) argued that differences in *d'* cannot be taken to show that the perceptual process is being affected since *d'* differences can be obtained when the effect is generated by a process that imposes a criterion shift alone. Similarly, Hollingworth & 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 these 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' 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.

Hollingsworth & Henderson (in press) have argued that the Biederman, Mezzanotte & Rabinowitz (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. Hollingsworth & 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 — i.e., 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 & 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 a very general one and is 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 $<F_i; C_i>$ and $<F_j; C_j>$ than to feature vectors $<F_i; C_j>$ and $<F_j; C_i>$ (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 these categories either will not occur in 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 phoneme 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,

[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.

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— ran 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

[10] We cannot think of this as an "absolute" change in sensitivity to the class of stimuli since 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.

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 — i.e., 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, Scharf & Buus, 1991) and in the spatial domain (Shulman & Wilson, 1987; Julesz & Papathomas, 1984), as well as on features defined by color (Green & Anderson, 1956; Friedman-Hill & Wolfe, 1995), shape (Egeth, Vizri & Garbart, 1984), motion (McCleod, Driver, Dienes & Crisp, 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 may 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 (Lupker & Massaro, 1979; Shaw, 1984; Muller & Findlay, 1987; Bonnel, Possamai & Schmidt, 1987; and Downing, 1988). 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 Section 6.4), it is much more dubious 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 phoneme-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 usual 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" a 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 3 dimensional world to our 2 dimensional retinas is many-to-one and therefore non-invertable. 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 3D-to-2D 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 non-invertibility 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 which 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, Marr,

[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 paper, 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.

1982, Ullman & Richards, 1990, Brown, 1984, Richards, 1988). 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 3 or more (orthographic) 2-D views of 4 noncoplanar points which maintain fixed 3-D interpoint distances, uniquely determines 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 which, 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 that the 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 — i.e., 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 *appeal* to the constraint: it simply does what it is wired to do, which, as it happens, means that it works in accordance 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 that points which 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 which 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.

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,

[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.

has become an important principle in computer vision. 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, Hoffman & Prakash, 1989). Although there are different ways to state the constraints — e.g., in terms of properties of the world or in terms of some world-independent mathematical principle, such as "regularization" (Poggio, Torre & Koch, 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., Marr & Poggio, 1979; Rosenfeld, Hummel & Zucker, 1976; Dawson & Pylyshyn, 1988).

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 that 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". In what follows we shall 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. Figure 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 different 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

[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 over-rides the usual early vision mechanisms. This could indeed be a case of cognitive over-ride. 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, McDermott & Chun, 1997).

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 light common 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, Ullman, 1976; Maloney & Wandell, 1990).

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 & 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 is not perceived (Figure 2b). Interestingly, if the "covering" surface presented over the phi dots is perceived as a transparent surface, then the illusory phi motion persists (Figure 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.

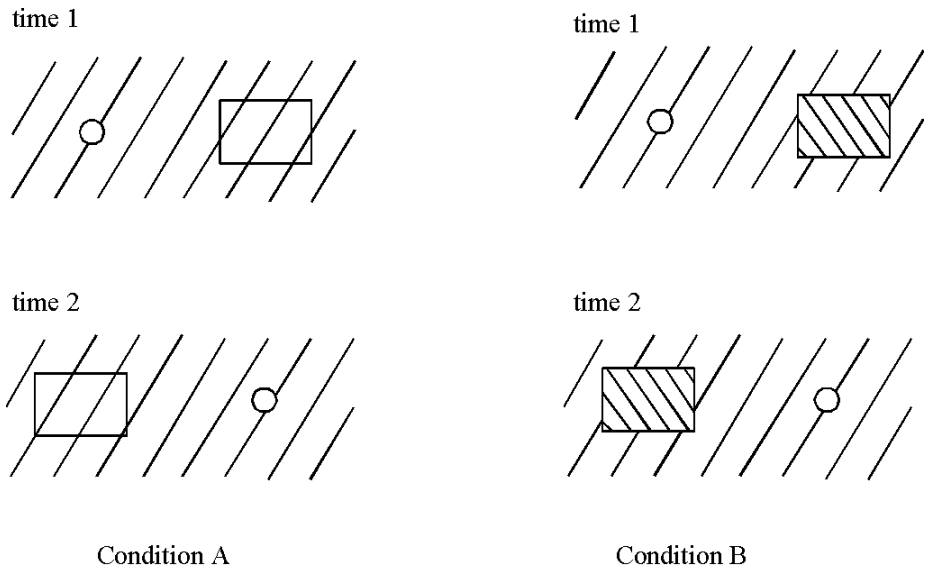

Figure 2: In the left figure (Condition A), the texture seen through the rectangle makes it appear to be an outline, so phi motion is perceived, whereas in the figure on the right (Condition B), the distinct texture on the rectangle makes it appear opaque, so phi motion is not perceived (after Sigman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

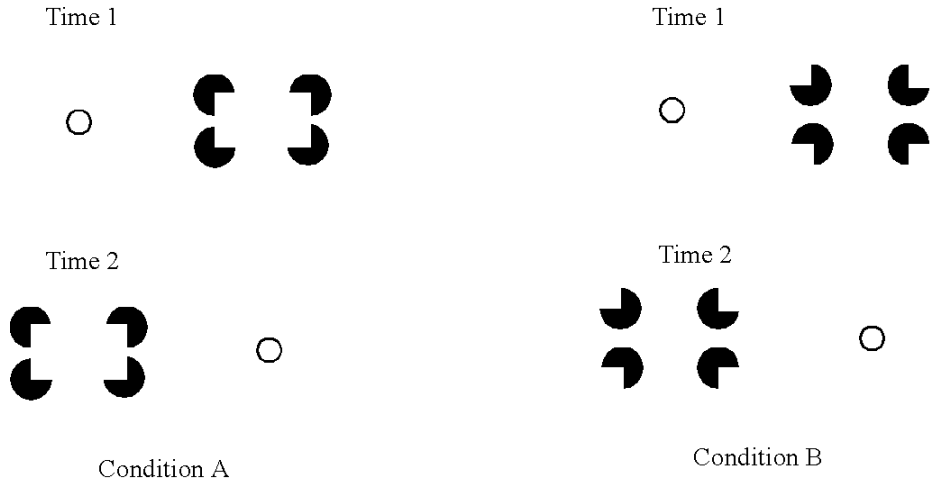

Figure 3: In the figure on the left (Condition A), the illusory rectangle appears to be opaque and to alternately cover the two dots so apparent motion is not perceived, whereas in the control display on the right (Condition B), apparent motion is perceived (after Sigman & Rock, 1974).

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 Figure 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).

What such examples are taken to show is 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. (a) A natural constraints 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 take 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. (b) 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 them 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 as 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, p319). 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 that 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 in 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, Rock & Egatz, 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.

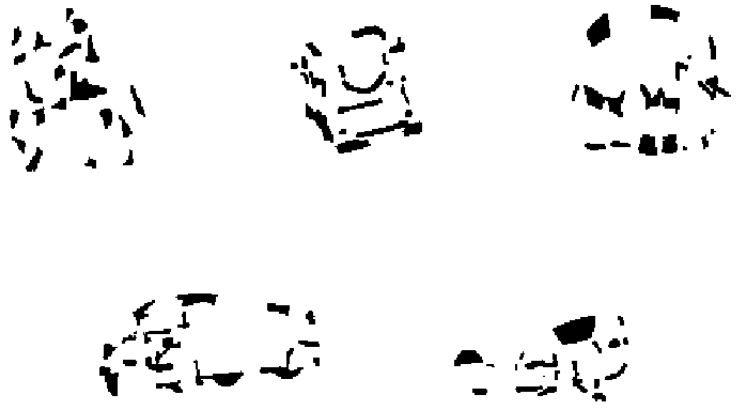

Figure 4: Examples of fragmented or "closure" figures, taken from Street's (1931) "Gestalt Completion Test".

Although verbal hints may have little effect on recognizing fragmented figures, other kinds of information can be beneficial. For example, Snodgrass & 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 may 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 (in 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, since it represents what is really a pre-perceptual 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. Stark & Ellis, 1981; Gale & Findlay (1983) 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 images, 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 while other parts have a bias towards 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 & 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 & 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 & Clatworthy (1975) found that repeated presentation had a beneficial effect that lasted for at least 3 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 notice patterns that novices cannot see (bird watchers, art authenticators, radiologists, aerial-photo interpreters, sports analysis, 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 & 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 & Simon interpret this as showing 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 which 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., Starkes, Allard, Lindley & O'Reilley, 1994; Abernethy, Neal & Koning, 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 & Umilta, 1992; Greenfield, deWinstanley, Kilpatrick & Kaye, 1994; Nougier, Ripoll & Stein, 1989) visual expertise in sports, like the expertise found in the Chase & Simon studies of chess skill, appears to be based on non-visual 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 & 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 importantly, 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 became 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 Section 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 & 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 most likely was 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, Edelman & Poggio, 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, Nealey & Maunsell, 1992). In this task 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 only improvement 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, Goldstone & Thibaut (in press) 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., Section 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 & 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 to 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 Section 6.3) and for the encoding of combinations of features (this is the "attention as glue" hypothesis of Treisman, 1988; see also Ashby, Prinzmetal, Ivry & Maddox, 1996)

In addition to the single spatial locus of enhanced processing that has generally been referred to as focal attention (discussed in Section 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 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 & 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 figure 4), since 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 such 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 "complied 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 an 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 since 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 retinal 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 effect 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 non-retinal 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 outputs 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-one means that the visual system induces a partition of the visual world into equivalence classes (many-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, e.g., 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 depend 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 might be thought of as generating a set of one or more shape-descriptions which in many cases might be sufficient (perhaps in concert with other contextual information) to identify objects stored in memory — presumably along with other information about these objects. 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

[14] The question of what constitutes similarity of appearance is being completely begged in this discussion. 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 since, 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, Mervis, Gray, Johnson, & Boyes-Braem, 1976). It also seems reasonable that the shape-classes provided by vision are ones whose names can be learned by ostension — i.e., 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.

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 which 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 non-visual 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 don't 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, He & Shimojo (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 in a number of separate subprocesses that compute color, luminance, motion, form and 3-D depth and that these subprocesses are restricted in their intercommunication (Livingston & Hubel, 1987; Cavanagh, 1988). In other words, the visual process is highly complex 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

[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 tradeoff 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 (c.f., 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 note 7).

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 it 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 enable 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).

[16] Needless to say, not everyone agrees on the precise status of subjective experience in visual science. It 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, Thompson & Noë (in press) and the associated commentaries.

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 & Lishman (1975) showed that posture can be controlled by the oscillations of a specially designed room whose walls were suspended inside a real room and could 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, Pelisson & Prablanc (1986) also showed a dissociation between information that is noticed by a subjects and information to which the motor system responds. In reaching for a target, subjects first make a saccadic eye movement towards 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 & Mack (1981) and subsequently Bridgeman (1992) showed that the judgment and motor system can even be given conflicting visual information. The Wong & 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 & 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, Bracewell & Andersen (1991) and in the case of reaching and grasping by Goodale, Jakobson & Keillor (1994). Aglioti, DeSouza & Goodale (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 which, 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 & Goodale (1995).

Milner & 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 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 towards 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 can not 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, Warrington, Sanders & Marshall, 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 towards points of light that they insisted he could not see, and at least in the case of Weiskrantz' 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, DN 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.

While it is still too early to conclude, as Milner & Goodale (1995), as well as 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 review 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 note 5), 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 while 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 paper 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) and 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 3D 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 3D 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.

References

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.

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.

Aglioti, S., DeSouza, J. F. X., & Goodale, M. A. (1995). Size-contrast illusions deceive the eye but not the hand. *Current Biology*, **5,** 679-685.

Ahissar, M., & Shaul, H. (1995). How early is early vision? Evidence from perceptual learning. *In* T. V. Papathomas, C. Chubb, A. Gorea, & E. Kowler (Eds.), *Early vision and beyond* (pp. 199-206). Cambridge, MA: MIT Press (A Bradford Book).

Ashby, F. G., Prinzmetal, W., Ivry, R., & Maddox, W. T. (1996). A formal theory of feature binding in object perception. *Psychological Review*, **103,** 165-192.

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,** 1141-1147.

Bennett, B., Hoffman, D., & Prakash, K. (1989). *Observer Mechanics: A formal theory of perception.* New York: Academic Press.

Biederman, I. (1987). Recognition-by-components: A theory of human image interpretation. *Psychological Review*, **94,** 115-148.

Biederman, I., Mezzanotte, R. J., & Rabinowitz, J. C. (1982). Scene perception: Detecting and judging objects undergoing relational violations. *Cognitive Psychology*, **14,** 143-177.

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,** 640-645.

Blake, R. (1995). Psychoanatomical strategies for studying human visual perception. *In* T. V. Papathomas, C. Chubb, A. Gorea, & E. Kowler (Eds.), *Early vision and beyond* (pp. 17-25). Cambridge, MA: MIT Press (A Bradford Book).

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,** 757-776.

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

Bridgeman, B. (1992). Conscious vs. unconcious processes: The case of vision. *Theory Psychology*, **2,** 73-88.

Bridgeman, B. (1995). Dissociation between visual processing modes. *In* M. A. Arbib (Ed.), *The Handbook of Brain Theory and Neural Networks*. Cambridge, MA: MIT Press.

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

Brown, C. M. (1984). Computer vision and natural constraints. *Science*, **224,** 1299-1305.

Bruce, V. (1991). Face recognition. Hillsdale, NJ: Lawrence Erlbaum Associates.

Bruner, J., & Postman, L. (1949). On the perception of incongruity: a paradigm. *Journal of Personality*, **18,** 206-223.

Bruner, J. S. (1957). On perceptual readiness. *Psychological Review*, **64,** 123-152.

Bruner, J. S., & Minturn, A. L. (1955). Perceptual identification and perceptual organization. *Journal of General Psychology*, **53,** 21-28.

Burkell, J., & Pylyshyn, Z. W. (1997). Searching through subsets: A test of the visual indexing hypothesis. *Spatial Vision*, **11,** 225-258.

Castiello, U., & Umiltá, C. (1992). Orienting attention in volleyball players. *International Journal of Sport Psychology*, **23,** 301-310.

Cavanagh, P. (1988). Pathways in early vision. *In* Z. W. Pylyshyn (Ed.), *Computational Processes in Human Vision: An interdisciplinary perspective* (pp. 239-262). Norwood, NJ: Ablex Publishing.

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

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, **5,** 55-81.

Cheng, K. (1986). A purely geometric module in the rat's spatial representation. *Cognition*, **23,** 149-178.

Churchland, P. M. (1988). Perceptual plasticity and theoretical neutrality: A reply to Jerry Fodor. *Philosophy of Science*, **55,** 167-187.

Clowes, M. B. (1971). On Seeing Things. *Artificial Intelligence*, **2,** 79-116.

Connine, C. M., & Clifton, C. (1987). Interactive use of lexical information in speech perception. *Journal of Experimental Psychology: Human Perception And Performance*, **13,** 291-299.

Dai, H., Scharf, B., & Buss, S. (1991). Effective attenuation of signals in noise under focused attention. *Journal of the Acoustical Society of America*, **89,** 2837-2842.

Dawson, M., & Pylyshyn, Z. W. (1988). Natural constraints in apparent motion. *In* Z. W. Pylyshyn (Ed.), *Computational Processes in Human Vision: An interdisciplinary perspective* (pp. 99-120). Norwood, NJ: Ablex Publishing.

Desimone, R. (1996). Neural mechanisms for visual memory and their role in attention. *Proceedings of the National Acedemy of Science, USA*, **93,** 13494-13497.

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

Egeth, H. E., Virzi, R. A., & Garbart, H. (1984). Searching for conjunctively defined tagets. *Journal of Experimental Psychology*, **10,** 32-39.

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,** 143-165.

Epstein, W. (1982). Percept-percept couplings. *Perception*, **11,** 75-83.

Fahle, M., Edelman, S., & Poggio, T. (1995). Fast perceptual learning in hyperacuity. *Vision Research*, **35,** 3003-3013.

Farah, M. J. (1989). Semantic and perceptual priming: How similar are the underlying mechanisms? *Journal of Experimental Psychology: Human Perception and Performance*, **15,** 188-194.

Farah, M. J. (1990). *Visual Agnosia: Disorders of Object Recognition and What They Tell us About Normal Vision*. Cambridge, MA: MIT Press.

Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*, **1,** 1-47.

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.

Feyerabend, P. (1962). Explanation, reduction and empiricism. *In* H. Feigl, & G. Maxwell (Eds.), *Minnesota Studies in Philosophy of Science (Volume 3)*. Minneapolis, MN: University of Minnesota Press.

Fodor, J. A. (1983). *The Modularity of Mind: An Essay on Faculty Psychology*. Cambridge, Mass.: MIT Press, a Bradford Book.

Fodor, J. A. (1998). *Concepts: Where cognitive science went wrong*. Oxford: Oxford University Press.

Fodor, J. A., & Pylyshyn, Z. W. (1981). How Direct is Visual Perception? Some Reflections on Gibson's 'Ecological Approach'. *Cognition*, **9,** 139-196.

Freuder, E. C. (1986). Knowledge-Mediated Perception. *In* H. C. Nusbaum, and Schwab, E. C. (Ed.), *Pattern Recogniton by Humans and Machines: Visual Perception*. Orlando: Academic Press Inc.

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,** 531-551.

Frisby, J. P., & Clatworthy, J. L. (1975). Learning to see complex random-dot stereograms. *Perception*, **4,** 173-178.

Gale, A. G., & Findlay, J. M. (1983). Eye movement patterns in viewing ambiguous figures. *In* R. Groner, C. Menz, D. F. Fisher, & R. A. Monty (Eds.), *Eye Movements and Psychological Functions: International Views* (pp. 145-168). Hillsdale, NJ: Lawrence Ernlaum Associates.

Gallistel, C. R. (1990). *The Organization of Learning*. Cambridge, MA: MIT Press (A Bradford Book).

Gibson, E. J. (1991). *An Odyssey in learning and perception*. Cambridge, MA: MIT Press.

Gibson, J. J. (1979). *An Ecological Approach to Visual Perception*. Boston: Houghton Mifflin.

Gilchrist, A. (1977). Perceived lightness depends on perceived spatial arrangement. *Science*, **195,** 185-187.

Girgus, J. J., Rock, I., & Egatz, R. (1977). The effect of knowledge of reversibility on the reversibility of ambiguous figures. *Perception and Psychophysics*, **22,** 550-556.

Gogel, W. C. (1973). The organization of perceived space. *Psychologische Forschung*, **37,** 195-221.

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

Goldstone, R. L. (1995). The effect of categorization on color perception. *Psychological Science*, **6,** 298-304.

Goodale, M. A. (1983). Vision as a sensorimotor system. *In* T. E. Robinson (Ed.), *Behavioral approaches to brain research* (pp. 41-61). New York: Oxford University Press.

Goodale, M. A. (1988). Modularity in visuomotor control: From input to output. *In* Z. W. Pylyshyn (Ed.), *Computational Processes in Human Vision: An interdisciplinary perspective* (pp. 262-285). Norwood, NJ: Ablex Publishing.

Goodale, M. A., Jacobson, J. S., & Keillor, J. M. (1994). Differences in the visual control of pantomimed and natural grasping movements. *Neuropsychologia*, **32,** 1159-1178.

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,** 748-750.

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

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,** 105-123.

Grimson, W. E. L. (1990). The combinatorics of object recognition in cluttered environment using constrained search. *Artificial Intelligence*, **44,** 121-166.

Haber, R. N. (1966). Nature of the effect of set on perception. *Psychological Review*, **73,** 335-351.

Haenny, P., & Schiller, P. (1988). State dependent activity in money visual cortex I. Single cell activity in V1 andV4 on visual tasks. *Experimental Brain Research*, **69,** 225-244.

Hanson, N. R. (1958). *Patterns of Discovery*. Cambridge: Cambridge University Press.

Hebb, D. O. (1968). Concerning Imagery. *Psychological Review*, **75,** 466-477.

Hernandez-Péon, R., Scherrer, R. H., & Jouvet, M. (1956). Modification of electrical activity in the cochlear nucleus during "attention" in unanesthetized cats. *cience*, **123,** 331-332.

Hollingworth, A., & Henderson, J. M. (in press). Does consistent scene context facilitate object perception? *Journal of Experimental Psychology: General*.

Holmes, G. (1918). Disturbances in visual orientation. *British Journal of Ophthalmology*, **2,** 449-506.

Howard, I. P. (1982). *Human Visual Orientation*. New York, NY: John Wiley & Sons.

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction, and functional architecture in the cat's visual cortex. *Journal of Physiology*, **160,** 106-154.

Humphreys, G. W., & Riddoch, M. J. (1987). *To see but not to see : a case study of visual agnosia*. Hillsdale, NJ: Lawrence Erlbaum.

Irwin, D. E. (1993). Perceiving an integrated visual world. *In* D. E. Meyer, & K. S. (Eds.), *Attention and Performance XIV*. Cambridge, MA: MIT Press.

Ittelson, W. H., & Ames, A. J. (1968). *The Ames Demonstrations in Perception*. New York: Hafner Publishng Co.

Johansson, G. (1950). *Configurations in Event Perception*. Boston: Houghton Mifflin.

Julesz, B. (1971). *Foundations of Cyclopean Perception*. Chicago: Univ. of Chicago Press.

Julesz, B. (1990). Early vision is bottom up, except for focal attention. *Cold Spring Harbor Symposium on Quantitative Biology*, **55,** 973-978.

Julesz, B., & Papathomas, T. V. (1984). On spatial-frequency channels and attention. *Perception and Psychophysics*, **36,** 398-399.

Kahneman, D., & Treisman, A. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology*, **24,** 175-219.

Kanizsa, G. (1969). Perception, past experience and the impossible experiment. *Acta Psychologica*, **31,** 66-96.

Kanizsa, G. (1985). Seeing and Thinking. *Acta Psychologica*, **59,** 23-33.

Kanizsa, G., & Gerbino, W. (1982). Amodal completion: Seeing or thinking? *In* B. Beck (Ed.), *Organization and Representation in Perception* (pp. 167-190). Hillsdale, NJ: Erlbaum.

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,** 4302-4311.

Karni, A., & Sagi, D. (1995). A memory system in the adult visual cortex. *In* B. Julesz, & I. Kovacs (Eds.), *Maturational Windows and adult cortical plasticity* (pp. 95-109). Reading, MA: Addison-Wesley.

Kawabata, N. (1986). Attention and depth perception. *Perception*, **15,** 563-572.

Kelly, M. D. (1971). Edge detection by computer using planning. *In* B. Meltzer, & D. Michie (Eds.), *Machine Intelligence 6*. Edinburgh: University of Edinburgh Press.

Kosslyn, S. M. (1994). *Image and Brain*. Cambridge. MA: MIT Press.

Kuhn, T. (1972). *The Stucture of Scientific Revolutions*. Chicago: Univ. of Chicago Press.

Kutas, M., McCarthy, G., & Donchin, E. (1977). Augmenting mental chronometry: The P300 as a measure of stimulus evaluation time. *Science*, **197,** 792-795.

Lee, D. N., & Lishman, J. R. (1975). Visual proprioceptive control of stance. *Journal of Human Movement Studies*, **1,** 87-95.

Leslie, A. M. (1988). The necessity of illusion: Perception and thought in infancy. *In* L. Weiskrantz (Ed.), *Thought Without Language*. Oxford: Oxford Science Publications.

Lindsay, P. H., & Norman, D. A. (1977). *Human Information Processing: An Introduction to Psychology*.

New York: Academic Press.

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,** 3416-3468.

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,** 417-434.

Lowe, D. (1987). Three-dimensional object recognition from single two-dimensional images. *Artificial Intelligence*, **31**.

Lupker, S. J., & Massaro, D. W. (1979). Selective perception without confounding contributions of decision and memory. *Perception and Psychophysics*, **25,** 60-69.

Lynch, J. C. (1980). The functional organization of posterior parietal cortex. *BBS*, **3,** 485-534.

Magnuson, S. (1970). Reversibility of perspective in normal and stabilized viewing. *Scandinavian Journal of Psychology*, **11,** 153-156.

Maloney, L. T., & Wandell, B. A. (1990). Color constancy: A method for recovering surface spectral reflectance. *In* S. Ullman, & W. Richards (Eds.), *Image Understanding 1989*. Norwood, NJ: Ablex Publishing.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco: W.H. Freeman.

Marr, D., & Poggio, T. (1979). A theory of human stereo vision. *Proceedings of the Royal Society of London*, **204**.

Massaro, D. W. (1988). *Experimental Psychology: An Information Processing Approach*. New York: Harcourt Brace Javanovich.

McCarthy, G., & Donchin, E. (1981). A metric for thought: A comparison of P300 latency and reaction time. *Science*, **211,** 77-80.

McClelland, J. L. (1991). Stochastic interactive processes and the effect of context on perception. *Cognitive Psychology*, **23,** 1-44.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, **18,** 1-86.

McLeod, P., Driver, J., & Crisp, J. (1988). Visual search for a conjunction of movement and form is parallel. *Nature*, **332,** 154-155.

McLeod, P., Driver, J., Dienes, Z., & Crisp, J. (1991). Filtering by movement in visual search. *Journal of Experimental Psychology: Human Perception & Performance*, **17,** 55-64.

Merigan, W. H., Nealey, T. A., & Maunsell, J. H. R. (1992). Qualititatively different effects of lesions of cortical areas V1 and V2 in Macaques. *Perception*, **21,** 55-56.

Miller, G. A. (1962). Decision units in the perception of speech. *Institute of Radio Engineers: Transactions on Information Therory*, **8,** 81-83.

Miller, G. A., Bruner, J. S., & Postman, L. (1954). Familiarity of letter sequences and tachistoscopic identification. *Journal of General Psychology*, **50,** 129-139.

Milner, A. D., & Goodale, M. A. (1995). *The Visual Brain in Action*. New York: Oxford University Press.

Moran, J., & Desimone, R. (1985). Selective attention gates visual processing in the extrastriate cortex. *Science*, **229,** 782-784.

Morton, J. (1969). Interaction of information in word recognition. *Psychological Review*, **76,** 165-178.

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,** 2239-2255.

Muller, H. J., & Findlay, J. M. (1987). Sensitivity and criterion effects effects in the spatial cuing of visual attention. *Perception and Psychophysics*, **42,** 383-399.

Nakayama, K., He, Z. J., & Shimojo, S. (1995). Visual surface representation: A critical link between lower-level and higher-level vision. *In* S. M. Kosslyn, & D. N. Osherson (Eds.), *Visual Cognition* (pp. 1-70). Cambridge, MA: MIT Press.

Nakayama, K., & Silverman, G. H. (1986). Serial and parallel processing of visual feature conjunctions. *Nature*, **320,** 264-265.

Newell, A. (1980). HARPY, Production Systems, and Human Cognition. *In* R. A. Cole (Ed.), *Perception and production of fluent speech*. Hillsdale, NJ: Erlbaum.

Newell, A. (1990). *Unified Theories of Cognition*. Cambridge, MA: Harvard University Press.

Norris, D. (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,** 935-939.

Nougier, V., Ripoll, H., & Stein, J.-F. (1989). Orienting of attention with highly skilled athletes. *International Journal of Sport Psychology*, **20,** 205-223.

O'Regan, J. K. (1992). Solving the "real" mysteries of visual perception: The world as an outside memory. *Canadian Journal of Psychology*, **46,** 461-488.

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 conrtex. *In* A. Blake, & T. Troscianko (Eds.), *AI and the Eye* (pp. 181-200). Chichester, England: John Wiley & Sons.

Perrett, D. I., Mistlin, A. J., & Chitty, A. J. (1987). Visual neurones responsive to faces. *Trends in Neuroscience*, **10,** 358-364.

Pessoa, L., Thompson, E., & Noe, A. (in press). Finding out about filling in: A giude to perceptual completion for visual science and the philosophy of perception. *BBS*.

Peterson, M. A., & Gibson, B. S. (1991). Directing spatial attention within an object: Altering the functional equivalence of shape description. *Journal of Experimental Psychology: Human Perception and Performance*, **17,** 170-182.

Poggio, T., Torre, V., & Koch, C. (1990). Computational vision and regularization theory. *In* S. Ullman, & W. Richards (Eds.), *Image Understanding 1989* (pp. 1-18). Norwood, N.J.: Ablex Publishing.

Potter, M. C. (1975). Meaning in visual search. *Science*, **187,** 965-966.

Proteau, L. (1992). On the specificity of learning and the role of visual information for movement control. *In* L. Proteau, & D. Elliot (Eds.), *Vision and motor control. Advances in Psychology No. 85* (pp. 67-103). Amsterdam, Netherlands: North-Holland.

Pylyshyn, Z. W. (1973). What the Mind's Eye Tells the Mind's Brain: A Critique of Mental Imagery. *Psychological Bulletin*, **80,** 1-24.

Pylyshyn, Z. W. (1978). Imagery and Artificial Intelligence. *In* C. W. Savage (Ed.), *Perception and Cognition: Issues in the Foundations of Psychology*. Minneapolis: Univ. of Minnesota Press.

Pylyshyn, Z. W. (1981). The imagery debate: Analogue media versus tacit knowledge. *Psychological Review*, **88,** 16-45.

Pylyshyn, Z. W. (1984). *Computation and cognition: Toward a foundation for cognitive science*. Cambridge, MA: MIT Press.

Pylyshyn, Z. W. (1989). The role of location indexes in spatial perception: A sketch of the FINST spatial-index model. *Cognition*, **32,** 65-97.

Pylyshyn, Z. W. (1994). Some primitive mechanisms of spatial attention. *Cognition*, **50,** 363-384.

Pylyshyn, Z. W., & Storm, R. W. (1988). Tracking of multiple independent targets: evidence for a parallel tracking mechanism. *Spatial Vision*, **3,** 1-19.

Ramachandran, V. S. (1976). Learning-like phenomena in stereopsis. *Nature*, **262,** 382-384.

Ramachandran, V. S. (1990). Visual perception in people and machines. *In* A. Blake, & T. Troscianko (Eds.), *AI and the Eye*. New York: John WIley & Sons.

Ramachandran, V. S., & Braddick, O. (1973). Orientation specific learning in stereopsis. *Perception*, **2,** 371-376.

Reddy, D. R. (1975). Speech Recognition. New York: Academic Press.

Reynolds, R. (1981). Perception of an illusory contour as a function of processing time. *Perception*, **10,** 107-115.

Reynolds, R. I. (1985). The role of object-hypotheses in the organization of fragmented figures. *Perception*, **14,** 49-52.

Rhodes, G., Parkin, A. J., & Tremewan, T. (1993). Semantic priming and sensitivity in lexical decision. *Journal of Experimental Psychology: Human Perception and Performance*, **19,** 154-165.

Rhodes, G., & Tremewan, T. (1993). The Simon then Garfunkel effect: Semantic priming, sensitivity, and the modularity of face recognition. *Cognitive Psychology*, **25,** 147-187.

Richards, W. (1988). Natural Computation. Cambridge, MA: MIT Press (A Bradford Book).

Riseman, E. M., & Hanson, A. R. (1987). A methodology for the development of general knowledge-based vision systems. *In* M. A. Arbib, & A. R. Hanson (Eds.), *Vision, Brain, and Cooperative Computation*. Cambridge, MA: MIT Press (A Bradford Book).

Roberts, L. G. (1965). Machine perception of three-dimensional solids. *In* J. P. Tippett (Ed.), *Optical and electro-optical information processing*. Cambridge, MA: MIT Press.

Rock, I. (1983). *The Logic of Perception*. Cambridge, Mass.: MIT Press, a Bradford Book.

Rock, I. (1997). Indirect Perception. Cambridge, MA: MIT Press.

Rock, I., & Anson, R. (1979). Illusory contours as the solution to a problem. *Perception*, **8,** 655-681.

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.

Rosenblatt, F. (1959). Two theorems of statistical separability in the perceptron. *In* N. P. Laboratory (Ed.), *Symposium on Mechanization of Thought Processes*. London: HM Stationery Office.

Rosenfeld, A., Hummel, R. A., & Zucker, S. W. (1976). Scene labeling by relaxation operators. *IEEE Transactions on Systems, Man, and Cybernetics*, **SMC-6,** 420-433.

Rumelhart, D. E. (1977). *Human Information Processing*. New York: John Wiley & Sons.

Samuel, A. G. (1981). Phonemic restoration: Insights from a new methodology. *Journal of Experimental Psychology: General*, **110,** 474-494.

Samuel, A. G. (1996). Does lexical information influence the perceptual restoration of phonemes? *Journal of Experimental Psychology: General*, **125,** 28-51.

Saye, A., & Frisby, J. P. (1975). The role of monocularly conspicuous features in facilitating stereopsis from random-dot stereograms. *Perception*, **4,** 159-171.

Schyns, Goldstone, & Thibaut (in press). The development of features in object concepts. *BBS*.

Sekuler, A. B., & Palmer, S. E. (1992). Visual completion of partly occluded objects: A microgenetic analysis. *Journal of Experimental Psychology: General*, **121,** 95-111.

Sekuler, R., & Blake, R. (1994). *Perception*. New York: McGraw-Hill.

Selfridge, O. (1959). Pandemonium: A paradigm for learning, *Symposium on Mechanization of Thought Processes: National Physical Laboratory Symposium*. London: HM Stationery Office.

Shaw, M. L. (1984). Division of attention among spatial locations: a fundamental difference between detection of letters and detection of luminance increments. *In* H. Bouma, & D. G. Bouwhuis (Eds.), *Attention and Performance, X*. Hillsdale, NJ: Erlbaum.

Shirai, Y. (1975). Analyzing intensity arrays using knowledge about scenes. *In* P. H. Winston (Ed.), *Psychology of Computer Vision*. Cambridge, MA: MIT Press.

Shulman, G. L., & Wilson, J. (1987). Spatial frequency and selective attention to local and global information. *Perception*, **16,** 89-101.

Sigman, E., & Rock, I. (1974). Stroboscopic movement based on perceptual intelligence. *Perception*, **3,** 9-28.

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,** 479-482.

Snodgrass, J. G., & Feenan, K. (1990). Priming effects in picture fragment completion: Support for the perceptual closure hypothesis. *Journal of Experimental Psychology: General*, **119,** 276-296.

Soloman, R. L., & Postman, L. (1951). Frequency of usage as a determinant of recognition thresholds for words. *Journal of Experimental Psychology*, **43,** 195-201.

Stark, L., & Ellis, S. R. (1981). Scanpaths revisited: Cognitive models direct active looking. *In* D. Fisher, R. Monty, & J. Senders (Eds.), *Eye Movements: Cognition and Visual perception*. Hillsdale, NJ: Lawrence Erlbaum Associates.

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,** 249-265.

Sternberg, S. (1969). The Discovery of Processing Stages: Extensions of Donders' Method. *Acta Psychologica*, **30,** 276-315.

Sternberg, S. (1998). Discovering mental processing stages: The method of additive factors. *In* D. Scarborough, & S. Sternberg (Eds.), *Methods, Models, and Conceptual issues* (pp. 703-864). Cambridge. MA: MIT Press (A Bradford Book).

Street, R. F. (1931). *A Gestalt Completion Test: A study of a cross section of intellect*. New York: Bureau of Publications, Teachers College, Columbia University.

Swets, J. A. (1998). Separating discrimination and decision in detection, recognition, and matters of life and death. *In* D. Scarborough, & S. Sternberg (Eds.), *Methods, Models, and Conceptual issues* (pp. 635-702). Cambridge, MA: MIT Press (A Bradford Book).

Tanner, W. P., & Swets, J. A. (1954). A decision-making theory of human detection. *Psychological Review*, **61,** 401-409.

Treisman, A. (1988). Features and objects: The fourteenth Bartlett memorial lecture. *The Quarterly Journal of Experimental Psychology*, **40A,** 201-237.

Ullman, S. (1976). On Visual Detection of Light Sources. *Biological Cybernetics*, **21,** 205-212.

Ullman, S. (1979). *The interpretation of visual motion*. Cambridge, MA: MIT Press.

Ullman, S., & Richards, W. (1990). Image Understanding. Norwood, NJ: Ablex Publishers.

Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. *In* J. Ingle, M. A. Goodale, & R. J. W. Mansfield (Eds.), *Analysis of visual behavior* (pp. 549-586). Cambridge, MA: MIT Press.

Uttley (1959). Conditional Probability Computing in the nervous system, *Mechanization of Thought Processes*. London: HM Stationery Office.

van Essen, D. C., & Anderson, C. H. (1990). Information processing strategies and pathways in the primate retina and visual cortex. *In* S. F. Zornetzer, J. L. Davis, & C. Lau (Eds.), *Introduction to Neural and Electronic Networks*. New York: Academic Press.

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.

Wallach, H. (1949). Some considerations concerning the relation between perception and cognition. *Journal of Personality*, **18,** 6-13.

Wallach, H., & O'Connell, D. N. (1953). The kinetic depth effect. *Journal of Experimental Psychology*, **45,** 205-217.

Weiskrantz, L. (1995). Blindsight: Not an island unto itself. *Current Directions in Psychological Science*, **4,** 146-151.

Weiskrantz, L., Warrington, E., Sanders, M. D., & Marshall, J. (1974). Visual capacity in the hemianopic field following restricted occipital ablation. *Brain*, **97,** 709-729.

Wickens, C. D., Kramer, A. F., & Donchin, E. (1984). The event-related potrntial as an index of processing demands of a complex target acquisition task. *Annals of the New York Academy of Sciences,* **425,** 295-299.

Wilson, J. A., & Robinson, J. O. (1986). The impossibly-twisted Pulfrich pendulum. *Perception*, **15,** 503-504.

Winston, P. H. (1974). New progress in artificial intelligence. Cambridge, MA: MIT Artificial Intelligence Laboratory.

Wong, E., & Mack, A. (1981). Saccadic programming and perceived location. *Acta Psychologia*, **48,** 123-131.

Woods, W. A. (1978). Theory Formation and Control in a Speech Understanding System with Extrapolations towards Vision, *Computer Vision Systems: Papers from the workshop on computer vision systems*. New York: Academic Press.

Yuille, A. L., & Ullman, S. (1990). Computational theories of low-level vision. *In* D. N. Osherson, S. M. Kosslyn, & J. M. Hollerbach (Eds.), *Visual cognition and action* (pp. 5-39). Cambridge, MA: MIT Press.

Zucker, S. W., Rosenfeld, A., & David, L. S. (1975). General purpose models: Expectations about the unexpected, *Fourth International Joint Conference on Artificial Intelligence*. Tbilisi, Georgia, USSR.