Chapter 2
THE INDEPENDENCE OF VISION AND COGNITION

2.1 Is Vision distinct from reasoning?

2.1.1 What determines what we see? Do we see what we expect to see?

If the first major seduction of our phenomenal experience is the belief that vision constructs an inner world or inner display, then the next one is that “seeing is believing”, or that there is little or no distinction between the visual system and the general reasoning system, other than that the former gets some of its initial information from the eyes. It is widely, if not universally, accepted that what we see is heavily conditioned by our beliefs and our expectations. The view that perception and cognition are continuous is particularly believable because it comports well with everyday experience as well as with the Zeitgeist – the spirit of the times – that celebrates the plasticity of the mind. The beginning of the second half of the 20th century was characterized by a belief in biologically limitless human potential. An advertisement on the American Public Television System declares, “If it can be imagined, it can be done: This is America!” In the debate between nature and nurture it was nature that had its day in the 1950s with the dominance of philosophical empiricism. That brought with it not only the view of the mind as a blank tabula rasa, upon which experience writes the entire adult intellectual capacity, but also the view that the newborn perceptual system provides the infant only with what William James called “A blooming buzzing confusion”. This picture of human nature has been thoroughly dismantled as new evidence shows ever more capacities in the newborn mind. One part of this prevalent ideology is a perceptual relativity: the view that how we perceive the world is conditioned by our beliefs, our culture, our moods and so on. The average person takes it for granted that how we see the world is radically influenced by our mental state; by our beliefs, our expectations, and our needs, hopes and fears, and above all, by our language and culture. And there are plenty of reasons to take this view. One of the more dramatic illustrations of this is magic, where the magician manipulates what we see by setting up certain false expectations. We are also told that when people are thirsty in a desert they often see oases. And who has not had the experience of being afraid and then being “spooked” by harmless shadows which we mistake for signs of something awful. The popularity of what is known as the Sapir-Whorf hypothesis of linguistic relativity among the literate public (which gave us the apocryphal story that Eskimos have 17 different words for snow because their perceptual discrimination of types of snow is more refined – i.e., that they have 17 different ways of seeing snow) also supports this general view, as does the widespread belief in the cultural effect on our ways of seeing (e.g., the books by Carlos Castaneda). The remarkable placebo effect of drugs and of authoritative suggestions by people wearing white coats also bears witness to the startling malleability of perception, not to mention the effects of suggestions given under hypnosis.

In this chapter I will examine some of the reasons why this view has been generally held, not just by the public, but also by psychologists, and I will also examine some evidence against this view of vision. To anticipate the conclusion, what I will argue is that visual perception, in the everyday sense of the term, does indeed merge seamlessly with reasoning and other aspects of cognition. But the everyday sense of the term is too broad to be of scientific value precisely because it fails to make distinctions that “cut nature at her joints”. I will argue that within the broad category of what we call “vision” is an information processing system, which some have called “early vision”, that is highly complex, but which functions independently of what we believe. This system picks out or individuates objects in a scene and computes the spatial layout of visible
surfaces and the 3D shape of the objects in the scene.\textsuperscript{10} It thus covers a lot of what one means by visual perception. What it does not do is identify the things we are looking at, in the sense of relating them to things we have seen before, to the contents of our memory. And it does not make judgments about how things really are. In other words this aspect of vision is not the sole determiner of our perceptual beliefs – our beliefs about what we are seeing. In this sense then, seeing\textsuperscript{11} is not the same as believing; far from it. Believing depends on all your wits and intelligence and knowledge of how the world works, whereas early vision does not. It depends on how the visual system was wired up by evolution as well as by biological, chemical and physical principles, and on the incoming patterns of light, but on little else. In would not be a great exaggeration to say that “early vision” – the part of visual processing that is prohibited from accessing general knowledge – computes just about everything that might be called a “visual appearance” of the world except the identities and names of the objects. This view, which I have referred to as the “independence thesis” or the “modularity of vision thesis,” flies in the face of a great deal of received wisdom, both in and out of visual science, and requires some elaboration. In particular, it requires drawing some conceptual distinctions and explaining a great deal of apparently contradictory evidence. By the time we are through with this topic we will have distinguished between seeing and believing, between “seeing” and “seeing as,” and between the processes carried out by visual system and pre-visual process of deciding where to focus attention or the post-visual process of deciding what the visual system reveals about the scene.

But first, let us look more closely at the basis for the belief in the relativity and plasticity of vision. The reasons for believing that vision and cognition are very closely linked go deeper than just our everyday experience. The view is backed by an enormous amount of experimental evidence from psychology and from social sciences more broadly (including cross-cultural observations). It is also supported by work in artificial intelligence that attempts to build computer systems that can recognize objects and scenes. Below I present some of this experimental evidence, which persuaded many scientists that vision is continuous with cognition. I 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.

2.2 The case for the continuity of vision and cognition

2.2.1 The “New Look” in the psychology of perception

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 perceivers’ beliefs about the world and about the

\textsuperscript{10} Notice that it is critical for our thesis that the notion of early vision include, among other things, such functions as the individuation of objects (which will be the topic of Chapter 5), as well as the computation of what are referred to as “surface layouts” – the shape of the visible surfaces in our field of view. If it were not for the fact that early vision includes such complex features of the visible world, the independence theses would be trivial since everyone believes that something is detected by the visual system without regard for beliefs and expectations, otherwise we would only see whatever we wished! The more conventional view (e.g., that was held by empiricists and perhaps by “New Look” theorists discussed in section 2.2.1) is that the only thing that is isolated from cognition is the operation of the sensors. We will return to the question of the nature of the output of early vision in the next chapter.

\textsuperscript{11} Having made the distinction between the ordinary sense of vision and what I called “early vision” I will often fall into the habit of referring to the latter as “vision” and of this process as resulting in “seeing”. I do this because the narrow technical sense of “vision” is precisely that part of visual perception that is unique to, and therefore constitutive of, visual perception. The broader sense is simply a confounding of several distinct processes. The policy of referring to the part of a process that forms the unique core of that process by the more general term is precisely what Chomsky does when he uses the term “language” to refer to that part of our entire linguistic ability that is unique to language, even though understanding and generating linguistic signs clearly involves all of our cognitive faculties (Chomsky, 1976).
particular scene before them: 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 formed the basis of what became known as the “New Look in Perception,” a movement that flourished in the 1960s and 1970s and continues to be influential even today in many areas of social science. According to this view, we perceive in conceptual 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 essential distinction between the processes of perception and thought.12

This view opposed the earlier position that had been adopted by psychologists from as wide a perspective as structuralists (like Titchener, 1915), Gestaltists (like Wallach, 1976) and even the early work of James J. Gibson (Gibson, 1966). Adherents of all these schools had accepted some sort of distinction between a pure stimulus event and a stimulus event that had interacted with past experience, or as it was sometimes put, between the “visual field” and the “visual world”, or between perception and conception (or, as the latter was sometimes called, apperception – the process of assimilating perception into one’s cognitive world). It would not be a misstatement to say that these thinkers accepted that there is a difference between appearances and beliefs, even though they would not have put it in those terms.

The New Look fit well with the Zeitgeist, which saw the organism as highly malleable by its environment, both in the short term and in the longer term (in the latter case this malleability was called ‘learning’). It was also a reaction against sense-data views of perception, views that had attained some currency with the structuralist and introspectionist schools of perception. Sense-data theories assumed that percepts were constructed out of more basic elements of experience called “sensations”. Sensations were different in kind from percepts; they were not perceptual categories but the raw material of the senses. This version of perceptual atomism lost much of its popularity with the fall of the introspective method, as well as with the general disenchantment with perceptual atomism promoted by Gestalt psychology. It seemed that basic perceptual atoms, especially ones that could be consciously sensed, were not to be had; perception was, as the Gestalt psychologists would say, always greater than the sum of its more elementary parts. In this spirit, the field was ripe for a holistic all-encompassing theory of perception that integrated it into the general arena of induction and reasoning.

There were literally thousands of experiments performed from the 1950s through the present time showing that the perception of almost any pattern, from the perception of sentences in noise to the recognition of familiar stimuli at short exposures, could be influenced by observers’ 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 cited 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 concluded that cognitive context affects such simple psychophysical tasks as magnitude judgments. In the case of more complex patterns there is even more evidence for the effects of what Bruner calls “readiness” on perception. The recognition

12 (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.
threshold for words decreases as the words become more familiar (Soloman & Postman, 1952). The exposure time required to report a string of letters shown in the flash-exposure instrument known as the tachistoscope, which was becoming extremely popular in psychology laboratories at that time, 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); the higher the order of approximation, the shorter the exposure required for recognition. The level of background noise at which listeners can still recognize a word is higher if that word is part of a sentence (where it could be predicted more easily) or even if it occurs in a list of words whose order statistically approximates English (Miller, 1962). More generally, words in sentences can be recognized more easily and in more adverse conditions that can words alone or in random lists.

Similar results were 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 higher than the time required to recognize a normal card (Bruner & Postman, 1949). Also, as in the letter and word recognition cases, the perceptual thresholds reflect the likelihood of occurrence of the stimuli in a particular context, and even their significance to the observer (the latter being illustrated by studies of so-called “perceptual defense”, in which pictures previously associated with shock, show elevated recognition thresholds). The reconstruction of partially-occluded figures was also taken as showing that vision makes use of knowledge in order to restore familiar shapes, as in the “New York” example in Chapter 1, Figure 1-4.

The results of these experiments were explained in terms of the accessibility of perceptual categories and the hypothesize-and-test nature of perception (where the “hypotheses” can come from any source, including immediate context, memory and general knowledge). There were also experiments that investigated the hypothesize-and-test view more directly. One way this was done was by manipulating the “availability” of perceptual hypotheses. For example, (Bruner & Minturn, 1955) manipulated what they called the “readiness” of the hypothesis that stimuli were numbers as opposed to being 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 impoverished conditions (e.g., with an unfocused picture), then the perception of the same stimulus is impaired, so it takes longer or requires a better display before a figure is correctly recognized when it is first presented unfocused. When subjects make incorrect guesses as the focus is gradually improved, they eventually perform worse compared with when a false hypothesis is not made (Bruner & Potter, 1964; Wyatt & Pola, 1979).

Because of these and other types of experiments showing contextual effects in perception, the belief that perception is thoroughly contaminated by such cognitive factors as expectations, judgments, beliefs and so on, became the received wisdom in much of psychology, with virtually all contemporary elementary texts in human information processing and vision taking that point of view for granted (Lindsay & Norman, 1977; Sekuler & Blake, 1994). The continuity view also became widespread within philosophy of science. Thomas Kuhn (Kuhn, 1972) 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. The explanation was that theoretical notions in different theories are essentially incommensurable and so evidence itself is contaminated by the theoretical systems within which scientists worked. Philosophers of science like (Feyerabend, 1962) and (Hanson, 1958) argued that there was no such thing as objective observation since every observation was what they called “theory laden”. These scholars frequently cited the New Look experiments showing cognitive influences on perception to support their views. Mid-nineteenth-century philosophy of science welcomed the new holistic all-encompassing view of perception that integrated it into the general framework of induction and reasoning.

2.2.2 The perspective of neuroscience

During the heyday of the New Look (in the 1960s and the 1970s), speculative neuropsychology such as the influential work of the influential Canadian psychologist Donald Hebb, was in general sympathy with the
interactionist view (see Hebb, 1949). However, the important discovery of single-cell receptive fields and the hierarchy of simple, complex, and hypercomplex cells (Hubel, 1962) gave rise to the opposing 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 by the mere existence of cells that responded to more and more abstract properties, 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 top-down or centripetal influences (i.e., from higher brain regions to sensors) also occurred in the nervous system. For example, (Hernandez-Péon, Scherrer, & Jouvet, 1956), showed that the auditory response in a cat’s cochlear nucleus (the first neural way-station from the ear) was attenuated when the cat was paying attention to some interesting visual stimulus. More recently, the notion of focal attention has begun to play a more important role in behavioral neuroscience theorizing and some evidence has been obtained showing that the activity of early parts of the visual system can indeed be influenced by selective attention (e.g., Haenny & Schiller, 1988; Moran & Desimone, 1985; Mountcastle, Motter, Steinmetz, & Sestokas, 1987; Sillito, Jones, Gerstein, & West, 1994; van Essen & Anderson, 1990). There is even recent evidence that attention can have long-term effects as well as transitory ones (Desimone, 1996). Some writers (e.g., Churchland, 1988) have argued that the presence of outgoing (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 unknown except where it has been shown that they are concerned with the allocation of attention. I will argue later that allocation of focal attention is indeed a principal means by which top-down effects can occur in vision, but that these do not constitute cognitive penetration.

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 I claim is prior to the operation of early vision – in particular, they constitute an attentional selection of relevant properties. By and large the neuroscience community (at least since the influential work of neuro-computationalist, David Marr) is now interested in how the visual system decomposes into separate modules and is comfortable with the idea that vision itself is far less dependent on cognition than assumed by the more behavioral psychologists of the same era. I will look at some of the evidence that is considered relevant to this newer perspective in section 2.3 below.

2.2.3 The perspective of robot vision

Another line of support for the idea that vision implicates reasoning and memory comes from 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 (i.e., robots) 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 to 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 (block-shaped) 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 Perceptron (Rosenblatt, 1959), Uttley’s Conditional Probability Computer (Uttley, 1959), and Selfridge’s Pandemonium (Selfridge, 1959).

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 is that if a uniform intensity-gradient threshold is used as a criterion for the existence of edges in the image, it invariably results in one of two undesirable situations. If the threshold is set low it leads to the extraction of a large number of features that corresponded to shadows, lighting and reflectance variations, noise, or other luminance differences unrelated to the existence of real edges in the scene. On the other hand, if the threshold is set higher, then many real scene edges that are clearly perceptible by human vision are 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 (e.g., Kelly, 1971; Shirai, 1975) based on this approach. The idea that knowledge is needed at every level in order to recognize objects was strongly endorsed by (Freuder, 1986) in his proposal for a system that would use a great deal of specialized knowledge about certain objects (e.g., a hammer) in order to recognize these objects in a scene. Riseman and Hanson also took 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” (Riseman & Hanson, 1987, p 286)). Indeed a rather heated debate ensued between supporters of the bottom-up view (Clowes, 1971) that utilized line-finders in the initial stage of processing, and those who believed that vision systems would have to be heavily knowledge-based all the way down (Michie, 1986).

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 for 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 of possible objects that could occur in the input scene. 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 that object. 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

13 An alternative, that is sometimes also referred to as a “model based” approach, that uses some form of “general purpose” model of objects (Lowe, 1987; Zucker, Rosenfeld, & David, 1975) — or even of parts of such object (Biederman, 1987) – does not fall into this category because the models are not selected on the basis of expectations about the particular situation being observed (where the latter depends on what the observer knows and believes). This type of constrained perception falls into the category of “natural constraint” approaches that will be discussed in section Chapter 3.
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 early speech recognition systems developed at BBN (Woods, 1978) (known as SPEECHLIS or HWIM, for “Hear What I Mean”) is 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, including HEARSAY (described by Reddy, 1975), and Harpy, (described by Newell, 1980a), also use multiple sources of knowledge and introduced a general scheme for bringing knowledge to bear in the recognition process. These 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 (the relatively fixed structural properties of the system) permits any relevant source of knowledge to contribute to the recognition process at every stage. This general scheme has also been used as the basis for vision systems such as those developed by (Freuder, 1986; Riseman & Hanson, 1987). Figure 2-1 below shows the structure of such a system (showing both speech and vision experts) illustrating how each expert can have an input at any stage in the analysis, providing a completely open system.

Figure 2-1: Sketch of the “blackboard architecture” used by the HEARSAY speech understanding system, as well as by some computer vision systems.

The idea of complete freedom of communication among different “experts” (through the common “blackboard”) received wide recognition in many areas of psychology and artificial intelligence. In fact it was for a time the received wisdom for how such pattern recognizers as systems for reading text might be organized. A popular idea, very closely related to the blackboard architecture, was based fairly directly on Selfridge’s Pandemonium idea, in which various experts competed for the attention of an executive decision maker, as illustrated in Figure 2-2 in a popular text on human information processing (Lindsay & Norman, 1977).
While it is true that computational systems that make use of knowledge do better than ones that do not, I will 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 independence thesis and will be discussed in Chapter 3.

2.2.4 Seeing and knowing: Where do we stand?

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 vision is separate and distinct from cognition. The answer, I claim, lies not in denying the evidence that shows the importance of knowledge for visual apprehension (although in some cases we will need to reconsider the evidence itself), but in making certain distinctions. It is clear that what we believe about the world we are looking at does depend on what we know and expect. In that sense we can easily be deceived— as in magic tricks. But as noted earlier, seeing is not the same as believing, the old adage notwithstanding. In order to understand visual perception it is essential to distinguish certain stages in the process, in particular a stage – which I call early vision (after David Marr and other vision scientists) – that is prohibited from
accessing relevant knowledge of the world or of the particular scene – and other stages that are permitted, or in some cases are even required by the nature of the task (e.g., recognizing a familiar face) to access such knowledge. The knowledge-dependent (or cognitively penetrable) stages include a pre-perceptual stage, wherein vision is directed at relevant places or objects in a scene, and a post-perceptual stage, in which memory is accessed and judgments are made about what is in the scene. The idea of drawing a sharp distinction between parts of visual perception and cognition, or between stages of visual perception has been anathema to much of psychology and contemporary scholarship in general, although (Fodor, 1983) has done much to revive its popularity. In what follows I will suggest some reasons why such a distinction is empirically justified.

2.3 Some reasons for questioning the continuity thesis

Before getting into the details of some methodological and experimental findings supporting the thesis that a major part of the visual process is cognitively impenetrable, I provide a brief summary of why I believe this to be the case despite the sorts of evidence already sketched. Here are four general reasons why it makes sense to consider the possibility that there is a principled demarcation between early vision and cognition.

2.3.1 Evidence of the cognitive impenetrability of illusions

As Bruner himself noted (see note 12): perception appears to be rather resistant to rational influence. It is a remarkable fact about the perceptual illusions that knowing about them does not make them disappear: Even after you have had a good look at the Ames room—perhaps even built it yourself—it still looks as though the person on one side is much bigger than the one the other side (Ittelson & Ames, 1968). Knowing that you measured two lines to be exactly equal does not make them look equal when arrowheads are added to them to form the Müller-Lyer illusion, or when a background of converging perspective lines are added to form the Ponzo illusion, as shown in Figure 2-3.

![Figure 2-3. Illustrations of the Ponzo illusion (on the left) and the Müller-Lyer illusion (on the right). In both cases the horizontal lines above one another are the same length.](image)

For another example in which the visual system’s internal mechanisms override your knowledge, consider the blocks in Figure 2-4. Which of the top faces of blocks, A or C, is identical in size and shape (except for being rotated) to face B? If you check using a ruler or by cutting the figures out of paper you will find that the face labeled A is identical to the face labeled B while C is quite different. Notice that such illusions are not just stubborn, in the way some people appear unwilling to change their minds in the face of contrary evidence: it is simply impossible to make something look to you the way you know it really is. 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. Later in section 2.5 I will examine
(and reject) claims that certain kinds of illusions (e.g. reversals of ambiguous figures and perceptual closure of difficult percepts) are susceptible to cognitive influences.

![Figure 2-4. Which of these figures has the same top face as B (A or C)? (after Shepard, 1981).](image)

### 2.3.2 Evidence of the independence of principles of visual organization and of inference

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 quite different from the principles of rational reasoning. These principles of perception differ from the principles of inference in two ways.

1. First, unlike the principles of inference, perceptual principles are responsive only to visually presented information. One way to think of the difference between visual representations and thoughts is to think of the representations that occur within the visual system as being in a proprietary vocabulary, distinct from the vocabulary that occurs in representations of thoughts and beliefs. This vocabulary encodes such perceived properties as which regions of a scene go together as a single object, which contours go with which surfaces, which surfaces partially occlude other surfaces, and so on. Perceptual principles specify how certain encoded properties (or “scene labels”) go with other encoded properties. In computer vision a major part of early vision is concerned with what is called scene labeling or label-propagation (Chakravarty, 1979; Rosenfeld, Hummel, & Zucker, 1976), wherein principles of label-consistency are applied to represented features in a scene in order to compute the correct labeling (we will see examples of this sort of labeling in the next chapter, section 3.1.1.1). Principles of visual interpretation explain why it is that the way you perceive some aspect of a visual scene determines the way you perceive another aspect of the scene. When a percept of an ambiguous figure (like a Necker Cube) 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 as “perceptual couplings” (Epstein, 1982; Rock, 1997). (Gogel, 1997/1973) attempted to capture some of these regularities in what he called perceptual equations. Such equations provide no role for what the perceiver knows or expects because the perception of visual form is not sensitive to the beliefs that perceivers have about what the scene being examined should look like, given what they know about the circumstances of that scene. The question of why certain principles (or equations) should be embodied in our visual systems is one that can only be answered by examining the function that vision has served in allowing organisms to survive in our kind of world. The particular equations or couplings may be understood in relation to the organism’s needs and the nature of world it typically inhabits (see Chapter 3 for a discussion of perceptual coupling). A more extensive discussion of the notion of natural constraints is provided in section 3.1.1).

2. Second, the principles of visual organization are quite different from those of logical inference and often do not appear to conform to what might be thought of as tenets of “rationality”. 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, first studied by Michotte, 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 (see Figure 1-6). 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; 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 2-5 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. In this and very many other such examples (many are discussed by Kanizsa in the papers cited) the simplest figure is not the one chosen, but rather one that conforms to some special principle that applies to local regions of the image. It is precisely because of this that there are visual illusions; the visual system cannot ignore its wired-in geometrical principles in favor of what the perceiver knows to be true.\textsuperscript{14}

\textbf{Figure 2-5: Kanizsa “amodal completion” figure. The completion preferred by the visual system is not the simplest figure despite the flanking examples of such figures. (After Kanizsa, 1985).}

\subsection*{2.3.3 Neuroscience evidence for top-down attenuation and gating of visual signals}

In section 2.2.2 I mentioned the existence of top-down neural pathways in vision. I also pointed out that the primary function of these pathways appears to be to allow attention to selectively sensitize or gate certain objects or regions of the visual field, as well as certain physical properties of the stimuli. Even if such effects ultimately originate from “higher” centers, they constitute one of the forms of influence that I claimed was prior to the operation of early vision – i.e., they constitute an early attentional selection of relevant 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 I have already alluded – modulation associated with focal attention, which can originate either from events in the world which attract attention automatically (exogenous control) or from voluntary 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 pathway) 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),

\textsuperscript{14} Of course if the principles by which vision constructs a representation of the world were totally capricious we would find ourselves walking into walls and generally not faring well in our commerce with the world, so the principles must be ones that more often than not yield a true description of the world in typical situations, even if they do not do it by rational inferential means, but by virtue of local geometrical principles. We will have occasion to return to this point (in Chapter 3), referred to as the tendency of the visual system to embody \textit{natural constraints}, since this is an important general principle which explains why vision, which is insensitive to knowledge and processes of rational inference, nevertheless manages to solve visual problems mostly the way they need to be solved for the purposes of survival.
as well as the review in (Lynch, 1980). There is now a great deal of evidence suggesting that the dorsal system is specialized for what (Milner & Goodale, 1995) call “vision for action”. What has not been reported, to my knowledge, 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 (I will take up this topic again in Chapter 7).

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. 15 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 experimenter 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 [for the firing of certain cells] 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 terms of the present thesis, I would say that although what such cells encode may be complex, it is not sensitive to the cognitive context.

2.3.4 Evidence of dissociation of visual and cognitive functions in the brain

One intriguing source of evidence that vision can be separated from cognition comes from the study of pathologies of brain function that demonstrate dissociations among various aspects of vision and cognition. Even when, as frequently happens, no clear lesion can be identified, the pattern of deficits can provide evidence of certain dissociations and co-occurrence patterns of skill. They thus constitute at least initial evidence for the taxonomy of cognitive skills. The discovery that particular skill components can be

---

15 Changes in attenuation/sensitivity can result in different percepts. Whether or not this constitutes cognitive penetration depends on two things: (a) what kinds of perceptual changes can be underwritten by changing activation levels and (b) whether the particular changes can be attributed to appropriate differences in the contents of beliefs and expectations. As for (a) it is dubious that the sort of influence that allows people to see a word whose meaning is predicted from the context, to see monsters when they are afraid or to see food when they are hungry (as claimed by the New Look school) can be supported by differences in attenuation or sensitivity (see the discussion of this issue in Chapter 4). But (b) is even more germane because it matters whether or not the influence comes from within the visual system or from outside. We know that how you interpret a certain portion of an image depends on how you interpret other parts of the image. Such intra-visual effects do not constitute cognitive penetration. Unfortunately, the neuroanatomical data do not cast light on that question.
dissociated from other skill components (particularly if there is evidence of a double-dissociation in which each component occurs without the other), provides a prima facie reason to believe that these skills might constitute independent systems.

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 in a book by Glyn Humphreys and Jane Riddoch (Humphreys & Riddoch, 1987). After suffering a stroke that resulted in bilateral damage to his occipital lobe, the patient was unable to recognize familiar objects, including faces of people well-known to him (e.g., his wife), and found it difficult to discriminate among simple shapes, despite the fact that he did not exhibit any intellectual deficit. As is typical in visual agnosias, this patient showed no purely sensory deficits, showed normal eye movement patterns, and appeared to have close to normal stereoscopic depth and motion perception. Despite the severity of his visual impairment, the patient could do many other visual and object-recognition tasks. For example, even though he could not recognize an object in its entirety, he could recognize its features and could describe and even draw the object quite well – either when it was in view or from memory. Because he recognized the component features, he often could figure out what the object was by a process of deliberate problem-solving, much as the continuity theory claims occurs in normal perception, except that for this patient it was a painstakingly slow process. From the fact that he could describe and copy objects from memory, and could recognize objects quite well by touch, it appears that there was no deficit in his memory for shape. These deficits seem to point to a dissociation between the ability to recognize an object (from different sources of information) and the ability to compute an integrated pattern from visual inputs 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. The pattern of dissociation shows the intact capacity to extract features together with the capacity to recognize objects from shape information is insufficient for visual recognition so long as the unique visual capacity for integration is absent. 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.

2.3.5 Evidence that cognitive penetration occurs in pre-perceptual and post-perceptual stages

Finally, there are methodological questions that can be raised in connection with the interpretation of empirical evidence favoring the continuity thesis. There are certain 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 after what I have called early vision, some of which are summarized in the next section, but are treated at greater length in (Pylyshyn, 1999). Thus 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. There is also evidence that in other cases where beliefs and past experience appear to influence what we see, 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 the perceiver learns to allocated attention to different features or places in a stimulus. Both these cases are briefly reviewed in the next section.
2.4 Distinguishing perceptual and decision stages: Some methodological issues

Some of the arguments among researchers, concerning whether early vision is cognitively penetrable or encapsulated, rest on certain considerations of experimental methodology. The question of whether vision is encapsulated might be approached by trying to divide visual processing into stages, for example a stage that might correspond to what I have called early vision and a stage that is concerned with making decisions about what the stimulus was and what response to make to it. Such distinctions raise questions of experimental method: Given that context affects the way observers respond to a stimulus, how does one tell whether this is because the context affects how they see it, what they recognize it to be (i.e., what they see it as), how they classify it in relation to things they have seen before, or what they decide to do in a particular experimental context (i.e., what response they decide to make). The attempt to distinguish such stages has a venerable history in the study of perception, going back to the earliest days of experimental psychology. The question received a major impetus in the 1950s with the development of such theoretical instruments as Signal Detection Theory and the use of certain patterns of electrical potentials on the scalp (called Event-Related Potentials, or ERPs) and other techniques for dividing the information process into stages and for determining which stage is responsible for certain observed phenomena.

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 the newly developed technique of Signal Detection Theory (SDT) (Tanner & Swets, 1954) in support of the conclusion that psychophysical functions involve decisions. What Bruner glossed over, however, is that the work on signal detection analysis not only shows that decisions are involved in such psychophysical tasks as threshold measurements, it also shows that such tasks typically involve at least two stages, one of which, sometimes called “stimulus detection,” is immune from cognitive influences, while the other, sometimes called “response selection”, is not. In principle, the theory provides a way to separate these two stages and to assign independent performance measures to them: To a first approximation, detection is characterized by a sensitivity measure, usually denoted as \( d' \), while response selection is characterized by a response bias or response criterion measure denoted as \( \beta \). Only the second of these measures was thought to capture the decision aspect of certain psychophysical tasks, and therefore, according to this view, it is the only part of the process that ought to be sensitive to knowledge and utilities.

The idea of factoring visual information processing into roughly a stimulus-processing 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), mathematical techniques such as the use of the “attention operating characteristic” (Sperling & Melchner, 1978), the use of event-related potentials (ERPs) and other methods devised for specific situations. Numerous experiments have shown that certain kinds of cognitive malleability observed in experiments on perception is due primarily to the second of these stages—the stage at which a response decision is made (Samuel, 1981). But not all the results support this conclusion; a few studies have also shown that stages prior to response selection are also changed by changes in expectations (i.e. by the cognitive context). When I began a review of techniques for analyzing information processing into stages (which led to the analysis presented in, Pylyshyn, 1999) I had hoped that these techniques would allow us to separate those effects attributable to the perceptual stage from those attributable to subsequent decision stages. This goal was only partly achieved, however, because it turned out that the stages distinguished by these techniques are too coarse for this purpose and do not correspond exactly to the distinction that is relevant to the independence thesis. What the techniques are able to show is that certain influences, sometimes taken to demonstrate the cognitive penetration of vision, have their locus in the selection of a response and in the preparation of an actual overt response. When the effect is found to lie outside the response selection stage, however, the methods are unable to distinguish whether this occurs in what I have been calling early vision or in some other part of the information-processing stream. There is clearly much more going on when we perceive than detecting a stimulus followed by the preparation of the action of responding. For example, apart from
various shape and surface computations there is the recognition of the stimulus pattern as one that has been seen before. Since this sort of recognition (seeing the stimulus as something familiar) inherently involves accessing memory, it falls outside what I call early vision, and yet it is not a case of preparing an actual response. Thus the use of methods such as those of signal detection theory and ERP provides an asymmetrical test of the cognitive penetrability of vision. Showing that the effect of changing beliefs or expectations operates entirely at the response selection stage (i.e., it affects $\beta$ but not $d'$) shows that in this case the belief change does not influence early vision. On the other hand, showing that the effect operates at the so-called stimulus detection stage (or operates by influencing the sensitivity measure $d'$ and not $\beta$) does not show that early vision is cognitively penetrable, because the detection stage includes more than just early vision.

The problem with all the techniques of stage analysis that have so far been proposed is this: Whether the technique is simple or sophisticated, it usually ends up distinguishing a stage of response preparation from everything else concerned with processing visual information. But that distinction is too coarse for our purposes if our concern is whether an intervention affects the visual process or the post-perceptual recognition-inference-decision-response process. To determine whether early vision is cognitively penetrable one needs to make further distinctions within what stage-theorists call the stimulus detection or stimulus evaluation stage. In particular one needs to factor out functions such as categorization and identification, which require accessing general memory, from functions of early vision, such as individuating objects and computing spatial relations among them, which, by hypothesis, do not. That is why we find, not surprisingly, 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. A more detailed technical discussion of this claim is provided in (Pylyshyn, 1999) and the reader who is interested in the underlying assumptions is invited to consult that paper.

2.5 Some examples in which knowledge is claimed to affect perception

2.5.1 “Intelligent” interpretations of inherently ambiguous information

A number of writers (Gregory, 1970; Rock, 1983) have noted that the visual system delivers unique interpretations of visual (optical) information that is inherently ambiguous, and that when it does so it invariably produces an interpretation that is “intelligent” in that it appears to take into account certain cues in a way which suggests that, to use an anthropomorphic phrase, “it knows how things in the world work”. These examples are indeed among the most interesting cases to consider from the perspective of the independence thesis, both because they constitute impressive demonstrations of smart vision and also because they provided major challenges to theories of computer vision and well as some of its most impressive successes. The successes followed the seminal work of David Marr (Marr, 1982) and are based on the discovery of certain constraints inherent in the visual system (so-called “natural constraints”). Because of the importance and far-reaching implications of this idea I will postpone this discussion until the next chapter where I consider the nature of the “architecture,” or relatively fixed structural properties of the vision system.

2.5.2 Experience and “hints” in perceiving ambiguous figures and stereograms

So far I have suggested that many cases of apparent penetration of visual perception by cognition are either cases of top-down effects occurring within the visual system (discussed in section 2.3.3), or are cases in which knowledge and utilities are brought to bear at the pre-perceptual stage (by determining where to focus attention) or at the post-perceptual stage (by determining which possible interpretation provided by early vision to favor). 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” (such as illustrated in Figure 2-6 or Figure 2-7), ambiguous figures, and stereograms,
including the famous Magic Eye® posters that show 3D images based on what are known as autostereograms. I will now suggest 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.

Figure 2-6. A famous picture of a Dalmatian dog, by R.C. James, which typically comes into view suddenly.

It is widely believed that providing “hints” can improve a person’s ability to recognize a fragmented figure such as that shown in Figure 2-6 (and other so-called “closure” figures such as those devised by Street, 1931), some of which are shown in Figure 2-7 below. Yet that claim has rarely been tested experimentally. There is some evidence that priming with completed pictures does improve recognition performance, although (Gollin, 1960) showed that training with the fragmented pictures using an ascending method of limits produced more learning than extensive training with completed pictures. In a related finding, (Snodgrass & Feenan, 1990) showed that perceptual closure can best be primed by showing fragmented figures that were partially completed. Priming by complete figures was much poorer than priming by partially completed figures. These cases, however, only show the effect of visually presented information on perceptual closure, and even then they find that partial information is better than complete information in priming identification. Since complete pictures carry full information about the identity of the objects it seems that it is not knowing the identity of the objects that is most effective, but some other perceptual factor. Perhaps the attempt to obtain closure in the partially-completed figures initiates a search for visual cues which then helps to focus attention. The partially-completed figures used by Snodgrass and Kelly allow the viewer to see both the identity of the complete figure and also the nature and location of some of the fragments that form part of the completed figure, so it provides better cues as to where to focus attention in the fragmented figure.

The effectiveness of verbal “hints” was investigated directly by (Reynolds, 1985). Reynolds used the figures taken from the Street’s Gestalt Completion Test (Street, 1931) (a few of which are shown in Figure 2-7) and 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 eventual recognition but did not decrease the time it took to do so.
Seeing and Visualizing: It's Not What You Think

Zenon Pylyshyn

(9/27/2002 2-17)

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 typically 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 Girgus, Rock, & Egatz, 1977; Rock & Anson, 1979). 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. If anything, the fact that the principle effect on perception comes from knowing that a meaningful figure could be seen (in the case of fragmented figures), or the effect of knowing that a figure could be ambiguous, adds more credence to the earlier suggestion that such information initiates a search for cues as to where to focus attention, as opposed to a perceptual process. This would also explain why closure of fragmented figures takes so long to attain, compared to the very short time to perceive a picture (which appears to be on the order of a tenth of a second, see for example, Potter, 1975).

Figure 2-7 Examples of fragmented or “closure” figures used by (Reynolds, 1985) (based on Street’s “Gestalt Completion Test” Street, 1931).

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 (see note 16) 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 simply 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 by voluntary cognitive processes – though not in terms of the content of the percept but in terms of the location at which the independent visual process is to be applied. In Chapter 4, I

---

16 Keith Humphrey, at the University of Western Ontario, frequently provides class demonstrations in which he shows fragmented figures in class to a group of students, some of whom were given written hints in advance. Judging by the latency of hand-raises, the hints have tended to speed up the recognition. Given the informal nature of this demonstration the results should be received with caution and a controlled version of the experiment seems merited. But an interesting observation that Humphrey reported is that the time taken for closure was very long – in the order of minutes, rather that the fractions of a second required for normal perception of such simple forms. Clearly something other than visual recognition is involved in such cases – most likely some sort of problem-solving based on searching for critical cues.
will argue that an important way in which cognition can affect the outcome of visual perception is by directing the independent visual system to focus attention at particular (and perhaps multiple) places in the scene.

A very similar story concerning the locus of cognitive intervention applies in the case of other ambiguous displays. When one percept is attained and observers know that 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 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.” This is not what has been reported in experiments with naïve subjects, where the only factor that has been found to be relevant is the knowledge that the stimulus is ambiguous. Moreover, as I have already suggested, there is a simple mechanism available for some degree of control 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. Some parts of a figure tend to have a bias toward one interpretation while other parts have a bias towards another interpretation; consequently changing the locus of attention may change the interpretation (e.g., there appears to be a bias toward seeing the attended parts of a figure as being closer to the viewer). If there is a bias in the interpretation of a part of the figure this will tend to influence the interpretation of the remainder of the figure. But 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 where it has been suggested that hints and prior knowledge affect perception. For example, the fusion of stereo images, such as Figure 3-18 and even more so with stereo images made of random dots – a form of presentation invented by (Julesz, 1971) called “random dot stereograms” – is often quite difficult and was widely thought to be improved by giving people information about what they should see (the same is true of the popular autostereograms).17 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) found 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 of the eyes required in order to fuse the display. One might surmise that it may also be the case that directing focal attention to 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.

2.5.3 Learning to “see” differently: A case of controlling focal attention?

There is general agreement that focal attention can be directed, either extrinsically (by virtue of certain properties of the stimulus) or intrinsically (by central processes, operating voluntarily). Whichever form of attention allocation occurs, both the conditions triggering automatic allocation and the conditions that determine voluntary allocation may be modified over time. There is evidence that just as eye movements can be modulated by experience (e.g., Shapiro & Raymond, 1989), so can the allocation of attention. This being the case, one might reasonable wonder whether some or all of the effects studied under the category of “perceptual learning” might not be attributable to the indirect effect of learning to allocate attention.

17 A commercial version of these autostereograms are the Magic Eye® 3-D posters. As of this writing, examples of these may be seen at: http://www.magiceye.com. Other stereo demonstrations can be found at http://www.vision3d.com/optical/index.shtml#stereogram.
Clearly we can decide to focus our attention on certain parts of a scene rather than others. Earlier I claimed that this attention allocation was one of the two principle ways for cognition to intervene in influencing visual perception; the other being the post-visual determination of the identity of familiar scene contents. I suggested that such attention- allocation was responsible for the speedup in achieving perceptual closure after exposure to a previous instance of the closure figure. Similarly, the effect of certain types of experience on our ability to fuse random-dot stereograms and autostereograms was attributed to learning to focus attention on features that guide appropriate eye vergence. For example, 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 retinal location and orientation 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 & Braddock, 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.

The improvement in a perceptual skill following learning which features are critical to the skill is a quite general phenomenon. In chapter 1 this claim was illustrated in the case of chess masters, whose memory for chess positions was found to be highly dependent on whether or not the chess positions could be interpreted in terms of a real chess match. This was taken as evidence that chess masters differ from beginners because (1) masters have a code for a large number of positions and (2) masters know where to look for significant chess-relevant patterns. It is frequently the case that expertise in a visual skill depends on finding and encoding certain special features that are diagnostic for the particular distinctions that define the expertise in question. I will present some quite general cases of this principle below.

2.5.3.1 Cultural and linguistic effects on perception

It was widely believed that how people see the world is conditioned by their culture and even by the language they speak. This thesis, called Linguistic Relativity was articulated by the anthropologist Benjamin Lee Whorf and linguist Edward Sapir and in sometimes referred to as the Sapir-Whorf hypothesis. The thesis was widely accepted during the time when the New Look in perception held sway. It was congenial to the tabula rasa view of human nature that pervaded both philosophical and psychological behaviorism, as well as view of the “perfectibility of human nature” that grew out of the enlightenment. It is the view that human nature is largely the product of our experiences. Although linguistic relativity in the strong Worflian form has been discredited in the past several decades, a modified and somewhat weaker version of this hypothesis has once again seen some support (e.g., Gumperz & Levinson, 1996). Most of the recent work only supports the claim that language provides resources over and above the nonlinguistic concepts that are needed to encode the visual world (and even so, many believe that the new relativism also rests on shaky foundations, see Li & Gleitman, 2002; Pinker, 1994). Notice that the new linguistic relativism is not the same as the New Look view, which claims that what we see, and how we see, depends on our beliefs. Even if the newer evidence and arguments are sustained, the claims that they make only concern long term effects of language on visual interpretation and encoding, not the sort of effects that support the idea that vision and cognition merge continuously into one another.

2.5.3.2 The case of “expert” perceivers

Another apparent case of penetration of vision by knowledge occurs in the case of “expert” perceivers of various kinds – people who are able notice patterns that the rest of us fail to 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 typically highly deft, rapid, and unconscious. When asked how they do it, experts typically say that they can “see” certain properties by “just looking”. But
what research is available shows that often what the expert has learned is not a “way of seeing” as such, but rather some combination of task-relevant mnemonic skills (knowing what kinds of patterns to look for) with knowledge of where to look or where to direct their attention.

The first type of skill is reminiscent of the conclusion reached by (Haber, 1966) that “preparatory set” operates primarily through mnemonic encoding strategies. It is most clearly illustrated in 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, see Figure 1-19e). 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 that they can use to classify or encode genuine chess positions, though 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 I 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 faster visual information processing capabilities has been found among athletes (e.g., Abernethy, Neil, & Koning, 1994; Starkes, Allard, Lindley, & O’Reilly, 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 & Umiltá, 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 abilities related to the learned skills of identifying, predicting and therefore attending to the most relevant places. Indeed, a report published in the May 5, 1969 issue of *Sports Illustrated* gave the reaction time of the great heavyweight boxer, Muhammed Ali, in responding to a light as 190 ms, which is within the normal range (reported in Keele, 1973, p76). Presumably Ali’s lightening-fast performance in the ring was due in part to his being able to anticipate his opponent’s moves – that and the fact that he could move his arm 16.5 inches following his initial reaction time in only 40 ms! Being a fast reactor in a visual-motor skill does not always mean processing visual information better or more quickly.

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. A remarkable case of such expertise was investigated by (Backus, 1978; 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. Beiderman and Shiffrar found that telling novices where the relevant information was located allowed them to quickly become experts themselves. What the “telling” does – and what the experts had tacitly learned – is how to bring the independent visual system to bear at the right spatial location, and what types of patterns to encode into memory, both of which are functions lying outside the visual system itself.

Note that this is exactly how I 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. I 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 below and section 6.5 for more on this claim).

2.5.3.3 Attention and perceptual learning

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. For example, the way people categorize objects and properties – and even the discriminability of features – can be altered through prior experience with the objects. In this same tradition, recent studies (Goldstone, 1994, 1995) showed that the discriminability of stimulus properties is altered by pre-exposure to different categorization tasks. (Schyns, Goldstone, & Thibaut, 1998) have argued that categorization does not rely on a fixed vocabulary of features but that feature-like properties are “created under the influence of higher-level cognitive processes ...when new categories need to be learned ...”.

This work is interesting and relevant to the general question of how experience can influence categorization and discrimination. The claim that a fixed repertoire of features at the level of cognitive codes (though clearly not at the level of basic sensory receptors) is inadequate for categorization 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 I have been developing it here for the following reasons:

1. Although some perceptual-learning researchers have claimed that what may be altered is the basic early-vision processes, this is far from clear from the results themselves. In all reported cases what is actually being measured is observers’ differential use of certain features compared with their use of other available features. But this could be the result of one of the two processes lying outside of early vision that were discussed earlier. These are: (a) selective attention being applied to one physically-specifiable dimension over another (e.g. one location over another, color over luminance, height over width, and so on) or (b) one perceptual feature being weighted more or less than another and/or combined in different (nonlinear) ways in the post-perceptual categorization decision-process (which could, in effect, result in the creation of different feature-clusters).

2. Even if the claim that there is low-level learning in early vision can be sustained (i.e., even if the effect is due to neither learning to allocate focal attention nor to post-perceptual decision processes) the claims are not directly relevant to the discontinuity thesis. The shaping of basic sensory processes by experience is not the same as cognitive penetration. The latter entails that the content of percepts (not just the discriminability of features) is rationally connected to beliefs, expectations, values, and so on-regardless of how the latter are arrived at or altered. That is the type of influence that the New Look was concerned with and that is the kind of influence that concerns us here. It is this kind of influence that is at the core of the discontinuity thesis and the whole issue of architectural constraints. The tuning of the sensitivity of sensory systems (not only organs but also cortical centers) by evolution or by prolonged experience is not the same as cognitive penetration, the determination of what we see in any particular instance by our immediate expectations, beliefs, needs, or inferences.

3. Finally, with regard to the genesis of visual expertise, it should be noted that there is no a priori reason why a post-perceptual decision process might not, 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, 1984a) referred to as “compiled transducers”. Compiling complex new transducers is a process by which formerly post-perceptual processing may become part of early vision by becoming automatized and encapsulated, and may even have its own local storage (see Pylyshyn, 1984a, chapter 9, for more on this point). If the resulting process is cognitively impenetrable – and therefore systematically loses the ability to access general memory, then, according to the view being advocated here, it becomes part of the visual system. Consequently, it is consistent with the present framework, that new complex processes could become part of the early vision system over time: Cognitive impenetrability and diachronic change are not incompatible. How processes can become “compiled” into the visual system remains unknown, although according to the Allan Newell’s levels-taxonomy (Newell, 1990) the
process of altering the encapsulated visual system should take one or two orders of magnitude longer to accomplish than the duration of basic cognitive operations themselves (i.e. in the order of minutes as opposed to fractions of a second) – and would very likely require repeated experience, as is the case with many of the perceptual learning phenomena.

### 2.5.4 Focal attention as an interface between vision and cognition

One of the features of perceptual learning already noted is that learning may lead to attention being focused on the most relevant objects, regions or properties of the visual field. Spatial, or object-based (see chapter 4) focusing of attention is perhaps the most important mechanism by which the visual system adjusts rapidly to an informationally-dense and dynamic 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). Visual attention is a fundamental mechanism in both vision and visualization (or mental imagery) and a careful analysis of the various forms and functions of focal attention goes a long way towards explaining many phenomena of perceptual learning, perceptual organization, mental imagery and even the source of certain phenomenal experiences of the spatial extension and spatial stability of both visual percepts (discussed in chapter 1 and chapter 5) and mental images (discussed in section 6.5). Consequently chapters 4 and 5 will be devoted to a discussion of these issues. For present purposes I wish merely to note that there are logically two possible loci where a mechanism like selective attention could operate. First, it might operate on the input to the visual system, by enhancing and/or attenuating certain properties of the input. For example, it might enhance certain spatial locations or certain regions that constitute individual visual objects. Secondly, it might operate by enhancing and/or attenuating the availability of certain perceptual categories, as proposed by Bruner. This is what Bruner meant by "perceptual readiness" – it is the ready availability of categories of perception. From our perspective, the latter locus of attention occurs after the early visual process and may even include selective encoding into memory or selective retrieval from memory. If selective attention operates over the input prior to early vision, then it must operate over properties that are primitive and pre-perceptual, such as spatial location or perhaps spatial frequency, otherwise appeal to selective attention would be circular. In that case perceptual learning might consist in learning where to look or something equally basic (as suggested in section 2.5.3). On the other hand, if attention is viewed as being at least in part a post-perceptual process, ranging 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 within the visual system, but rather involves post-perceptual decision processes based on knowledge and experience, however tacit and unconscious these may be. In the everyday use of the term “attention” these two loci are not distinguished. Attention is viewed as any process that selects some aspect of perception from other aspects. As a consequence, the attempt to judge whether attention occurs early or late in vision has had mixed results: Evidence for both “early selection” and “late selection” theories has been found. But it remains possible, and indeed seems reasonable, that attention operates at both loci. Our position is that attention operates at both loci but not in between – i.e. not within the early vision system itself.

### 2.6 Conclusions: Early vision as a cognitively impenetrable system

In this chapter I have considered the question of whether early visual perception is continuous with cognition or whether it is best viewed as a separate process, with its own principles and possibly its own internal memory, isolated from the rest of the mind except for certain well-defined and highly circumscribed modes of interaction. In the course of this analysis I have touched on many reasons why it appears on the surface that vision is part of general cognition and thus thoroughly influenced by our beliefs, desires and utilities. Opposed to this interactionist view we found 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 Irvin Rock, one of the champions of the view that vision is “intelligent,” has said, “Perception must rigidly adhere to the appropriate internalized rules, so
Most of this chapter 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 some alleged cases of “hints” affecting perception, of perceptual learning, and of perceptual expertise were examined. I 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.

A number of clinical findings were also noted concerning the dissociation of cognition and perception which tend to substantiate the view that vision and cognition are independent systems (although some of this evidence will be presented in the next chapter). Very little has been said about the general issue of the nature of the output from the visual system. This question is taken up in the next chapter where it will be 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 looking up parts of the stimulus in a shape-indexed memory for identification. I will also considered the possibility that more than one form of output is generated, directed at various distinct post-perceptual systems. In particular I will examine the evidence that motor control functions may be served by different visual outputs than recognition functions – and that both are cognitively impenetrable.

In examining the evidence that vision is affected by knowledge and expectations some space was also devoted to methodological issues concerned with distinguishing various stages of perception. Although the preponderance of evidence locates cognitive influences in a post-perceptual stage, we found that the sort of stage analysis methods in general use within experimental psychology, provide a decomposition that is too coarse to establish whether the locus of cognitive effects is inside or outside early vision proper. In particular, both signal detection measures and event-related potentials fail to provide a way to examine a stage that corresponds to what I have been calling early vision. 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-visually processes.

The idea that early vision (which we might as well call “vision” since it is the part of the extensive overall process of acquiring knowledge through the visual modality that is special or proprietary to visual processing) is both complex and impervious to cognitive influences such as expectations and beliefs and desires will play an important role in our subsequent discussion about how vision interacts with the world and with what we know, with our cognitive representations. Henceforth in speaking about the connection between vision and the world or vision and visualization (mental imagery) “vision” will refer to the system that was identified in this chapter – the early vision system. The rest of what goes on when we visually perceive, being part of reasoning, recall, judgment and so on, will not be counted as part of the visual system proper. If the entire visual-cognition process were to count as “vision” then one would be forced to conclude that there is no such thing as vision proper, only cognizing. In that case there would be nothing particular to say about whether (and how) we reason using the visual system, or how what we see connects with the perceived world; at least no more to say about those connections than about the connection between how we reason about fashion and how we reason about physics, or how we reason about any pair of topics or subject matters. But as we have seen, vision is not just another subject matter to be thought about by the reasoning system. It is a system that has well-defined properties and principles of its own, so it is of considerable interest to ask how
(if at all) these principles interact with the principle by which we reason, recall, make judgments and understand language. Such questions could not be studied with any precision if it had turned out that there was no such thing as the visual system as a separate and autonomous set of processes. Luckily for us, and for a science of visual perception, this appears to be far from the case. Rather there appears to be a rich set of mechanisms and processes that we may say are proprietary to vision. The question of just how rich and complex these mechanisms and processes are is the burden of much that follows. In particular, if this technical sense of vision is to be of much interest, it must include a lot more than merely converting a retinal image into a mental image; it must include the process that allows us to see as much as we do see without benefit of what we know about the scene we are observing. We begin to explore this question in the next chapter.