26. Vision, Thinking, and Model-Based Inferences

Athanassios Raftopoulos

Model-based reasoning refers to the sorts of inferences performed on the basis of a knowledge context that guides them. This context constitutes a model of a domain of reality, that is, an approximative and simplifying to various degrees representation of the factors that underlie, and the interrelations that govern, the behavior of this domain.

This chapter addresses both the problem of whether vision involves model-based inferences and, if yes, of what kind; and the problem of the nature of the context that acts as the model guiding visual inferences. It also addresses the broader problem of the relation between visual processing and thinking. To this end, the various modes of inferences, the most predominant conceptions about visual perception, the stages of visual processing, the problem of the cognitive penetrability of perception, and the logical status of the processes involved in all stages of visual processing will be discussed and assessed.

The goal of this chapter is, on the one hand, to provide the reader with an overview of the main broad problems that are currently debated in philosophy, cognitive science, and visual science, and, on the other hand, to equip them with the knowledge necessary to allow them to follow and assess current discussions on the nature of visual processes, and their relation to thinking and cognition in general.

Helmholtz [26.1] famously maintained that perception is a form of inference; the brain uses probabilistic knowledge-driven inferences to induce the causes of the sensory input from this input, that is, to extract from the bodily effects of the light emanating from the objects in a visual scene as it impinges on our transducers the various aspects of the world that cause the input. The brain integrates computationally the retinal properties of the image of an object with other relevant sources of information to determine the object's intrinsic properties. *Rock* [26.2] claimed that the perceptual system combines inferential information to form the percept. From visual angle and distance information, for example, the perceptual system infers and perceives size. This inference may be automatic and outside the authority of the viewer who does not have control over it, but is an inference nevertheless.

Similarly, *Spelke* [26.3] suggests "perceiving objects may be more akin to thinking about the physical world than to sensing the immediate environment". The reason is that the perceptual system, to solve the underdetermination problem of both the distal object from the retinal image and of the percept from the retinal image, employs a set of object principles (the Spelke principles) that reflect the geometry and the physics of our environment. Since the principles can be thought of as some form of knowledge about the world, perception engages in inferential processes from some pieces of worldly knowledge and visual information to the percept, that is, the object of our ordinary visual encounters with the world.

Recently *Clark* [26.4] argued that:

> "To perceive the world just is to use what you know to explain away the sensory signal across multiple spatial and temporal scales. The process of perception is thus inseparable from rational (broadly Bayesian) processes of belief fixation [...] As thought, sensing, and movement here unfold, we discover no stable or well-specified interface or interfaces between cognition and perception. Believing and perceiving, although conceptually distinct, emerge as deeply mechanically intertwined."

The aim of this conglomeration of faculties that constitute perception is, therefore, to enable perceivers to respond, modify their responses, and eventually adapt their responses as they interact with the environment so as to tune themselves to the environment in such a way that this interaction be successful; success in such an endeavor relies on inferring correctly (or nearly so) the nature of the source of the incoming signal from the signal itself.

In all these views, the visual system constructs the percept in the way thinking constructs new thoughts on the basis of thoughts that are already entertained. In this sense, vision is a cognitive, that is, thought involving, process.

If perception is to be thought of as some sort of thinking, its processes must necessarily first include transformations of states that are expressed in symbolic or propositional form, and, second, these transformations must be inferences from some states that function as premises to a state that is the conclusion of the inference. That is to say, visual processes must be inferences or arguments, exactly like the processes of rational belief formation. These two conditions follow directly from the claim that perception is some sort of thinking, since the characteristic trait of thinking is drawing inferences (whether it be deductive, abductive, or inductive) operating on symbolic forms by means of inference rules that are represented in the system, although thinking is not reduced to drawing inferences this way. In view of these considerations, the principles guiding the transformations of perceptual states, that is, the principles (such as Spelke's principles) acting as the inference rules in perceptual inferences, must be expressed in the system and, specifically, must be represented in a symbolic form. Whenever the system needs some of the principles to draw an inference, it simply activates and uses them. In addition, the premises and the conclusion of a visual argument be represented in the viewer in a propositional-like, symbolic form.

If these conditions are met, perception involves discursive inferences, that is, drawing propositions or conclusions from other propositions acting as premises by applying (explicitly or implicitly) inferential rules that are also represented in the system. Clark's view quoted above seems to echo this thesis in so far as Clark conceives the processes of visual perception as a rational process of belief fixation. It follows that the inferences used in perception are no different from the inferences used in thought. That is, they are discursive inferences.

A short digression is needed here, however, lest we attribute to Clark intentions that he may not have. The previous analysis assumes the standard view of the brain as a physical machine that processes symbols in purely formal or syntactic way on the basis of the physical properties of the symbols; the brain performs digital computations. These symbols have meaning, of course, and so do the transformations of these symbols, but the processes in the brain are independent of any meaning. To put it differently, the brain is a syntactic machine that processes symbols that have meaning. The standard view can be modified by adding the thesis that digital computations are not merely formal syntactic manipulations but also involve semantics, that is, the contents of the states that participate in computations

are causally relevant in the production of the computations' outputs [26.5].

Although this is the standard, algorithmic, view of cognition, it is by no means unequivocally endorsed. There is another, competing view of cognition, according to which the brain is not a syntactic machine that processes symbols through algorithms. The brain represents information in a nonsymbolic, analogue-like form, as activation patterns across a number of units. Furthermore, the processes in the brain do not assume the form of algorithmic but of algebraic transformations; this is the connectionist view of cognition, of which Clark is a stern proponent. This is not the place to expand and explain connectionism, but I wish to stress that in this view of cognition, the brain does not use at all discursive inferences, although some of its behavior certainly simulates the usage of discursive inferences. If this is so, Clark's thesis that perception is inseparable from the rational processes of belief fixation does not commit him to the view that perception employs discursive inferences for the simple reason that thinking itself does not implicate such inferences.

Furthermore, given the propositional or symbolic form of the format in which the states of the visual system must be represented if vision is akin to thinking, the contents of these states, that is the information carried by the states, consists of concepts that roughly correspond to the symbols implicated; it is conceptual content. If vision is some sort of thinking, therefore, its contents must be conceptual contents. This means two things. Either the visual circuits store conceptual information that they use to process the incoming information, or they receive from the inception of their function such information from the cognitive areas of the brain while they are processing the information impinging on the retina. Spelke's principles that guide visual processing and render the percept possible are examples of conceptual content.

It should be noted that discursive inferences are distinguished from *inferences* as understood by vision scientists according to whom any transformation of signals carrying information according to some rule is an inference [26.6]:

> "Every system that makes an estimate about unobserved variables based on observed variables performs inference [...] We refer to such inference problems that involve choosing between distinct and mutually exclusive causal structures as causal inference."

One could claim, therefore, that although inferences, in this liberal sense, occur in the brain during visual perception, they are not like the inferences used in thought. One might even go further than that and claim that these inferences, or rather state transformations, do not involve representational states at all [26.7]. Although the percept is certainly a representational state, the processes that lead to its formation are not representations. It follows that visual perception is not a cognitive process, if *cognitive* is taken to entail the use of mental representations; "a system is cognitive because it issues mental representations" [26.7].

In this chapter, I examine vision and its processes and discuss the relation of vision with thinking. I do not have the space here to discuss the problem of whether visual processes involve representations. I proceed by assuming that they do although, first, as I will argue, the state transformations do not presuppose the application of inference rules that are represented in the system, and, second, not all visual states are representational.

In Sect. 26.1, in view of the close relationship between thinking and inference, I chart and briefly discuss inference and its modes, namely, deduction, induction, and abduction or inference to the best explanation.

In Sect. 26.2, I sketch an overview of the main conceptions concerning vision, to wit constructivism, direct or ecological theory of vision, and the more recent proposals that view vision as inseparable from action.

In Sect. 26.3, I present the two stages of which visual perception consists, namely early vision and late vision.

In Sect. 26.4, I discuss the problem of the cognitive penetrability (CP) of perception, because if vision is akin to thinking, visual processes necessarily involve concepts and are thus cognitively penetrated. If it turns out that some stage of vision is cognitively impenetrable (CI) and conceptually encapsulated, the status of the logical characterization of the visual processes of that stage remains open since, being nonconceptual in nature, they cannot be discursive inferences. I am going to argue that a stage of vision, early vision, is CI and has nonconceptual content. This content is probably iconic, analogue-like and not symbolic. By not being symbolic, the contents of the states of early vision cannot be transformed to some other contents by means of discursive inferences in so far as the latter operate on symbolic forms. The second main visual stage, namely late vision, is CP and implicates concepts. I also address in this section two problems with my claim that early vision is conceptually encapsulated. The first is raised by the existence of some general regularities that seem to guide the functioning of the perceptual system, of which the Spelke principles are a subset, and which operate at all levels of visual processing. The problem is, first, whether the existence of such principles entails that at least some part of the information

processed in early vision is inherently conceptual, and, second, whether the existence of such principles entails that vision in general is theory-laden. The second concerns the effects of perceptual learning, since one might argue that through perceptual learning some concepts are embedded in the perceptual circuits of early vision. If either of these two is correct, the states of early vision have conceptual contents and thus the processes of early vision may involve discursive inferences rendering early vision akin to thought and belief formation. I argue, however, that neither the principle nor the effects of perceptual learning entail that early vision has conceptual content.

In Sect. 26.5, I examine the logical status of the processes of early and late vision and argue that the processes of early vision are abductive nondiscursive inferences that do not involve any concepts, while the processes of late vision despite the fact that they are abductive inferences guided by concepts, are not discursive inferences either. I argue that the abductive inferences involved in visual perception are not sentential inferences but, instead, they rely on pattern-matching mechanisms that explore both iconic, analogue-like information and symbolic information. In this sense, visual abduction could be construed as consisting of a series of model-based inferences.

26.1 Inference and Its Modes

Let us grant that vision is like thinking and, therefore, involves discursive inferences. The question that arises concerns the nature of the inferences involved; are they deductive, inductive, or abductive? (Appendix 26.A for a definition of deductive, inductive, and abductive inference).

I think it is safe to assume that the whole visual process fits better the description of an abductive inference. The main reason for this thesis is that vision constructs a representation, (i. e., the percept) that best fits the visual scene. Specifically, given that the retinal image is sparse and thus underdetermines both the distal object and the percept, the visual system has to fill in the missing information to arrive at the best explanation, that is, the percept that best fits the retinal information. In essence, given the sparsity of the incoming information in the retinal image, the brain attempts to construct a representation that consists of the properties that an object should have in order to produce the specific retinal image. That is, the brain works back from the information that the retinal image contains to the object that could produce such a retinal image. Many objects could produce this image and the brain attempts to figure out which one of them best fits the retinal image. This is the trait par excellence of an abductive inference. Recent work (see [26.4] for an overview) suggests that this abductive inference or inference to the best explanation is a Bayesian inferences in which the brain constructs the percept that best explains the visual input by selecting the hypothesis that has the highest probability given the visual input.

It follows that the inference is ampliative, that is, the conclusion has a wider content than that of the premises and thus is not implicitly included in the premises; as such, the inference is not deductive. This is easy to grasp if we consider that the only information impinging on the retina consists of differences of light intensities and electromagnetic wavelengths. The percept that which the visual processes output (and since we have assumed that vision is a complex inference, the premises of the inference consist in the impinging information and the percept is the conclusion of this inference), however, is the object of our ordinary experience with its shape, size, color, motion, texture, etc. All these properties far exceed the impinging information concerning light intensities and wavelengths.

Moreover, and related to the first consideration, even if the premises of a visual inference that outputs the percept are correct, that is, even if the principles that guide perception reflect correctly the physical and geometrical regularities, and the impinging information being what it is, the percept may still not be a correct representation of the object in the environment that emanated the light rays and caused the perception. In other words, the conclusion may be wrong even though the premises are correct. This is why vision should be better understood as an abductive process or as an inference to the best explanation. Traditionally, abduction is thought as synonymous to the inference to the best explanation (for a recent reaffirmation see [26.8]). Recently, however, this thesis has come under attack mainly on the ground that abduction is for the generation of theories, whereas the inference to the best explanation is for their evaluation [26.9, 10]. Although I agree with Lipton, I will not dwell on this issue here any further. I will continue to use *abduction* as synonymous to *inference to the best explanation* because nothing important in the discussion in this chapter hinges on the outcome of this debate.

One may wonder why this ampliative, non-truth-preserving inference should be construed as an abductive inference and not as an inductive inference. One

might argue that all inductions are abductions or inferences to the best explanation [26.11]. Most authors, however, think that abduction is a subspecies of induction since it bears the basic marks of induction as it is ampliative and does not preserve truth. However, it is more specific than induction since it aims exclusively to pinpoint the cause or causes for some phenomena, that is, it aims to yield an explanation of a set of phenomena. Not all inductions are focused towards this aim. Several times a good induction leads to a generalization that subsumes a set of phenomena under the heading of a generalization, which, however, does not explain the phenomena. Consider the following induction.

Bird α is a crow and is black ($Ca\&Ba$)
Bird β is a crow and is black ($Cb\&Bb$)
...
Bird κ is a crow and is black ($CK\&Bk$)
Therefore (inductively)
All crows are probably black $((x)(Cx \rightarrow Bx))$

Under certain conditions this is a good induction in which from the colors of specific specimens of crows one infers the color of all crows. This is hardly a good explanation though. A good explanation seeks to explain, that is, make us or the scientific community understand why crows α and β are black. The generalization *All crows are probably black* fails to accomplish this since all that it does is gather together all instances of black crows in a generalization. Moreover, a good explanation of a set of phenomena is expected to have a wider range than these specific phenomena in the sense that it can be used as a springboard to explain a wider class of phenomena. In our case, a good explanation of why crows α and β are black should certainly involve genetics. Such an account not only would provide understanding of the correlation of crows with the color black, but it could also be used to explain the colors of other species. Now, it is widely agreed that the discovery of the relevant laws of genetics would fall within the purview of abduction. To put this point differently, all abductions are inductive inferences but not all inductions are abductions.

When I examine in Sects. 26.3 and 26.5 the visual processes in some detail, I shall adduce more evidence supporting the claim that visual processing is an abductive inference.

26.2 Theories of Vision

I have claimed that vision is a complex process that starts when light impinges on the retina and culminates with the formation of the percept, that is, the object of our ordinary experience and its properties. If vision as a whole is a complex process, it consists of a series of processes, or, in other words, in a series of state transformations in which one state containing some information is transformed via the visual mechanisms to a state containing some other sort of information. According to this view, vision is a process in which the visual system constructs the percept from the incoming visual information. All these processes take place within the visual system and although information from the other modalities and the actions of the viewer may either facilitate or inhibit the visual processing, vision in principle is autonomous from the other modalities and action.

This thesis can be assaulted from at least two fronts. The first is to deny that vision is a complex process involving information processing. It may be the direct retrieval of visual information from the environment without any need for mediating processes. The proponents of this view are divided into two camps. The first maintain that the retrieval of information from the environment is mediated by representations, while the second deny the necessity of invoking representations to explain how visual perception works. The second is to claim that although vision necessarily involves inferences, vision cannot be separated from action in that actions figure inherently and constitutively in vision. In this section, I present the three different conceptions of vision.

26.2.1 Constructivism

Visual perception begins with information impinging on the retina, this is the stimulation of the sensory organs, and culminates with the construction of the percept, which is a visual representation of the worldly objects (they are called *distal objects*) that emanate the light that stimulates the sensory organs. This is made possible through a series of transformations whereby the information impinging on the retina is progressively transformed into a final visual representation, the percept. The construction of the final visual representation is preceded by the construction of a host of intermediate visual representations of increasing complexity.

The transformation from one visual representation to the other, which are both mental representations being located in the brain, is effectuated through the

processes of vision that consist of the application of transformational rules that take as input representation $r1$ at time $t1$ and output representation $rt+1$ at time $t2$. These rules could be construed as abductive inferences since the brain is called upon to fill in the gaps in the information contained in the retinal image in order to construct a representation of the distal object that is the most likely candidate for being the object that could have produced the retinal image. It could be argued, hence, that the brain guesses which object is the best fit to explain the retinal image.

Since visual perception consists of a series of constructions of visual representations, vision is a constructive process. Let us call this construal of visual perception *constructivism*. According to one of the most influential visual scientists that espouse constructivism, *Marr* [26.12], there are three levels of representation. The initial level of representation involves Marr's *primal sketch*, which consists of the *raw primal sketch* and the *full primal sketch*. The *raw primal sketch* provides information about the edges and blobs present in a scene, their location and their orientation; this information is gathered by locating and coding individual intensity changes. Grouping procedures applied to the edge fragments formed in the *raw primal sketch* yield the *full primal sketch*, in which larger structures with boundaries and regions are recovered. Through the *primal sketch* contours and textures in an image are captured. The primal sketch can be thought of as a description of the image of a scene but not as a description of the real scene. This latter involves the relative distances of the objects and their motions. This information is provided by the viewer-centered representation, which is Marr's $2^{1/2}$ *sketch*. At this level information about the distance and layout of each surface is computed using various depth cues and by means of analysis of motion and of shading. This information describes only the parts of the object that are visible to the viewer and thus is relative to the viewer.

The computations leading to the formation of the $2^{1/2}$ *sketch* are determined by three factors:

1. The input to the visual system, that is, the optical array
2. The physiological mechanisms involved in vision, and the computations they allow, and
3. Certain principles that restrict and guide the computation.

These principles are constraints that the system must satisfy in processing the input. These constraints are needed because perception is underdetermined by any particular retinal image; the same retinal image could lead to distinct perceptions. Thus, unless the observer makes some assumptions about the physical world that give rise to the particular retinal image, perception is not feasible.

It is important at this juncture to stress that according to Marr, all the processes that lead to the formation of the $2\frac{1}{2}$D sketch are data-riven; they are driven solely by the input.

One of the aims of vision is the recognition of objects. This requires the matching of the shape of a structure with a particular object, a matching that requires an object-centered representation. This is Marr's *three-dimensional (3-D) model*. The recovery of the objects present in a scene cannot be purely data-driven, since what is regarded as an object depends on the subsequent usage of the information, and thus is task dependent and cognitively penetrable. Most computational theories of vision [26.12, 13] hold that object recognition is based on part decomposition, which is the first stage in forming a structural description of an object. It is doubtful, however, whether this decomposition can be determined by general principles reflecting the structure of the world alone, since the process appears to depend upon knowledge of specific objects [26.14]. Object recognition, which is a top-down process and requires knowledge about specific objects, is accomplished by the high-end vision. The construction of the percept, which is the end product of visual perception, therefore requires the synergy of both top-down and bottom-up transfer of information between the visual circuits and the cognitive centers of the brain. Object recognition requires matching the internal representation of an object stored in memory against the representation of an object generated from the image. In Marr's model of object recognition the 3-D model provides the representation extracted from the image that will be matched against the stored structural descriptions of objects (perceptual classification). (It should be emphasized that these *object recognition* units are not necessarily semantic, since we may recognize an object that we had seen before, even though we have no idea of its name, of what it does and how it functions, that is, even if we have no semantic and lexical information about it. Ref. [26.15] introduces a distinction between the *perceptual classification* and *semantic classification* and *naming*. These processes are independent one of the other.). See Appendix 26.B for an overview of constructivism.

Marr's and Biederman's hypothesis that object recognition occurs through part decomposition is based on the conception of three-dimensional objects as arrangements of some set of primitive 3-D shapes. According to Marr, these primitive 3-D shapes are generalized cylinders (Fig. 26.1) that are defined in terms of major axes and radii of objects.

According to Biederman, the primitive 3-D shapes are the so-called geons (Fig. 26.2). All objects can be decomposed into a set of 36 specific geons related in various ways. The properties that identify geons and allow them to function as volumetric perceptual primitives are viewpoint invariant, that is, they do not change as the angle of view changes. As such, they are called nonaccidental features since they are features not only of the image but also of the worldly objects (that is, they are properties that exist in the environment outside the viewer) that do not depend on what the viewpoint may be accidentally. Examples of nonaccidental properties are parallel lines and collinearity. If an object has parallel lines many rotations of this object yield an image in which these lines are still nearly parallel; that is to say, parallelism is a property that is rotation- or perspective-invariant.

Let me close the account of constructivism by reminding the reader that the theories of visual perception presented in this part of the chapter are some among the many different theoretical accounts of visual processing. The differences between the various theories notwithstanding, all constructivist theories share a common core, namely that visual perception involves state transformations in the course of which visual representations of increasing complexity are being gradually constructed by the visual system. The visual processes start from the meager information contained in the retinal image and which consists of local distributions of light intensities and wavelengths. These transformations can also be construed as computations in which the brain computes an output state given an input state. Many of these transformations (but not all of them) act on and therefore essentially involve mental representations that are within the brain of the viewer, and can be independent of any other activities on the part of the viewer. The transformations are made possible through the application of transformational rules, such as, for example, the rule that abrupt changes in light intensity signify the presence of edges that is used by the perceptual system to construct the raw primal sketch. Such a rule takes as input states that carry information about various light intensities distributed in space and delivers states that carry information about edges. It follows that the transformations taking place in visual processing are information-processing operations. (I said that not all of the transformations operate on representations because many of these transformations operate on states that are not representational. It would require another chapter to discuss the conditions under which a state is representational or not and, of course, much depends on how one defines the term *representation*. I confine myself to pointing out

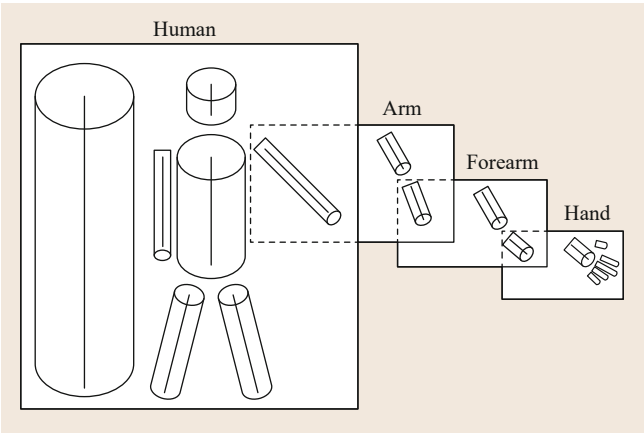

Fig. 26.1 Marr's generalized cylinders

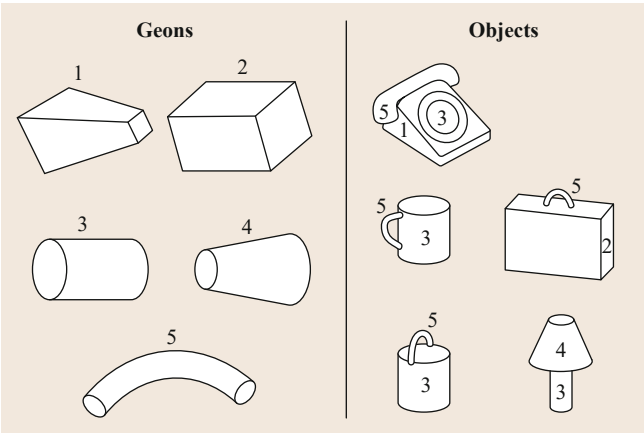

Fig. 26.2 Biederman's geons

that many of the earlier visual states are probably not representational because they do not meet the criteria that an adequate definition of representation posits, and to referring the reader to the discussion in Chap. 4. As we shall see in Sect. 26.4, one could claim that there is a sharp distinction between internal probabilistic dependencies between states that can be explained by internal causal connections between the circuits of the brain and those that cannot; only those that cannot be explained internally carry information about the external world and thus involve representational states.)

The fact that the visual brain transforms states to other states through the usage of some rules means that the function of the brain can be understood as a series of inferences from some state/premises to some other states/conclusion. In view of our discussion in the beginning of this section, as well as in the previous one, the inferences most likely are abductive in nature.

26.2.2 Theory of Direct Vision or Ecological Theory of Visual Perception

Gibson [26.16] started from a very different assumption than that of constructivism. In contradistinction to the latter, Gibson argued that perception begins not with the informationally sparse retinal image but with the informationally rich *optic array*. The spatial pattern of light intensities and the mixture of wavelengths that impinge on the receptors of the retina form the optic array. This light, however, carries a lot of information about the solid objects in the environment (the distal objects) because the intensities of light and its wavelengths vary from one solid visual angle to another (as the rays of light emanating from solid objects travel in space and between the surfaces of the objects that fill the space, given that at any point in space light converges from all directions, at each point in space there is a nested set of solid visual angles). As a result, the optical array is determined by, and therefore carries information about, the nature and location of the three-dimensional worldly surfaces from which it is being reflected.

Unlike the retinal image, the optic array is an external source of information, or, better, an external information-bearing structure since it exists outside the viewer, is independent of the constitution of the brain of the viewer, and carries information about the environment. Gibson's central claim is that the information contained in the optic array suffices to allow perceivers to specify the visual scene that causes the optic array, that is, to specify the solid surfaces that surround them, and to use the information included in the optic array to interact with their environment.

When perceivers move in their environment, moreover, the entire optic array is transformed to reflect the new environment since it depends exclusively on it. As perceivers move around, they sample different optic arrays and therefore receive a variety of information about the changing environment, since the transformations of the optic array as perceivers move contain information about the layout of the objects in the environment as well. As in realistic situations perceivers are not static, motion enriches the visual information that the perceivers receive from the environment enabling them to recover the visual scene much easier than if they were static. Furthermore, this motion by effecting transformations of the optic array allows the perceivers to identify those aspects of objects that remain invariant under movement (the nonaccidental properties that we have discussed). It goes without saying that this information is made available only to perceivers that move in their environment and effect a change in the optic array that they receive from the environment; a static perceiver would never be able to detect the properties of objects that remain invariant under motion. Note that information about the invariant properties is available in the environment, but viewers can retrieve or detect it only as they move. This entails that perception becomes entangled with action, since moving around is a form of action.

The richer the information that the light impinging on the retina carries, the less information processing the visual brain is required to do in order to form the percept. Taking this view to its extreme end, one might claim that if the optic array suffices all by itself to enable viewers to recover the visual scene, there is no need to posit any internal information processing on information-bearing states. Visual perception involves no information processing and no inferences of any sort; it just recovers the visual scene directly from the information contained in the optic array (which explains coining this theory a theory of direct vision). This interpretation of the theory received a devastating criticism in *Fodor* and *Pylyshyn*'s [26.17] paper entitled *How Direct is Visual Perception*. I think it safe to assume that the radical interpretation that excludes information processing from visual perception has not recovered from this critique since most of the counterarguments raised in that paper have not been adequately answered. Whether, however, Gibson subscribed to this radical view is debatable. Be that as it may, the radical interpretation is not the only possible interpretation of direct vision.

The fact that the input to the visual system may contain more information than that envisaged by constructivism does not entail that visual perception does not involve any internal information processing. It only entails that the internal information processing needed for the formation of the percept is less than in constructivist theories since a part of it is being replaced by the manipulation through motion and transformation of the optic array, which as you recall is an external information-bearing structure. Nor does the richness of the information in the input entail that no representations are needed; it entails that visual perception allows positing less representations than those required in constructivist theories. As *Rowlands* [26.18] remarks:

> "Here is nothing in Gibson's theory itself – as opposed, perhaps, to his statements about his theory – that entails or even suggests that all of the role traditionally assigned to manipulation and transformation of internal information-bearing structures can be taken over by the manipulation and transformation of external information-bearing structures."

In this moderate interpretation of Gibson's theory of direct vision, the need for some information processing over internal representational states still persists, except

that, in view of the fact that the information contained in the visual input is richer than previously thought, this need is attenuated. Therefore, visual perception involves some sort of inferences.

Gibson's theory was coined the *theory of direct perception* because it relinquished the need for internal information processing; instead, the viewers retrieve all the information they need to detect the environment directly from the environment without any internal processing of any sort mediating the process of information retrieval. If, however, some information processing over internal representations is needed as well, as a moderate form of Gibson's theory asserts, can the qualification *direct* be salvaged?

There is a sense in which it might. Suppose that *direct* is construed so as to emphasize not the lack of information processing operating on internal representations, but the fact that the information processing is entirely data-driven, that is, guided by environmental input and some principles that reflect regularities in the environment, and the whole process is not influenced by other internal nonvisual states of the viewer, such as the viewer's cognitive or emotional states. If this supposition is borne out, then visual perception is direct in the sense that the whole process is data-driven and, as such, the information processing used operates over information retrieved exclusively from the environment. Note that this presupposes that the principles guiding visual processing do not constitute some form of intervention on the part of the viewer whose contribution exceeds what is given in the environment.

This assumption is borne out if visual perception or at least some stage of it, is purely data-driven, that is, cognitively and emotionally impenetrable. If cognitive states penetrate and thus influence perceptual processing, the viewer's cognitive states actively contribute to the formation of the percept and the visual processing does not retrieve information directly from the environment but only through some cognitive intervention; visual perception, in this case, is not direct. *Norman* [26.19] has argued along this line that the processing along the dorsal visual pathway that guides our on-line interactions with the environment, owing to the fact that when it operates immediately on the visual input it is entirely data-driven, is a visual function that conforms very closely to Gibson's *direct* theory. The ventral visual pathway, in contradistinction that is responsible for object recognition and categorization is clearly affected by cognition and, in this sense, is not a *direct* visual function. Since both visual pathways are found in the brain, the constructivist and the ecological theories of perception can be reconciled.

Even though it seems abundantly clear that visual perception requires a significant amount of information processing, and in this sense one of Gibson's main insights is considered to be wrong, several of Gibson's insights have been incorporated in the constructivist information-processing research program. For example, most, if not all, information-processing theories hold that most of the ambiguities that occur during the information processing of the retinal input cannot be resolved by that input alone and need top-down assistance only when information comes from a static monocular image. When additional information can be derived from stereopsis and motion of real scenes, then the information-processing program can resolve the ambiguity without the need of a top-down flow of information. If one takes into account the real input to human vision, which is binocular and dynamic, there are few ambiguities that cannot be resolved through a full consideration of the products of the early visual processing *modules* [26.20]. This shows that the dynamic and interactive character of vision solves several problems encountered within the information-processing research program.

Our discussion about direct vision revealed an aspect of visual processing that traditional constructivist theories did not initially consider, namely the interaction of perception and action. The next kind of theory of perceptual processing that we will examine views visual perception as inextricably linked with action and uses the most recent neuropsychological evidence, vision science research, and computer modeling to both substantiate this claim, and draw the details of how the active visual brain works in order to provide a fully fledged unifying model of perception and action. Although this model aims to cover all modalities, for the purpose of this chapter I will restrict the presentation and discussion to visual perception.

26.2.3 Predictive Visual Brain: Vision and Action

The basic tenet of the theory of ecological or direct vision is that all the information viewers need to recover the visual scene that causes the retinal image is already included in the incoming information in the optic array. Little or no information processing is required for the construction of the percept. The constructivist theories of visual perception, in contradistinction, underline the necessity of information processing and state transformations in the brain. The flow of information in the brain is bidirectional; both top-down and bottom-up signals are transmitted and the ensuing percept is the result of the synergy between top-down and bottom-up processing. This class of theories assumes that the representation constructed at some level is transmitted bottom-up to the neuronal assembly at the next

immediate level where it is further processed. Moreover, recurrent signals return top-down to earlier levels mainly to test hypotheses concerning aspects of the visual scene (recall that visual perception aims to recover the visual scene that causes the retinal image and does that by constructing increasingly more complex representations of the probable aspects of the visual scene at various spatial and temporal scales) until the percept is constructed.

Recent empirical findings and modeling shed light on the way the brain actually effectuates these processes. These details, as we shall see, entail certain deviations from the traditional constructivism image, which concern (a) the sort of information transmitted bottom-up; only prediction errors are transmitted to the next level, (b) the nature of the representations constructed; they are distributions of probabilities rather than having a unique value (note that this new approach emphasizes the indispensable role of representations in visual processing), and (c) the interaction between perception and cognition. This last trait is very important has important repercussions for our discussion on the relation between visual processing and thinking.

According to this view of visual perception, brains are predictive machines [26.4]:

> "They are bundles of cells that support perception and action by constantly attempting to match incoming sensory inputs with top-down expectations or predictions. This is achieved using a hierarchical generative model that aims to minimize prediction error within a bidirectional cascade of cortical processing."

A hierarchical generative model as applied to visual processing is a model of perceptual processes according to which the brain uses top-down flow of information (enabled by top-down neural connections) in an attempt to generate a visual (meaning, in the brain) representation of the visual scene (in the environment) that causes the light pattern impinging on the transducers and the low-level visual responses to this light pattern. The brain attempts to recover gradually the causal matrix (the various aspects of a visual scene) that causes and thus is responsible, for the retinal image seen as a data-structure (i. e., the sensory data). The brain does that by capturing the statistical structure of the sensory data, that is, by discovering the deep regularities underlying the retinal structure, on the very plausible assumption that the deep structure underneath the sensory data reflects, so to speak, the causal structure of the visual scene.

Hierarchical generative models attempt to achieve this by constructing, at each level, hypotheses about the probable cause of the information represented in the immediately previous level, and testing these hypotheses by matching their predictions with the actual sensory data at the preceding processing level. Suppose, for example that a neuronal assembly at level l receives from level l-1 information concerning differences in light intensities. The higher level attempts to recover the probable edges that cause the variation in light intensity and forms a hypothesis involving such edges. Now, and this is the crucial part, if this hypothesis were correct, that is, if the edges as represented in the hypothesis were present in the environment, then a certain pattern of variation of light intensities at the appropriate local scale would have been present in the sensory data. This prediction is transmitted top-down to level l-1 and matched against the actual pattern of variations in light intensities. If there is a match (with an acceptable degree of error deviation due to the inherent noise of the signal, of course) no further action is needed since the perceptual system *assumes* that it has constructed the correct, at this spatial scale, representation of the relevant environmental input. If the match reveals a discrepancy, that is, if an error in the prediction is detected, this prediction error is transmitted bottom-up to level l so that a new hypothesis be formulated and tested until, eventually, no unacceptable prediction error persists. If one thinks of the discovered error as a surprise for the system, the system strives to correct its hypotheses so that by making correct predictions, the testing of the hypotheses yields no surprises; this is a typical error-driven learning process where a system learns, i. e., constructs a correct representation, by gradually reducing error. The hierarchical generative models hence generate, in essence, low-level states (the predictions they make about the activities at the lower levels) from high-level causes (the hypotheses that would, if correct, explain the activity at the lower levels).

Bidirectional hierarchical structure allows the system to [26.4]:

> "infer its own priors (the prior beliefs essential to the predicting routines) as it goes along. It does this by using its best current model at one level as the source of the priors for the level below, engaging in a process of *iterative estimation* that allows priors and models to coevolve across multiple linked layers of processing so as to account for the sensory data."

To form hypotheses concerning the probable cause of the sensory data at a certain level, at a specific spatial and temporal scale, the neuronal assembly at the next level, say level l, uses information not only about the sensory data at the previous level (or, to be precise, information regarding its prediction error) that is transmitted bottom-up, but also higher-level information that

is transmitted to *l* either laterally, that is, from neuronal assemblies at the same level (neurons in V1 processing wavelengths inform other neurons in V1 processing shape information, for example), or top-down from levels higher in the hierarchy (neurons in V4, for instance, are informed about the color of incoming information from neurons in the inferotemporal cortex in the brain (IT) as a result of precueing – that is, when a viewer has been informed about the color of an object that *will* appear on a screen). This higher-level information may and usually does concern general aspects of the world (such as *solid objects do not penetrate each other*, or *solid objects do not occupy exactly the same space at the same time*, etc.), and may also reflect knowledge about specific objects learned through experience. All this lateral and top-down flow of information provides the context in which each neuronal assembly constructs the most probable hypothesis that would explain the sensory data at the lower level. Thus, context-sensitivity is a fundamental and pervasive trait of the processing of hierarchical predictive coding; the contextualized information significantly affects, and on occasions (as in hallucinations) may override, the information carried by the input.

The hierarchical predictive processing model can be naturally extended to include action and thus closely ties perception with action [26.4]. This is the action-oriented predictive processing. Action-oriented predictive processing extends the standard hierarchical predictive model by suggesting that motor intentions actively elicit, via the motor actions they induce, the ongoing streams of sensory data at various levels that our brains predict. In other words, once a prediction is made about the state in the world that causes the transduced information, the action-oriented predictive processes engage in a search in the environment of the appropriate worldly state. Suppose, for example, that owing to bad illumination conditions, a perceiver is unsure about the identity of an object in view. Its brain makes a prediction about the putative object that causes the sensory data the perceiver receives, and the perceiver moves around the object in order to acquire a better view that will confirm the prediction. By moving around, the perceiver's expectations about the proprioceptive consequences of moving and acting directly cause the moving and acting since where and when the perceiver moves is guided by the aim that the perceiver's action brings the object into a better view.

It is worth pausing at this point to discuss briefly the problem of nature of the relation between visual perception and action and, specifically, motion. Is this relation constitutional, which means that if someone cannot or does not move they cannot visually perceive anything? This claim was initially made by *Noe* although, in view of vehement criticism, Noe has attempted to modify it without compromising the main tenets of his views [26.21]:

> "When you experience an object as cubical merely on the basis of its aspect, you do so because you bring to bear, in this experience, your sensorimotor knowledge of the relation between changes in cube aspects and movement. *To experience a figure as a cube*, on the basis of how it looks, *is to understand how it looks changes as you move* (emphasis added)."

The sensorimotor knowledge consists of the expectations of how our perception of an object changes as we move around it, or as this object moves with respect to us. These expectations constitute a form of practical knowledge, a *knowing how* as opposed to a *knowing that*. Thus, to be able to experience visually an object, one needs to have the ability to move around the object and explore it. Visually experiencing the object literally consists of grasping the relevant sensorimotor contingencies, that is, the sensorimotor knowledge associated with this specific object. There are two ways to read this claim. According to the first reading, which Noe seems to espouse judging from the previously cited passage, to be able to visually perceive requires the actual exercise of the ability to probe the world. According to the second reading, visually perceiving an object only requires the ability to probe the world but not the actual exercise of this ability. The first reading entails immediately that prior to exercising this ability, one does not visually perceive the object. Since this is absurd, one has to concede that viewers do not need to exercise actually the ability to probe the environment, it suffices that they take recourse to their experience with similar objects and retrieve the requisite sensorimotor contingencies from experience. Even if one takes this line, however, the problem remains that at the time of a first encounter with an object to be able to see its, say, shape, one should be able to probe the object either by moving around the object, or by having the object move around them. Thus, when stationary viewers perceive a stationary novel object, lacking any knowledge of sensorimotor contingencies, they do not see its shape or its other properties.

It follows that infants upon opening their eyes for the first time and facing the world, by lacking any sensorimotor knowledge and by not probing the environment, they do not see anything. This claim flies to face of countless empirical evidence, which shows that there is something fundamentally wrong with equating visual perception with understanding sensorimotor contingencies and deploying the relevant practical knowledge. This entails, in turn, that the relation between visual

perception and action, no matter how important it is, is not a constitutional relation; one gets to see the world even if both they and the world are stationary, although it goes without saying that their experience will be restricted compared to other viewers who can probe the environment. They could not visually experience, for example, Marr's 3-D sketch because they lack knowledge of the unseen surfaces of objects.

This unity between perception and action emerges most clearly in the context of *active inference*, where the agent moves its sensors in ways that amount to actively seeking or generating the sensory consequences that their brains expect. "Perception, cognition, and action work closely together to minimize sensory prediction errors by selectively sampling, and actively sculpting, the stimulus array" [26.4]. Their synergy moves a perceiver in ways that fulfill a set of expectations that constantly change in space and time. Accordingly, perceptual inference is necessary to induce prior expectations about how the sensorium unfolds and action is engaged to resample the world to fulfill these expectations.

Since the construction of the representations of the putative causes of the sensory inputs is made possible through a synergy of bottom-up processing transmitting the prediction errors and top-down processing transmitting for testing the hypotheses concerning the probable causes of the input and in so far as the processes constructing these hypotheses are informed by high-level knowledge of the sort discussed above, visual perception unifies cognition and thinking with sensation; these two become intertwined. This means that perception inextricably involves thinking. Notice that this account of visual perception necessarily involves representations; it requires that each level retain a representation of the data represented at this level so that the top-down transmitted predictions of the hypotheses formed at subsequent higher levels be matched against the information represented at the lower level in order for the hypothesis to be tested. It also requires the representation of the putative causes of the sensory data at the preceding level; these are called the representation-units, which operate along the error units (the units that compute the error signal, that is, the discrepancy between prediction and actual data) in a hierarchical generative system.

Furthermore, testing hypotheses and altering them as a result of any prediction errors until the prediction error is minimized and thus until the most probable cause of the sensory data has been discovered, is an inference. Being a probabilistic inference that aims to discover the most probable hypothesis that explains away a set of data, it is most likely a Bayesian inference. It is very plausible, therefore, that the computational framework of hierarchical predictive processing realizes a Bayesian inferential strategy (see Appendix 26.C for an analysis of Bayes' theorem). Indeed, recent work on Bayesian causal networks [26.22] presents the brain as a Bayesian net operating at various space and time scales.

What Bayes' theorem, on which this strategy is based, ensures is that a hypothesis is eventually selected that makes the best prediction about the sensory data minimizing thereby the prediction error and thus best explains them away; that is a hypothesis that by having the highest posterior probability provides the best fit for the sensory data. The construction of this hypothesis crucially and necessarily involves the context, as it is clearly expressed in Bayes' equation in the form of the prior probability for the hypothesis $P(A)$, whose value depends on the context. That is to say, it is the context that provides the initial plausibility of a hypothesis before the hypothesis is tested.

This enables *Clark* [26.4] to claim that in the framework of predictive brains that use hierarchical generative processing perception becomes theory-laden in the specific sense that what viewers perceive depends crucially on the set of priors (that is, the hypotheses that guide the predictions about the matrix of the sensory data at the lower processing levels, which the hypothesis projects) that the brain brings to bear in its attempt to predict the current sensory data. This remark brings us back to the main theme of this chapter, namely, the relation between perceiving and thinking. If thinking necessarily implicates discursive inferences and deploying concepts, as it usually does, Clark's claims entail that perception employs from its onset concepts and draws discursive inferences. To assess this dual claim, we must examine the processes of vision to determine first whether concepts are used and if the answer is affirmative the extent to which they are being used, and second, whether the inferences that are undoubtedly used in perception must necessarily be discursive. I hasten to note that, with respect to this last problem, nowhere in his account does Clark suggest that the inferences must be discursive. In fact, the sources he refers to, especially those concerning connectionist neural networks, suggest that the inferences on which perception relies may take another form and need not necessarily involve propositionally structured premises and conclusions.

26.3 Stages of Visual Processing

I said above that we must examine the processes of vision with a view to determine whether and, depending on the answer to this question, the extent to which, cognition penetrates visual perception in the sense that perceptual processing uses conceptual information that is either transmitted top-down to perceptual circuits, or is inherently embedded in visual circuits. In the literature, visual processing is divided into two main stages, to wit, early vision and late vision.

26.3.1 Early Vision

Early vision is a term used to denote the part of perceptual processing that is preattentive, where attention means top-down, cognitively driven attention. *Lamme* [26.23, 24] argues for two kinds of processing that take place in the brain, the feedforward sweep (FFS) and recurrent processes (RP). In the FFS, the signal is transmitted only from the lower (hierarchical) or peripheral (structural) levels of the brain to the higher or more central ones. There is no feedback; no signal can be transmitted top-down as in RP. Feedforward connections in conjunction with lateral modulation and recurrent feedback that occurs and is restricted within the early perceptual areas (local recurrent processing – LRP) extract high-level information that is sufficient to lead to some initial categorization of the visual scene and selective behavioral responses.

When a visual scene is being presented, the feedforward sweep reaches *V1* in about 40 ms. Multiple stimuli are all represented at this stage. The unconscious FFS extracts high-level information that could lead to categorization, and results in some initial feature detection. LRP produces further binding and segregation. The representations formed at this stage are restricted to information about spatiotemporal and surface properties (color, texture, orientation, motion, and perhaps to the affordances of objects), in addition to the representations of objects as bounded, solid entities that persist in space and time. (*Affordances* is the term *Gibson* [26.16] used to refer to the functional properties of objects (an object affords eating to an organism, grasping to an organism, etc.). *Clark* [26.25] defines *affordance* as "the possibilities for use, intervention and action which the physical world offers a given agent and are determined by the *fit* between the agent's physical structure, capacities and skills and the action-related properties of the environment itself". Affordances are directly perceivable by an organism in the sense that an object does not have to be classified as a member of a certain category in order for the organism to draw the conclusion, or use the relevant knowledge, that this object can be used in a certain way by the organism; the organism just perceives the affordance, that is, the opportunity of action on this specific object. Affordances have two important properties. First, they are determined by the functional form of an object, that is, a combination of the object's visible properties should suffice to determine whether this object has an affordance relative to some viewer. Affordances are based on certain invariant characteristics of the environment. Second, the affordance is always relative to the viewing organism; this is a consequence of the fact that affordances provide organisms with the opportunity to interact with objects in their environment. This interaction depends on the objects' properties but it also depends on the needs and the constitution of the organism. A fly, for instance, affords eating to a frog but not to a human.)

At this level there are nonattentional selective mechanisms that prevent many stimuli from reaching awareness, even when attended to. Such stimuli are the high temporal and spatial frequencies, physical wavelength (instead of color), crowded or masked stimuli and so forth. FFS results in some initial feature detection. Then this information is fed forward to the extrastriate areas. When it reaches area V_4 recurrent processing occurs. Horizontal and recurrent processing allows interaction between the distributed information along the visual stream. At this stage, features start to bind and an initial coherent perceptual interpretation of the scene is provided. Initially, RP is limited to within visual areas; it is local. At this level one can be phenomenally aware of the content of perceptual states. At these intermediate levels there is already some competition between multiple stimuli, especially between close-by stimuli. The receptive fields that get larger and larger going upstream in the visual cortical cannot process all stimuli in full and crowding phenomena occur. Attentional selection intervenes to resolve this competition. Signals from higher cognitive centers and output areas intervene to modulate processing; this is global RP and signifies the inception of late vision.

Lamme [26.23, 24] discusses the nature of information that has achieved local recurrent embedding. He suggests that local RP may be the neural correlate of binding or perceptual organization. However, it is not clear whether at this preattentional stage the binding problem has been solved. The binding of some features, such as its color and shape, may require attention, while other feature combinations are detected preattentively. So, before attention has been allocated, the percept consists of only tentatively but uniquely bound features that form the proto-objects [26.26]. *Lamme* [26.24] argues that Marr's $2\frac{1}{2}$D surface representation of objects and

their surface properties are extracted during the local RP stage. Other research [26.27] suggests that spatial relations are extracted at this recurrent stage. In addition motion and size are represented in cortical areas in which local RP take place.

It should be added that Marr thought of the $2\frac{1}{2}$D sketch as the final product of a cognitively unaffected stage of visual processing, since, as we have seen, the formation of the 3-D sketch relies on semantic, conceptual knowledge. If, as is usually thought, cognitive effects on perception are mediated by cognitively-driven top-down attention, Lamme's proposal that early vision is not affected by this sort of attention echoes Marr's view that early vision is not affected by cognition and is thus CI, a view shared by *Pylyshyn* [26.28].

Current research (see [26.4] for a discussion) sheds light on the nature of inferences involved in the hypothesis testing implicated in early vision. Specifically, the top-down and lateral effects within early vision aim to test hypotheses concerning the putative distal causes of the sensory data encoded in the lower neuronal assemblies in the visual processing hierarchy. This testing assumes the form of matching predictions made on the basis of this hypothesis about the sensory information that the lower levels should encode assuming that the hypothesis is correct, with the current, actual sensory information encoded at the lower levels. Eventually, the hypothesis that best matches the sensory data is selected and the whole process of hypothesis selection can be construed as an abductive inference or inference to the best explanation, which could very well be carried through by Bayesian nets. One should note that this account of early vision shows that the standard constructivist theories of visual processing can be reconciled and greatly benefit from the recent conceptions of the brain as a generative, predictive machine.

There seems to be, however, a crucial discrepancy between the account of early vision presented here and Clark's account of generative hierarchical predictive models. It concerns the role of context, or previously acquired knowledge, in the formation of the working hypotheses and its direct consequence that because of this trait, visual perception and discursive thinking are inseparable. If early vision is restricted to processes occurring within the visual cortex and excludes any cognitive influences, then first, previous knowledge seems to play no role in the formation of the working hypotheses, and second, early vision does not involve any thinking since the latter requires the participation of the cognitive centers of the brain. Moreover, the representations in early vision are analogue-like, iconic and not symbolic and this entails that early vision cannot be some sort of discursive thinking since the latter operates on symbolic forms.

With respect to the first point, there is actually no real discrepancy. Recall that lateral and local recurrent processes play a fundamental role in the formation of the hypotheses that are constructed in early vision. Moreover, as we shall see in the next section, all visual processes including those of early vision, are restricted by certain principles, or better constraints, that reflect general regularities about the world and its geometry. Now, one could say that these constraints constitute a body of knowledge that informs early vision processing and affects early vision from the within and not in a top-down manner, since as we saw there are no cognitive top-down effects in early vision. This as we shall see, however, is misleading because these constraints do not constitute some form of knowledge that by affecting early vision renders it theory-laden, as Clark claims. Finally, early vision is also affected by associations of object properties that reflect statistical regularities in the environment and are stored in the early visual circuits through perceptual learning. I argue in the next section that these associations do not constitute a body of knowledge that affects early vision rendering it theory-laden. The lateral and local recurrent processes, the constraints, and the associations built in the early visual circuits constitute a rich context that contributes significantly to the formation of the working hypotheses that early vision neuronal assemblies construct to explain the sensory data at the lower processing levels. This context, however, does not involve any body of knowledge that renders perception theory-laden, as theories are traditionally understood.

As far as the second point is concerned, there is indeed a discrepancy because the account of early vision and Clark's views. Early vision, by being CI and conceptually encapsulated does not involve thinking and is radically different from thinking. In fact, as I argue in Sect. 26.5, not even late vision that involves concepts and is affected by the viewers' knowledge about the world is like thinking.

26.3.2 Late Vision

The conceptually modulated stage of visual processing is called late vision. Starting at 150–200 ms, signals from higher executive centers including mnemonic circuits intervene and modulate perceptual processing in the visual cortex and this signals the onset of global recurrent processing (GRP). In 50 ms low spatial frequency (LSF) information reaches the IT and in 100 ms high spatial frequency (HSF) information reaches the same area. (LSF signals precede LSF signals. LSF information is transmitted through fast magnocellular pathways, while HSF information is transmitted through slower parvocellular pathways.)

Within 130 ms, parietal areas in the dorsal system but also areas in the ventral pathway (IT cortex) semantically process the LSF information and determine the gist of the scene based on stored knowledge that generates predictions about the most likely interpretation of the input. This information reenters the extrastriate visual areas and modulates (at about 150 ms) perceptual processing facilitating the analysis of HSF, for example by specifying certain cues in the image that might facilitate target identification [26.29–31]. Determining the gist may speed up the FFS of HSF by allowing faster processing of the pertinent cues, using top-down connections to preset neurons coding these cues at various levels of the visual pathway [26.32].

At about 150 ms, specific hypotheses regarding the identity of the object(s) in the scene are formed using HSF information in the visual brain and information from visual working memory (WM). The hypothesis is tested against the detailed iconic information stored in early visual circuits including V1. This testing requires that top-down signals reenter the early visual areas of the brain, and mainly V1. Indeed, evidence shows that V1 is reentered by signals from higher cognitive centers mediated by the effects of object- or feature-centered attention at 235 ms post-stimulus [26.33, 34]. This leads to the recognition of the object(s) in the visual scene. This occurs, as signaled by the P3 event-related-potentials (ERP) waveform, at about 300 ms in the IT cortex, whose neurons contribute to the integration of LSF and HSF information. (The P3 waveform is elicited about 250–600 ms and is generated in many areas in the brain and is associated with cognitive processing and the subjects' reports. P3 may signify the consolidation of the representation of the object(s) in working memory.)

A detailed analysis of the form that the hypothesis testing might take is provided by *Kosslyn* [26.35]. Note that one need not subscribe to some of the assumptions presupposed by Kosslyn's account, but these disagreements do not undermine the framework. Suppose that one sees an object. A retinotopic image is formed in the visual buffer, which is a set of visual areas in the occipital lobe that is organized retinotopically. An attentional window selects the input from a contiguous set of points for detailed processing. This is allowed by the spatial organization of the visual buffer. The information included in the attention window is sent to the dorsal and ventral system where different features of the image are processed. The ventral system retrieves the features of the object, whereas the dorsal system retrieves information about the location, orientation, and size of the object. Eventually, the shape, the color, and the texture of the object are registered in anterior portions of the ventral pathway. This information is transmitted to the pattern activation subsystems in the IT cortex where the image is matched against representations stored there, and the compressed image representation of the object is thereby activated. This representation (which is a hypothesis regarding the identity, that is, class membership of an object) provides imagery feedback to the visual buffer where it is matched against the input image to test the hypothesis against the fine pictorial details registered in the retinotopical areas of the visual buffer. If the match is satisfactory, the category pattern activation subsystem sends the relevant pattern code to associative or WM, where the object is tentatively identified with the help of information arriving at the WM through the dorsal system (information about size, location, and orientation). Occasionally the match in the pattern activation subsystems is enough to select the appropriate representation in WM. On other occasions, the input to the ventral system does not match well a visual memory in the pattern activation subsystems. Then, a hypothesis is formed in WM. This hypothesis is tested with the help of other subsystems (including cognitive ones) that access representations of such objects and highlight their more distinctive feature. The information gathered shifts attention to a location in the image where an informative characteristic or an object's distinctive feature can be found, and the pattern code for it is sent to the patternactivation subsystem and to the visual buffer where a second cycle of matching commences.

Thus, the processes of late vision rely on recurrent interactions with areas outside the visual stream. This set of interactions is called *global recurrent processing* (GRP). In GRP, standing knowledge, i. e., information stored in the synaptic weights is activated and modulates visual processing that up to that point was conceptually encapsulated. During GRP the conceptualization of perception starts and the states formed have partly conceptual and eventually propositional contents. Thus, late vision involves a synergy of perceptual bottom-up processing and top-down processing, where knowledge from past experiences guides the formation of hypotheses about the identity of objects. This is the stage where the 3-D sketch (that is, the representation of an object as a volumetric structure independently of the viewer's perspective) is formed. This recovery cannot be purely data-driven since what is regarded as an object depends on the subsequent usage of the information and thus depends on the knowledge about objects. Seeing 3-D sketches is an instance of amodal completion, i. e., the representation of object parts that are not visible from the viewer's standpoint. (Amodal completion is the perception of the whole of an object or surface when only parts of it affect the sensory receptors. An object will be perceived as a complete volumetric structure even if

only part of it, namely, its facing surface, projects to the retina and thus is viewed by the viewer; it is perceived as possessing internal volume and hidden rear surfaces despite the fact that only some of its surfaces are exposed to view. Whether this perception involves visual awareness, in which case the brain completes the missing features through mental imagery, or visual understanding only, which means that the hidden features are not present in the phenomenology of the visual scene but are thought of, is a matter of debate.). In amodal completion, one does not have a perceptual impression of the object's hidden features since the perceptual system does not fill in the missing features as happens in modal perception; the hidden features are not perceptually occurrent (see Appendix 26.D for a discussion of modal and amodal perception or completion).

One readily notices that Kosslyn's account of hypothesis testing naturally fits the schema of hierarchical generative predictive models as discussed in *Clark* [26.4]. The main themes of this schema are present in Kosslyn's account. These are: the generation of hypotheses at a higher level of visual processing, the crucial role of context or previously acquired knowledge in the formation of these hypotheses, and the testing of these hypotheses through their predictions against the rich iconic information stored in the lower levels in the visual hierarchy. The whole process fits the scheme of an abductive inference or inference to the best explanation that could be carried out by means of Bayesian networks.

There is a marked difference between the abductive inferences involved in early vision and those involved in late vision; the latter but not the former are informed by knowledge properly speaking, that is, by information that is articulated in thought and thus contains concepts. This might tempt one to think that late vision may be akin to thought and thus that there is a stage of visual processing that has the most crucial traits of thinking, i. e., it involves discursive inferences justifying thus in part Clark's, Spelke's and others' belief to that effect. Against this, I am going to argue in Sect. 26.5, that late vision despite its being informed by conceptually articulated knowledge, differs in significant ways from thinking, the most important difference being that late vision does not engage in discursive inferences.

I have claimed that late vision constructs gradually a representation that best matches the visual scene through a set of processes that test a series of hypotheses by matching these hypotheses against stored iconic information. In other words, the output of late vision, a recognitional belief, is the result of an abductive inference.

26.4 Cognitive Penetrability of Perception and the Relation Between Early Vision and Thinking

In assessing claims relating perception to thinking and cognition, it is of paramount importance to examine the role that concepts play in modulating perceptual processing. This is so because if the processes of visual perception are the same as those that lead to belief formation, which means that perception and thinking are of the same nature and cannot be separated, then since belief formation is a process that requires the deployment of concepts, so should perception; perception should be conceptual through and through.

I have argued elsewhere [26.36] that early vision, the first stage of visual processing, is CI and conceptually encapsulated in the sense that its processes are not affected directly, that is, in an on-line manner from cognitive states. There are, as a mater of course, many indirect cognitive effects on early vision, such as precueing effects and the effects of spatial attention in its capacity as a determinant of the focus of gaze, but these effects do not constitute cases of genuine CP [26.36] because, first, concepts do not enter the content of the states of early vision although they causally affect it, and second, because of the preceding fact, these sorts of cognitive effects can be mitigated and thus do not threaten the epistemological role of early vision as a neutral arbitrer of perceptual judgments. If this view is correct, early vision being CI does not employ any concepts and thus it cannot be like thinking, which necessarily involves concepts.

One might object that this claim overlooks the possibility that concepts are embedded in the circuits subserving early vision, rendering it conceptual from the within as it were and not because of any top-down cognitive influences. Being conceptually affected and by using inferences, there is no obstacle in thinking of early vision as akin to thinking. This objection is reinforced by two empirical facts. First, as we have seen, all stages of visual processing are restricted by a set of principles or constraints that aim to solve the various problems of underdetermination. These principles contain concepts and exemplify some form of knowledge that renders early vision theory-laden; it follows that early vision can be like thinking owing to its inherent structure. Second, as a result of perceptual learning, many an environmental regularity are

learned and stored in the early visual circuits to facilitate the processing of familiar input. Since these associations could arguably be construed as involving concepts, a claim could be made that early vision if affected by concepts.

In what follows, I examine these two objections and argue that both sorts of phenomena do not signify the CP and theory-ladenness of perception. This is so because, first, they do not entail that there are any concepts embedded in early vision, and second, because it is doubtful whether they contain any representations. This is also important for the wider claim that visual perception is like thinking, since thinking necessarily involves inferences driven by representations of both premises and the rules of inferences. If it turns out, as I argue here, that the transformation rules that visual perception employs to process its states are not represented anywhere in the system, this would severely undermine the claim that perceptual inferences are the same as the inferences used in belief formation.

26.4.1 The Operational Constraints in Visual Processing

There is extensive evidence that there is an important *body of information* that affects perception not in a top-down manner but from within and this might be construed as evidence for the CP of visual perception from its inception. The perceptual system does not function independently of any kind of internal restrictions. Visual processing at every level is constrained by certain principles or rather operational constraints that modulate information processing. Such constraints are needed because distal objects are underdetermined by the retinal image, and because the percept itself is underdetermined by the retinal image. Unless the processing of information in the perceptual system is constrained by some *assumptions* about the physical world, perception is not feasible. Most computational accounts hold that these constraints substantiate some reliable generalities of the natural physical world as it relates to the physical constitution and the needs of the perceiving agents. There is evidence that the physiological visual mechanisms reflect these constraints. Their physical making is such that they implement these constraints, which are thus hardwired in perceptual systems (see Appendix 26.E for a list of some of these constraints).

These are *Raftopoulos*' [26.36] *operational constraints* and *Burge*'s [26.37] *formation principles*. The operational constraints reflect higher-order physical regularities that govern the behavior of worldly objects and the geometry of the environment and which have been incorporated in the perceptual system through causal interaction with the environment over the evolution of the species. They allow us to lock onto medium size lumps of matter, by providing the discriminatory capacities necessary for the individuation and tracking of objects in a bottom-up way; they allow perception to generate perceptual states that present worldly objects as cohesive, bounded, solid, and spatiotemporally continuous entities.

The constraints are not available to introspection, function outside the realm of consciousness, and cannot be attributed as acts to the perceiver. One does not believe implicitly or explicitly that an object moves in continuous paths, that it persists in time, or that it is rigid, though one uses this information to parse and index the object. These constraints are not perceptually salient but one must be *sensitive* to them if one is to be described as perceiving their world. The constraints constitute the *modus operandi* of the perceptual system and not a set of rules used by the perceptual system as premises in perceptual inferences even though the *modus operandi* of the visual system consists of operations determined by laws describable in terms of computation principles. They are reflected in the functioning of the perceptual system and can be used only by it. They are available only for visual processing, whereas *theoretical* constraints are available for a wide range of cognitive applications. These constraints cannot be overridden since they are not under the perceiver's control; one cannot decide to substitute them with another body of constraints even if one knows that they lead to errors.

Being hardwired, the constraints are not even contentful states of the perceptual system. A state is formed through the spreading of activation and its modification as it passes through the synapses. The hardwired constraints specify the processing, i. e., the transformation from one state to another, but they are not the result of this processing. They are computational principles that describe transitions between states in the perceptual system. Although the states that are produced by means of these mathematical transformations have contents, there is no reason to suppose that the principles that specify the mathematical transformation operations are states of the system or contents of states in the system. If they are not states of the visual system, the principles that express them linguistically cannot be contents of any kind. Even though the perceptual system uses the operational constraints to represent some entity in the world and thus operates in accord with the principles reflected in the constraints (since the constraints are hardwired in the perceptual system, physiological conditions instantiate the constraints), the perceiver does not represent these principles or the constraints in any form. By the same token, these principles cannot be thought of as implicating concepts, since concepts are

representational. For this reason, perceptual operations should not be construed as inference rules, although they are describable as such, and they do not constitute either a body of knowledge or some theory about the world.

Recent work on Bayesian causal networks [26.4] draws a picture of the brain as a Bayesian net operating at various space and time scales, and suggests that there is a sharp distinction between internal probabilistic dependencies that can be explained by internal causal connections and those that cannot. Only those that cannot be explained internally carry information about the external world. Applying this to the case of the neuronal mechanisms that implement the operational constraints at work in visual processing, one could say that these mechanisms perform transformations that depend entirely on the internal probabilistic dependencies in the system as they are determined by the hardwired circuitry that realizes the internal causal connections and thus there is nothing representational about them.

These considerations allow us to address *Cavanagh*'s [26.38] claim that the processes that lead to the formation of a conscious percept constitute *visual cognition* in virtue of their use of inferences. The construction of a percept is "the task of visual cognition and, in almost all cases, each construct is a choice among an infinity of possibilities, chosen based on likehood, bias, or a whim, but chosen by rejecting other valid competitors" [26.38]. This process is an inference in that "it is not a guess. It is a rule-based extension from partial data to the most appropriate solution"; in the terminology of this chapter, the selection process is an abduction.

According to *Cavanagh* [26.38], for inference to take place the visual system should not rely to purely bottom-up analyses of the image that use only retinal information, such as sequences of filters that underlies facial recognition, or the cooperative networks that converge on the best descriptions of surfaces and contours. Instead, the visual system should use some object knowledge, which is nonretinal, context-dependent information. By *object knowledge* Cavanagh means any sort of nonretinal information that may be needed for the filling in that leads to the construction of the percept. This knowledge consists of rules that guide or constrain visual processing in order to solve the underdetermination problem that I mentioned above; they provide the rule-based extension from partial data that constitutes an inference. These rules do not influence visual processing in a top-down way, since they reside within the visual system; they are "from the side" [26.39].

The discussion concerning the nature of the operational constraints suggests that, their crucial role in perceptual processing notwithstanding, these constraints do not justify Cavanagh's characterization of visual perception as *visual cognition*, if cognition is thought of as involving discursive inferences.

26.4.2 Perceptual Learning

Evidence from studies showing early object classification effects suggests that to the extent that object classification presupposes object knowledge, this knowledge affects early vision in a top-down manner rendering it theory-laden. Moreover, even if one could show that these effects do not entail the CP of early vision, one could argue that since perceptual learning affects the way one sees the world, some experiences are learned and form memories that are stored in visual memory and affect perceptual processing from its inception. Our experiences shape the way we see the world.

Indeed, visual memories affect perception. Familiarity with objects or scenes that is built through repeated exposure to objects or scenes (sometimes one presentation is enough), or even repetition memory [26.40] facilitate search, affect figure from ground segmentation, speed up object identification and image classification, etc. [26.41–43].

Familiarity can affect visual processing in different ways. It may facilitate object identification and categorization, which are processes that take time since their final stage occurs between 300–600 ms after stimulus onset as is evidenced by the P3 responses in the brain, but their earlier stage starts about 150 ms after stimulus onset [26.44–46]. One notices that familiarity intervenes during the latest stage of visual processing (300–360 ms). These effects involve the higher cognitive levels of the brain at which semantic information and processing, both being required for object identification and categorization, occur [26.30]. In this sense, these sort of familiarity effects do not threaten the CI of early vision, which has ended about 120 ms after stimulus onset.

Familiarity, including repetition memory, also affects object classification (whether an image portrays an animal or a face), a process that occurs in short latencies (95–100 ms and 85–95 ms respectively) [26.47–49]. These early effects may pose a threat to the CI of early vision since they cannot be considered post-sensory. The threat would materialize should the classification processes either require semantic information to intervene or require the representations of objects in working memory to be activated, since that would, too, mean conceptual involvement.

Researchers however unanimously agree that the early classification effects in the brain result from the FFS and do not involve top-down semantic information,

nor do they require the activation of object memories. The brain areas involved are low-level visual areas (including the FEF – front eye fields) from V1 to no higher than V4 [26.48] or perhaps a bit more upstream to posterior IT [26.42] and lateral occipital complex (LOC) [26.49].

The early effects of familiarity may be explained by invoking contextual associations (target-context spatial relationships) that are stored in early sensory areas to form unconscious perceptual memories [26.50] which, when activated from incoming signals that bear the same or similar target-context spatial relationships, modify the FFS of neural activity resulting in the facilitating effects mentioned above. Thus, what is involved in the phenomenon are certain associations built in the early visual system that once activated speed up the feedforward processing. This is a case of rigging-up the early visual processing; it is not a case of top-down cognitive effects on early visual processing.

The early effects may also be explained by appealing to configurations of properties of objects or scenes. Currently, neurophysiological research [26.40, 49], psychological research [26.42], and computation modeling [26.51] suggest that what is stored in early visual areas are implicit associations representing fragments of objects and shapes, or *edge complexes*, as opposed to whole objects and shapes. One of the reasons that have led researchers to argue that it is object and shape fragments that are used in rapid classifications instead of whole objects and shapes is the following: If these associations reflecting some sort of object recognition can affect figure-ground segmentation as we have reasons to believe [26.42] in view of the fact that figure-ground segmentation occurs very early (80–100 ms) [26.52] these associations must be stored in early visual areas (up to V4, LO and posterior IT) and cannot be the representations stored in, say, anterior IT. The earlier visual areas store object and shape fragments and not holistic figures and shapes [26.40, 51].

The associations that are built in, through learning, in early visual circuits reflect in essence the statistical distribution of properties in environmental scenes [26.32, 53]. The statistical differences in physical properties of different subsets of images are detected very early by the visual system before any top-down semantic involvement as is evidenced by the elicitation of an early deflection in the differential between animal-target and nontarget ERP's at about 98 ms (in the occipital lobe) and 120 ms (in the frontal lobe). The low-cues could be retrieved very early in the visual system from a scene by analyzing the energy distribution across a set of orientation and spatial frequency-tuned channels [26.54]. This suggests that the rapid image classification may rely on low-level, or intermediate-level cues [26.51] that act diagnostically, that is, they allow the visual system to predict the gist of the scene and classify images very fast. These cues may be provided by coarse visual information, say by low-level spatial frequency information and thus the visual system does not have to rely on high-level fully integrated object representations in order to be able to classify rapidly visual scenes.

It follows that the classification of an object that occurs very early during the fast FFS at about 85–100 ms is due to associations regarding shape and object fragments stored in early visual areas and does not reflect any top-down cognitive effects on, that is, the CP of, early vision. Thus, early object classification is not a sign of the theory-ladenness of early vision, since the knowledge about the world does not affect it in a top-down manner.

To recapitulate the results of our discussion in this section, I have argued that neither the operational constraints operating in visual perception, nor perceptual learning entail that concepts affect early vision. Moreover, they do not entail that visual processing in general is theory-laden because of the role of these constraints, since they are not representational elements and any theory constitutively implicates representational elements. On the other hand, both the operational constraints and the effects of perceptual learning provide the context in which early vision constructs its hypotheses, and part of the context in which late vision operate, the other part being the viewer's knowledge of the world, which, as I have said, affects late vision but not early vision.

26.5 Late Vision, Inferences, and Thinking

Jackendoff [26.55] distinguishes visual awareness from visual understanding. There is a qualitative difference between the experience of a 3-D sketch and the experience of a $2\frac{1}{2}$D sketch. Although one is in some sense aware of the 3-D sketch or of category-based representations, however, this is not visual awareness but some other kind of awareness. Visual awareness is awareness of Marr's $2\frac{1}{2}$D sketch, which is the viewer-centered representation of the visible surfaces of objects, while the awareness of the 3-D sketch is visual understanding. Thus, the 3-D sketch, which includes the unseen surfaces that are not represented in the $2\frac{1}{2}$D sketch,

is a result of an inference. These views belong to the belief-based account of amodal completion: the 3-D sketch is the result of beliefs abductively inferred from the object's visible features and other background information from past experiences (see Appendix 26.D for an explanation of amodal and modal completion or perception).

The problem is whether the object identification that occurs in late vision (which, as we have seen most likely constitutes in essence an abductive inference) and depends on concepts should be thought of as a purely visual process or as a case of discursive understanding involving discursive inferences. If late vision involves conceptual contents and if the role of concepts and stored knowledge consists of providing some initial interpretation of the visual scene and in forming hypotheses about the identity of objects that are tested against perceptual information, one is tempted to say that this stage relies on inferences and thus differs in essence from the purely perceptual processes of early vision. Perhaps it would be better to construe late vision as a discursive stage involving thoughts, in the way of epistemic seeing, where *seeing* is used in a metaphorical nonperceptual sense, as where one says of his friend whom she visited *I see he has left*, based on perceptual evidence [26.56]. It is, also possible that *Dretske* [26.57, 58] thinks that seeing in the doxastic sense is not a visual but rather a discursive stage.

One might object, first, that abandoning this usage of *to see* violates ordinary usage. A fundamental ingredient of visual experience consists of meaningful 3-D solid objects. Adopting this proposal would mean that one should resist talking of seeing tigers and start talking about seeing viewer-centered visible surfaces. "By this criterion, much of the information we normally take to be visually conscious would not be, including the 3-D shape of objects as well as their categorical identity" [26.59].

More to the point, I think that one should not assume either that late vision involves abductive inferences construed as inferential discursive-state transformations that constitutively involve thoughts in the capacity of premises in inferences whose conclusion is a recognitional belief, or that late vision consists of discursively entertaining thoughts; if thinking is construed as constitutively implicating discursive argumentation, visual perception is different from thinking in some radical ways. The reason is twofold. First, seeing an object is not the result of a discursive inference, that is, a movement in thought from some premises to a conclusion, even though it involves concepts and intrastate transformations. Second, late vision is a stage in which conceptual modulation and perceptual processes form an inextricable link that differentiates late vision from discursive stages and renders it a different sort of a set of processes than understanding, even though late vision involves implicit beliefs regarding objects that guide the formation of hypotheses concerning object identity, and an explicit belief of the form *that O is F* eventually arises in the final stages of late vision. Late vision has an irreducible visual ingredient that makes it different from discursive understanding.

Let me clarify two terminological issues. First, judgments are occurrent states, whereas beliefs are dispositional states. To judge that O is F is to predicate F-ness to O while endorsing the predication [26.60]. To believe that O is F is to be disposed to judge under the right circumstances that O is F. This is one sense in which beliefs are dispositional items. There is also a distinction between standing knowledge (information stored in long term memory, LTM) and information that is activated in working memory (WM). The belief that O is F may be a standing information in LTM, a memory about O even though presently one does not have an occurrent thought about O. Beliefs need not be consciously or unconsciously apprehended, that is, activated in the mind, in order to be possessed by a subject, which means that beliefs are dispositional rather than occurrent items; this is a second sense in which beliefs are dispositional. When this information is activated, the thought that O is F emerges in WM; all thoughts are occurrent states.

It follows that a belief qua dispositional state may be either a piece of standing knowledge, in which case it is dispositional in the sense that when activated it becomes a thought, or a thought that awaits endorsement to become a judgment, in which case the belief is dispositional in the sense that it has the capacity to become a judgment. In the first case, beliefs differ from thoughts. In the second case, a belief is a thought held in WM, albeit one that has not been yet endorsed. In what follows, I assume that beliefs are either pieces of standing information or thoughts that have not been endorsed and thus are not judgments. Finally, by *implicit belief* I mean the belief held by a person who is not aware that she is having that belief.

As I said in the introduction, this chapter examines whether the abductive processes that take place in late vision should be construed as discursive inferences. Specifically, my claim is that the processes in late vision are not inferential processes where *inference* is understood as discursive, that is, as a process that involves drawing propositions or conclusions from other propositions, that are represented in the system, acting as premises by applying (explicitly or implicitly) inferential rules that are also represented. As we saw, these inferences are distinguished from *inferences* as understood by vision scientists according to whom

any transformation of signals carrying information according to some rule is an inference.

26.5.1 Late Vision, Hypothesis Testing, and Inference

I think that the states of late vision are not inferences from premises that include the contents of early vision states, even though it is usual to find claims that one infers that a tiger, for example, is present from the perceptual information retrieved from a visual scene. An inference relates some propositions in the form of premises with some other proposition, the conclusion. However, the objects and properties as they are represented in early vision do not constitute contents in the form of propositions, since they are part of the nonpropositional, iconic nonconceptual content of perception. In late vision, the perceptual content is conceptualized but the conceptualization is not a kind of inference but rather the application of stored concepts to some input that enters the cognitive centers of the brain and activates concepts by matching their content. Thus, even though the states in late vision are formed through the synergy of bottom-up visual information and top-down conceptual influences, they are not inferences from perceptual content.

Late vision involves hypotheses regarding the identity of objects and their testing against the sensory information stored in iconic memory. One might think that inferences are involved since testing hypotheses is an inferential process even though it is not an inference from perceptual content to a recognitional thought. It is, rather, an argument of the form of: if *A* and *B* then (conclusion) *C*, where *A* and *B* are background assumptions and the hypothesis regarding the identity of an object respectively, and *C* is the set of visual features that the object is likely to have. *A* consists of implicit beliefs about the features of the hypothesized visual object. If the predicted visual features of *C* match those that are stored in iconic memory in the visual areas, then the hypothesis about the identity of the object is likely correct. The process ends when the best possible fit is achieved. However, the test basis or evidence against which these hypotheses are tested for a match, that is, the iconic information stored in the sensory visual areas, is not a set of propositions but patterns of neuronal activations whose content is nonpropositional.

There is nothing inference-like in this matching. It is just a comparison between the activations of neuronal assemblies that encode the visual features in the scene and the activations of the neuronal assemblies that are activated top-down from the hypotheses. If the same assemblies are activated then there is a match. If they are not, the hypothesis fails to pass the test. This can be done through purely associational processes of the sort employed, say, in connectionist networks that process information according to rules and thus can be thought of as instantiating processing rules, without either representing these rules or operating on language-like symbolic representations. Such networks perform vector completion and function by satisfying soft constraints in order to produce the best output given the input into the system and the task at hand. Note that the algebraic and thus continuous nature of state transformations in neural networks, as opposed to the algorithmic discrete-like operations of classical AI (which assumes that the brain is a syntactic machine that processes discrete symbols according to rules that are also represented in the system) suits best the analogue nature of iconic representations.

In perceptual systems construed as neural networks, the fundamental representational unit is not some linguistic or linguistic-like entity but the activation pattern across a proprietary population of neurons. If one wishes to understand the workings of the visual brain, one should eschew sentences and propositions as bearers of representations and meanings and reconceptualize representations as activation patterns. This does not mean, of course, that the brain does not have symbolic representations but only that, first, these are a subset of the representations that the brain uses in its various functions, and, second and most importantly, the symbolic representations are constructed somehow out of the more fundamental context-dependent representations that the brain uses and are, consequently, a later construct, phylogenetically speaking. This has an important corollary for any theory of cognition that employs activation patterns as the fundamental units of representation, namely, that it must be able to explain the existence and usage of symbolic representations. This means also that the processing at work in the brain, that is, the transformation of the representational units to other representational units is not exclusively the transformation of complex or simple symbols by means of a set of syntactic rules as in the algorithms that, according to the classical view, the brain is supposed to run. Instead, it can be the algebraic transformation of activation patterns (in essence the algebraic transformations from one multidimensional matrix or tensor to another). The transformation is effected by the synaptic connections among the neurons as the signal passes from one layer to another. These connections have weights that constitute a filter through which the signal is transformed as it passes through.

The above also explain the holistic nature of the abductive visual processes that classical cognitive theories (the family of theories that assume that the brain is a syntactic machine that processes symbols that are

constant, context independent, and freely repeatable elements) have failed to capture. It is interesting that if I am right, Fodor's attempt to differentiate the perceptual systems from cognitive functions in order to protect the former from the abductive holistic reasoning implicated in the latter fails since late vision is abductive and holistic as well.

Since discursive inferences are carried out through rules that are represented in the system and operate on symbolic structures, the processing in a connectionist network does not involve discursive inferences, although it can be described in terms of inference making. Thus, even though seeing an object in late vision involves the application of concepts that unify the appearances of the object and of its features under some category, it is not an inferential process.

I have said that the noninferential process that results in the formation of a recognitional thought or belief can be recast in the form of an argument from some premise to a conclusion. However, this does not entail that the formation of the perceptual thought is a piece of reasoning, that is, a transition from a set of premises that act as a reason for holding the thought to the thought itself. Admittedly, the perceiver can be asked on what grounds she holds the thought that *O* is *F*, in which case she may reply *because I saw it* or *I saw that O is F*. However, this does not mean that the reason she cites as a justification of her thought is a premise from which she inferred the thought. The perceiver does not argue from her thought *I saw it to be thus and so* to the thought *It is thus and so*. She just forms the thought on the basis of the evidence included in her relevant perceptual state in the noninferential way I described above. What warrants the recognitional thought *O is F* is not the thought held by the perceiver that she sees *O* to be *F* but the perceptual state that presents to her the world as being such and such. "When one knows something to be so by virtue of seeing to be so, one's warrant for believing it to be so is that one sees it to be so, not one's believing that one sees it to be so" [26.57].

Spelke [26.3] who echoes *Rock*'s [26.2] views that the perceptual system combines inferential information to form the percept (for example, from visual angle and distance information, one infers and perceives size) – argues "perceiving objects may be more akin to thinking about the physical world than to sensing the immediate environment". The reason is that the perceptual system, to solve the underdetermination problem of both the distal object from the retinal image and of the percept from the retinal image, employs a set of object principles and that reflect the geometry and the physics of our environment. Since the contents of these principles consist of concepts, and thus the principles can be thought of as some form of knowledge about the world, perception engages in discursive, inferential processes. Against this, I argued above that the processes that constrain the operations of the visual system should not be construed as discursive inferences. They are hardwired in the perceptual circuits and are not represented in it. Thus, perceptual operations should not be construed as inference rules, although they are describable in terms of discursive inferential rules. It follows that the abduction that takes place in late vision is not an Aristotelian inference; it is better described by the ampliative vector completion of connectionism.

26.5.2 Late Vision and Discursive Understanding

Even if I am right that seeing in late vision is not the result of a discursive abductive inference but the result of a pattern-matching process that ensures the best fit with the available data, it is still arguable that late vision should be better construed as a stage of discursive understanding rather than as a visual stage. If object recognition involves forming a belief about class membership, even if the belief is not the result of an inference, why not say that recognizing an object is an experience-based belief that is a case of understanding rather than vision?

Late Vision Is more than Object Recognition

A first problem with this view is that late vision involves more than a recognitional belief. Suppose that *S* sees an animal and recognizes it as a tiger. In the parallel preattentive early vision, the proto-object that corresponds to the tiger is being represented amongst the other objects in the scene. After the proto-objects have been parsed, the object recognition system forms hypotheses regarding their identity. However, for the subject's confidence to reach the threshold that will allow her to form beliefs about the identity of the objects and report it, these hypotheses must be tested [26.61].

For this to happen, the relevant sensory activations enter the parietal and temporal lobes, and the prefrontal cortex, where the neuronal assemblies encoding the information about the objects in the scene are activated and the relevant hypotheses are formed. To test these hypotheses, the visual system allocates resources to features and regions that would confirm or disconfirm the hypotheses. To accomplish this, activation spreads through top-down signals from the cognitive centers to the visual areas of the brain where the visual sensory memory and the fragile visual memory store the proto-objects extracted from the visual scene. This way, conceptual information about the tiger affects visual processing and after some hypothesis testing the animal is recognized as a tiger through the synergy of

visual circuits and WM. At this point the explicit belief *O is F* is formed. This occurs after 300 ms, when the viewer consolidates the object in WM and identifies it with enough confidence to report it, which means that beliefs are formed at the final phases of late vision.

However, semantic modulation of visual processing and the process of conceptualization that eventually leads to object recognition starts at about 130–200 ms. There is thus a time gap between the onset of conceptualization and the recognition of an object, which is a prerequisite for the formation of an explicit recognitional belief. As *Treisman* and *Kanwisher* [26.62] observe, although the formation of hypotheses regarding the categorization of objects can occur within 130–200 ms after stimulus onset, it takes another 100 ms for subsequent processes to bring this information into awareness so that the perceiver could be aware of the presence of an object. To form the recognitional belief that *O is F*, one must be aware of the presence of an object token and construct first a coherent representation. This requires the enhancement through attentional modulation of the visual responses in early visual circuits that encode rich sensory information in order to integrate them into a coherent representation, which is why beliefs are delayed in time compared with the onset of conceptualization; not all of late vision involves explicit beliefs.

Late Vision as a Synergy of Bottom-Up and Top-Down Information Processing

A second reason why the beliefs formed in late vision are partly visual constructs and not pure thoughts is that the late stage of late vision in which explicit beliefs concerning object identity are formed constitutively involves visual circuits (that is, brain areas from LGN to IT in the ventral system). Pure thought, on the other hand, involves an amodal form of representation formed in higher centers of the brain, even though these amodal representations can trigger in a top-down manner the formation of mental images and can be triggered by sensory stimulation. The point is that amodal representations can be activated without a concomitant activation of the visual cortex. The representations in late vision, in contrast, are modal since they constitutively involve visual areas. Thus, what distinguishes late vision beliefs from pure thoughts is mostly the fact that the beliefs in late vision are formed through a synergy of bottom-up and top-down activation and their maintenance requires the active participation of the visual circuits. Pure thoughts can be activated and maintained in the absence of activation in visual circuits.

The constitutive reliance of late vision on the visual circuits suggests that late vision relies on the presence of the object of perception; it cannot cease to function as a perceptual demonstrative that refers to the object of perception, as this has been individuated though the processes of early vision. As such, late vision is constitutively context dependent since the demonstration of the perceptual particular is always context dependent. Thought, on the other hand, by its use of context independent symbols, is free of the particular perceptual context. Even though recognitional beliefs in late vision and pure perceptual beliefs involve concepts, the concepts function differently in the two contexts [26.37]:

> "Perceptual belief makes use of the singular and attributive elements in perception. In perceptual belief, pure attribution is separated from, and supplements, attributive guidance of contextually purported reference to particulars. Correct conceptualization of a perceptual attributive involves taking over the perceptual attributive's range of applicability and making use of its (perceptual) mode of presentation."

The attributive and singular elements in perception correspond to the perceived objects and their properties respectively. The attributive elements or properties guide the contextual reference to particulars or objects since the referent in a demonstrative perceptual reference is fixed through the properties of the referent as these properties are presented in perception.

Concepts enter the game in their capacity as pure attributions that make use of the perceptual mode of presentation. Burge's claim that in perceptual beliefs pure attributions supplement attributions that are used for contextual reference to particulars may be read to mean that perceptual beliefs are hybrid states involving both visual elements (the contextual attributions used for determining reference to objects and their properties) and conceptualizations of these perceptual attributives in the form of pure attributions. In this case, the role of perceptual attributives is ineliminable. In late vision, unlike in pure beliefs, there can be no case of pure attribution, that is, of attribution of features in the absence of perceptually relevant particulars since the attributions are used to single out these particulars.

The inextricable link between thought and perception in late vision explains the essentially contextual [26.63, 64] character of beliefs in late vision. The proposition expressed by the belief cannot be detached from the perceptual context in which it is believed and cannot be reduced to another belief in which some third person or objective content is substituted for the indexicals that figure in the thought (in the way one can substitute via Kaplan's characters the indexical terms with their referents and get the *objective* truth-evaluable content of the belief); the belief is tied to a idiosyncratic

viewpoint by making use of the viewer's physical presence and occupation of a certain location in space and time; the context in which the indexical thought is believed is essential to the information conveyed.

The discussion on late vision and the inferences it uses to construct the percept suggests that late vision, its conceptual nature notwithstanding, does not involve discursive inferences and in this sense is fundamentally different from thinking, if the latter is thought to implicate constitutively discursive inferences. Late vision employs abductive inferences, in that it constructs the representation that best fits the sensory image, but these inferences are not the result of the application of rules that are represented in the system. Even the operational constraints that restrict visual processing in late vision and could be thought of as transformation rules that the system follows to make inferences, are not, as we have seen, propositional structures or even representations in the brain. The inferences involved are informed and guided by conceptual information in pattern-matching processes but fall short of being discursive inferences. The fact that both conceptual and nonconceptual representations are in essence activation patterns allows us to understand how conceptual, symbolic information and nonconceptual iconic information could interact. The main difference between the two forms of representations is that the former are not homogeneous and have a syntactic structure that has a canonical decomposition, whereas the latter are homogeneous and lack a canonical decomposition. To appreciate the difference think of it in following way: the fact that a symbolic representation has a canonical decomposition means that not every subpart of the representation is a representation; only those subparts that satisfy the syntactic rules of the representational systems are symbols or representations. The expression (p&Q), for instance, is a symbol or a representation, but the expression (p(&q) is not. Any subpart of an image, on the other hand, is an image and thus a representation.

The output of late vision, namely the percept, enters the space of reasons and participates in discursive inferences and thus in thought.

26.6 Concluding Discussion

I have argued in this chapter that visual processing involves abductive inferences that aim to construct a representation, namely, the percept, that best matches the sensory information. To achieve this, the brain probably uses Bayesian strategies since abductive inferences are probabilistic in nature. I also argued that these inferences are not discursive inferences and since the latter are the characteristic trait of thinking, visual processing is not akin to thinking despite its usage of abductive inferences; my claim applies to both early vision and late vision.

The discussion in this chapter, and especially the view on the relation between perceptual inference and perceptual judgments, is in agreement with *Magnani*'s [26.65] elaboration on *Peirce*'s [26.66, 67] views on visual perception, which Peirce also conceived of perception as an abductive inference. In particular, I have tried to defend the thesis eloquently expressed by Peirce that the transition between abductive inferences and perceptual judgment is a continuous one without any sharp line of demarcation between them despite their many differences that I elaborated on in the previous section. My discussion also reinforces *Magnani*'s view that "judgments in perception are fallible but indubitable abductions we are not in any condition to psychologically conceive that they are false, as they are unconscious habits of inference" [26.65]. Most importantly, my account of the abductive inferences involved in visual perception fully justifies Magnani's claim that visual abduction is not sentential, that is, it does not employ symbolic, or discursive as I have called them, inferences. Instead it relies on pattern matching in which activation patterns that take on continuous values are compared. Thus, the representational medium employed is analogue and not symbolic in nature and the usage of stored knowledge in drawing inferences resembles more the use of models that put the incoming information in a context so that conclusions could be drawn rather than the recruitment of sentences and inference rules. In other words, visual abduction is model-based.

In constructing the percept, the brain uses a set of operational constraints that aim to solve the various underdetermination problems that the visual perception encounters in order to construct the percept. I have argued that these constraints should not be thought of as rules that are represented in the system or that have some representational contents and guide the perceptual inferences rendering them discursive. Instead, they are hardwired in the visual system and are not representations.

I also suggested, although I did not discuss this issue in full, that the recent developments in vision studies tend to bring together the different theories of vision by showing the points of contact between them, rather than to underline their differences.

To recapitulate, the main conclusion of this chapter is, first, that to the extent that thinking is associated with the use of discursive inferences, perception differs radically from thinking. If the meaning of thinking is extended to comprise nondiscursive inferences, the claim may be made that perception is thinking. In this case, however, a distinction should be drawn between discursive thinking that characterizes cognition and nondiscursive thinking that characterizes perceptual processes. Second, if thinking also necessarily involves the deployment of concepts, then there is a stage of visual processing, namely, early vision, which is not akin to thinking since its contents are nonconceptual. The other stage of visual processing, namely late vision, uses conceptual information. Since, as I will argue, the processes of late vision are not discursive inferences, if thinking is conceived as necessarily implicating discursive inferences late vision is not akin to thinking, notwithstanding the conceptual involvement. If the concept of thinking is extended to include other sorts of inferences, such as the model-based abductive inferences discussed in this chapter, late vision could be thought of as a sort of thinking, which, unlike early vision, implicates concepts (see Table 26.1 for a taxonomy).

Table 26.1 Visual perception and thinking

	Thinking	
Perception	**Thinking narrow**	**Thinking wide**
Early vision	No	Yes/no concepts
Late vision	No	Yes/yes concepts

26.A Appendix: Forms of Inferences

These are the three forms of inferences in which all syllogisms can be categorized.

26.A.1 Deduction

An inference is deductive if its logical structure is such that the conclusion of the inference is a logical consequence of the premises of the inference. This entails that if the premises of a deductive argument are true then its conclusion is necessarily true as well. In this sense, deductive arguments are truth preserving. This is equivalent to saying that in any interpretation of the inference in which the premises are true, the conclusion is true too. Differently put, if an argument is deductively valid, there is no model under which the premises are true but the conclusion is false. This is why deductive inferences are sometimes characterized as conclusive.

A typical example of a deductive argument is this: All men are mortal; Socrates is a man. Therefore Socrates is mortal.

26.A.2 Induction

An argument is inductive if its conclusion does not follow logically from the premises. The premises of an inductive argument may be true and still its conclusion false. The premises of an inductive argument provide epistemic support or epistemic warrant for its conclusion; they constitute evidence for the conclusion. By definition, inductive arguments are not truth preserving.

A typical example of an inductive argument is the following: Bird α is a crow and is black; bird β is a crow and is black; ... bird κ is a crow and is black. Therefore: All crows are probably black.

If the examined specimens are found in a variety of places and under different environmental conditions, the premises of the inference provide solid evidence for the conclusion. Yet, the conclusion may still be wrong since the next crow that we will examine may not be black. This example shows that the conclusion does not follow logically from the premises. It is still possible, no matter how good the premises, that is the evidence, are that the conclusion be false, which explains the qualification *probably* in the conclusion of an inductive argument. The world could be such that even crows α through κ are black, crow $\kappa + 1$ is white. For this reason inductions are considered to be nonconclusive but tentative [26.68].

26.A.3 Abduction or Inference to the Best Explanation

It is an inference in which a series of facts, which are either new, or improbable, or surprising on their own or in conjunction, are used as premises leading to a conclusion that constitutes an explanation of these facts. This explanation makes them more probable and more comprehensible in that it accounts for their appearance. As such, with abductive inferences the mind reaches conclusions that go far beyond what is given. For this reason, abductions are the main theoretical tools for building models and theories that explain reality. Ab-

duction is inductive since it is ampliative, does not preserve truth and is thus probabilistic in that the conclusion is tentative.

26.A.4 Differences Between the Modes of Inference

Induction versus Deduction

Induction is an ampliative inference, whereas deduction is not ampliative. This means that the information conveyed by the conclusion of an inductive argument goes beyond the information conveyed by the premises and, in this sense, the conclusion is not implicitly contained in the premises. In deduction, the conclusion is implicitly contained in the premises and the inference just makes it explicit. If all men are mortal and Socrates is a man, for example, the fact that Socrates is mortal is implicitly contained in these two propositions. What the deduction does is to render it explicit in the form of the conclusion. When we deduce that Socrates is mortal, our knowledge does not extend that which we already knew; it only makes it explicit. When, on the other hand, we inductively infer that all crows are probably black from the premise that all the specimens of crows that we have examined thus far are black, we extend the scope of our knowledge because the conclusion concerns all crows and not just the crows thus far examined.

The above discussion entails the main difference between deductive and inductive arguments. Deductive arguments are monotonous, while inductive arguments are not. This means that a valid deductive argument remains valid no matter how many premises we add to the argument. The reason is that the validity of the deductive argument presupposes that the conclusion is a logical conclusion of its premises. This fact does not change by the addition of new premises, no matter what these premises stipulate and thus the deductive argument remains valid. Things are radically different in induction. A new premise may change the conclusion even if the previous premises strongly supported the conclusion. For example, if we discover that crow $\kappa + 1$ is white, this undermines that previously drawn and well-supported conclusion that all crows are black.

Induction versus Abduction

Both abduction and induction are tentative forms of inference in that they do not warrant the truth of their conclusion even if the premises are true. They are, also, both ampliative in that the conclusion introduces information that was not contained implicitly in the premises. As we have seen, in abduction one aims to explain or account for a set of data. Induction is a more general form of inference. When, for instance, one successfully tests a hypothesis by making predictions that are borne out, the predicted data provide inductive, but not abductive, support for the hypothesis. In general, the evaluation phase in hypothesis, or theory, construction is considered to be inductive. Conceiving the explanatory hypothesis, on the other hand, is an abductive process that may assume the form of a pure, educated guess that need not have involved any previous testing. In this case, the abductively produced hypothesis is not, a priori, the best explanation for the set of data that need explanation; this is one of the occasions in which abduction can be distinguished from the inference to the best explanation. However, it should be stressed, although I do not have the space to elaborate on this problem, that in realistic scientific practice abduction as theory construction could not be separated from the evaluative inductive phase since they both form an inextricable link. This justifies the claim that abduction is an inference to the best explanation.

A further difference between abduction and induction is that even though both kinds of inference are ampliative, in abduction the conclusion may, and usually does, contain terms that do not figure in the premises. Almost all theoretical entities in science were conceived as a result of abduction. The nucleus of an atom, for example, was posited as a way of explaining the scattering of particles after the bombardment of atoms. Nowhere in the premises of the abductive argument was the notion of an atom present; the evidence consisted in measurements of the deviation of the pathways particles from their predicted values after the bombardment. The conclusion *all crows are probably black*, on the other hand, contains only terms that are available in the premises.

26.B Appendix: Constructivism

Some of Marr's particular proposals of his model have been criticized on many grounds (see, for example, [26.59]). In particular, against Marr's model of object recognition, it has been argued by several researchers that object recognition may be more image-based than based on object-centered representations, which means that the latter may be less important than Marr thought them to be. Neurophysiological studies [26.69] also suggest that both object-centered and viewer-centered representations play a substantial role in object recognition. Nevertheless, his general ideas about the construction of gradual visual representations

remain useful. According to this form of constructivism, vision consists of four stages, each of which outputs a different kind of visual representation:

1. The *formation of the retinal image*; the immediate stimulus for vision, that is the first stimulus that affects directly the sensory organs (this is called the proximal stimulus) is the pair of two-dimensional (2-D) images projected from the environment to the eyes. This representation is based on a 2-D retinal organization. At this stage, the information impinging on the retina (which as you may recall concerns intensity of illumination and wavelengths, and which is captured by the retinal receptors) is organized so that all of the information about the spatial distribution of light (i. e., the light intensity falling on each retinal receptor) be recast in a reference frame that consists of square image elements (*pixels*), each indicating with a numerical value the light intensity falling on each receptor. Sometimes, the processes of this stage are called *sensation*.
2. The *image-based stage*; it includes operations that receive as input the retinal image (that is, the numerical array of values of light intensities in each pixel) and process it in order to detect local edges and lines, to link these edges and lines in a more global scale, to match up corresponding images in the two eyes, to define 2-D regions in the image, and to detect line terminations and blobs. This stage outputs 2-D surfaces at some particular slant that are located at some distance from the viewer in 3-D space.

In general, the image-based representation has the following properties: First, it receives as input and thus operates first on information about the 2-D structure of the retinal image rather than on information concerning the physical, distal, objects. Second, its geometry is inherently two-dimensional. Third, the image-based representation of the 2-D features is cast in a coordinate reference system that is defined with respect to the retina (as a result, the organization of the information is called retino-topic). This means that the axes of the reference system are aligned with the eye rather than the body or the environment. This stage is the first stage of *perception* proper:

3. The *surface-based*; in this stage, vision constructs representations of the intrinsic properties of surfaces in the environment that might have produced the features constructed in the image-based model. At this stage, and in contradistinction to the preceding stage, the information about the worldly surfaces is represented in three dimensions. Marr's two-and-a-half-dimensional (2.5-D) sketch is a typical example of a surface-based representation. Note that the surface-based representation of a visual scene does not contain information about all the surfaces that are present in the scene, but only those that are visible for the viewer's current viewpoint.

In general, the surface-based representation has the following properties: First, The elements that the surface-based stage outputs consist of the output of the image-based stage, that is, in 2-D surfaces at some particular slant that are located at some distance from the viewer in 3-D space. Second, these 2-D surfaces are represented within a 3-D spatial framework. Third, the aforementioned reference framework is defined in terms of the direction and distance of the surfaces from the observer's standpoint (it is egocentric):

4. *The object-based*; this is the stage in which the visual system constructs 3-D representations of objects that include at least some of the occluded surfaces of the objects, that is, the surfaces that are invisible from the standpoint of the viewer, such as the back parts of objects. In this sense, this is the stage in which explicit representations of whole objects in the environment are constructed. It goes without saying that in order for the visual system to achieve this aim, it must use information about the whole objects that viewers have stored from their previous visual encounters with objects of the same type. The viewer retrieves from memory this information and fills in with it the surface-based image constructed at the previous stage.

In general, the object-based representation has the following properties: First, this stage outputs volumetric representations of objects that may include information about unseen surfaces. Second, the space in which these objects are represented is three-dimensional. Third, the frame of reference in which the object-based representations are cast is defined in terms of the intrinsic structure of the objects and the visual scene (it is scene-based or allocentric).

26.C Appendix: Bayes' Theorem and Some of Its Epistemological Aspects

Bayes' theorem is the following probabilistic formula (in its simple form because there is another formulation when one considers two competing hypotheses), where A is a hypothesis purporting to explain a set of data B

$$P(A/B) = P(B/A)P(A)/P(B) ,$$

where $P(A)$ is the prior probability, that is, the initial degree of belief in A; $P(A/B)$ is the conditional probability of A given B, or posterior probability, that is, the degree of belief in A after taking into consideration B; $P(B)$ is the probability of B. $P(B/A)$ is the likelihood of B given A, that is, the degree of belief that B is true given that A is true. The ratio $P(B/A)/P(B)$ represents the degree of support that B provides for A.

Suppose that B is the sensory information encoded by a neuronal assembly at level l-1, and A is the hypothesis that the neuronal assembly at level l posits as an explanation of B. Bayes' theorem tells us that the probability that A is true, that is, the probability that level l represent a true pattern in the environment given the sensory data B, depends first on the prior probability of hypothesis A, that is the probability of A before the predictions of A are tested. This prior probability depends on both the incoming signal to l but, also and most crucially because many different causes could have caused the incoming signal, on the contextual effects because these are the factors that determine which is the most likely explanation of the data among the various possible alternative accounts.

The probability of A also depends on the $P(B/A)$, that is, the probability that B be true given A. This reflects a significant epistemological insight, namely, that since a correct account of a set of data explains away, these data are a *natural consequence* of the explaining hypothesis, or naturally fit into the conceptual framework created by the hypothesis. The various gravity phenomena, for instance, become very plausible in view of the law of gravity; they are not so much so if the hypothesis purporting to explain these same phenomena involves some accidents of nature, even if they are systematic. To put in a reverse way, if gravity exists, then the probability that unsupported objects will fall down is greater than the probability of these objects falling down if some other hypothesis is postulated to explain the fall of unsupported objects.

The probability of the hypothesis A depends inversely on the probability of the data B. Since probabilities take values from (0 to 1), the smaller the probability in the denominator, that is, the more surprising and thus improbable B is, the greater the probability that A be true given B. This part of the equation also reflects an important epistemological insight, namely that the more surprising a set of data is, the more likely is to be true a hypothesis that successfully explains them. Finally, the ratio $P(B/A)/P(B)$ expresses the support B provides to A in the sense that the greater this ratio, the greater the probability that the hypothesis A is true.

26.D Appendix: Modal and Amodal Completion or Perception

There are two sorts of completion. In modal completion the viewer has a distinct visual impression of a hidden contour or other hidden features even though these features are not occurrent sensory features. The perceptual system fills in the missing features, which thus become as phenomenally occurrent as the occurrent sensory features of the object.

In amodal completion, one does not have a perceptual impression of the object's hidden features since the perceptual system does not fill in the missing features as it happens in modal perception, although as we shall see mental imagery can fill in the missing phenomenology; the hidden features are not perceptually occurrent.

There are cases of amodal perception that are purely perceptual, that is, bottom-up. In these cases, although no direct signals from the hidden features impinge on the retina (there is no local information available), the perceptual system can extract information regarding them from the global information contained in the visual scene without any cognitive involvement, as the resistance of the ensuing percepts to beliefs indicates. However, in such cases, the hidden features are not perceived. One simply has the visual impression of a single concrete object that is partially occluded and not the visual impression of various disparate image regions. Therefore, in these perceptually driven amodal completions there is no mental imagery involved, since no top-down signals from cognitive areas are required for the completion, and since the hidden features are not phenomenologically present.

There are also cases of amodal completion that are cognitively driven, such as the formation of the 3-D sketch of an object, in which the hidden features of the object are represented through the top-down acti-

vation of the visual cortex from the cognitive centers of the brain. In some of these cases, top-down processes activate the early visual areas and fill in the missing features that become phenomenologically present. In other cases of cognitively driven amodal completion, the viewer simply forms a pure thought concerning the hidden structure in the absence of any activation of the visual areas and thus in the absence of mental imagery.

26.E Appendix: Operational Constraints in Visual Processing

Studies by [26.3, 70–72] show that infants, almost from the very beginning, are constrained by a number of domain-specific principles about material objects and some of their properties. As *Karmiloff-Smith* [26.72] remarks, these constraints involve "attention biases toward particular inputs and a certain number of principled predispositions constraining the computation of those inputs". Such predispositions are the conception of object persistence, and four basic principles (boundness, cohesion, rigidity, and no action at a distance).

The *cohesion principle*: "two surface points lie on the same object only if the points are linked by a path of connected surface points". This entails that if some relative motion alters the adjacency relations among points at their borders, the surfaces lie on distinct objects, and that "all points on an object move on connected paths over space and time. When surface points appear at different places and times such that no connected path could unite their appearances, the surface points do not lie on the same object".

According to the *boundness principle* "two surface points lie on distinct objects only if no path of connected surface points links them". This principle determines the set of those points that define an object boundary and entails that two distinct objects cannot interpenetrate, because two distinct bodies cannot occupy the same place at the same time.

Finally the *rigidity* and *no action at a distance* principles specify that bodies move rigidly (unless the other mechanisms show that a seemingly unique body is, in fact, a set of two distinct bodies) and that they move independently of one another (unless the mechanisms show that two seemingly separate objects are in fact connected).

Further studies shed light on the nature of these principles or constraints and on the neuronal mechanisms that may realize them. There is evidence that the physiological mechanisms underlying vision reflect these constraints; their physical making is such that they implement these constraints, from cells for edge detection to mechanisms implementing the epipolar constraint [26.73, 74]. Thus, one might claim that these principles are hardwired in our perceptual system.

The formation of the *full primal sketch* in *Marr*'s [26.12] theory relies upon the principles of *local proximity* (adjacent elements are combined) and of *similarity* (similarly oriented elements are combined). It also relies upon [26.20] the more general principle of *closure* (two edge-segments could be joined even though their contrasts differ because of illumination effects).

Other principles used by early visual processing to solve the problem of the underdetermination of perception by the retinal image are those of *continuity* (the shapes of natural objects tend to vary smoothly and usually do not have abrupt discontinuities), *proximity* (since matter is cohesive, adjacent regions usually belong together and remain so even when the object moves), and *similarity* (since the same kind of surface absorbs and reflects light in the same way the different subregions of an object are likely to look similar).

The formation of the $2\frac{1}{2}$D *sketch* is similarly underdetermined, in that there is a great deal of ambiguity in matching features between the two images form in the retinas of the two eyes, since there is usually more than one possible match. Stereopsis requires a unique matching, which means that the matching processing must be constrained. The formation of the $2\frac{1}{2}$D *sketch*, therefore, relies upon a different set of operational constraints that guide stereopsis. "A given point on a physical surface has a unique position in space at some time" [26.69] and matter is cohesive and surfaces are generally smooth. These operational constraints give rise to the general constraints of *compatibility* (a pair of image elements are matched together if they are physically similar, since they originate from the same point of the surface of an object), of *uniqueness* (an item from one image matches with only one item from the other image), and of *continuity* (disparities must vary smoothly). Another constraint posited by all models of stereopsis is the *epipolar* constraint (the viewing geometry is known). *Mayhew* and *Frisby*'s [26.75] account of stereopsis posits some additional constraints, most notably, the principle of *figural continuity*, according to which figural relationships are used to eliminate most of alternative candidate matches between the two images.

References

26.1 H. von Helmholtz: *Treatise on Psychological Optics* (Dover, New York 1878/ 1925)
26.2 I. Rock: *The Logic of Perception* (MIT Press, Cambridge 1983)
26.3 E.S. Spelke: Object perception. In: *Readings in Philosophy and Cognitive Science*, ed. by A.I. Goldman (MIT Press, Cambridge 1988)
26.4 A. Clark: Whatever next? Predictive brains, situated agents, and the future of cognitive science, Behav. Brain Sci. **36**, 181–253 (2013)
26.5 M. Rescorla: The causal relevance of content to computation, Philos. Phenomenol. Res. **88**(1), 173–208 (2014)
26.6 L. Shams, U.R. Beierholm: Causal inference in perception, Trends Cogn. Sci. (Regul. Ed.) **14**, 425–432 (2010)
26.7 N. Orlandi: *The Innocent Eye: Why Vision Is not a Cognitive Process* (Oxford Univ. Press, Oxford 2014)
26.8 P. Lipton: *Inference to the Best Explanation*, 2nd edn. (Routledge, London, New York 2004)
26.9 D.G. Campos: On the Distinction between Peirce's Abduction and Lipton's inference to the best explanation, Synthese **180**, 419–442 (2011)
26.10 G. Minnameier: Peirce-suit of Truth-why inference to the best explanation and abduction ought not to be confused, Erkenntnis **60**, 75–105 (2004)
26.11 G. Harman: Enumerative induction as inference to the best explanation, J. Philos. **68**(18), 529–533 (1965)
26.12 D. Marr: *Vision: A Computational Investigation into Human Representation and Processing of Visual Information* (Freeman, San Francisco 1982)
26.13 J. Biederman: Recognition by components: A theory of human image understanding, Psychol. Rev. **94**, 115–147 (1987)
26.14 A. Johnston: Object constancy in face processing: Intermediate representations and object forms, Ir. J. Psychol. **13**, 425–438 (1992)
26.15 G.W. Humphreys, V. Bruce: *Visual cognition: Computational, Experimental and Neuropsychological Perspectives* (Lawrence Erlbaum, Hove 1989)
26.16 J.J. Gibson: *The Ecological Approach to Visual Perception* (Houghton-Mifflin, Boston 1979)
26.17 J. Fodor, Z. Pylyshyn: How direct is visual perception? Some reflections on Gibson's 'Ecological Approach, Cognition **9**, 139–196 (1981)
26.18 M. Rowlands: *The New Science of Mind: From Extended Mind to Embodied Phenomenology* (MIT Press, Cambridge 2010)
26.19 J. Norman: Two visual systems and two theories of perception: An attempt to reconcile the constructivist and ecological approaches, Behav. Brain Sci. **25**, 73–144 (2002)
26.20 V. Bruce, P.R. Green: *Visual Perception: Physiology, Psychology and Ecology*, 2nd edn. (Lawrence Erlbaum, Hillsdale 1993)
26.21 A. Noe: *Action in Perception* (MIT Press, Cambridge 2004)
26.22 J. Pearl: *Causality: Models, Reasoning and Inference* (Cambridge Univ. Press, Cambridge 2009)
26.23 V.A.F. Lamme: Why visual attention and awareness are different, Trends Cogn. Sci. **7**, 12–18 (2003)
26.24 V.A.F. Lamme: Independent neural definitions of visual awareness and attention. In: *The Cognitive Penetrability of Perception: An Interdisciplinary Approach*, ed. by A. Raftopoulos (Nova-Science Books, Hauppauge 2004)
26.25 A. Clark: An embodied cognitive science?, Trends Cogn. Sci. **3**(9), 345–351 (1999)
26.26 P. Vecera: Toward a biased competition account of object-based segmentation and attention, Brain Mind **1**, 353–384 (2000)
26.27 E.C. Hildreth, S. Ulmann: The computational study of vision. In: *Foundations of Cognitive Science*, ed. by M.I. Posner (MIT Press, Cambridge 1989)
26.28 Z. Pylyshyn: Is vision continuous with cognition? The case for cognitive impenetrability of visual perception, Behav. Brain Sci. **22**, 341–423 (1999)
26.29 M. Barr: The proactive brain: Memory for predictions, Philos. Trans. R. Soc. Lond. B Biol. Sci. **364**, 1235–1243 (2009)
26.30 K. Kihara, Y. Takeda: Time course of the integration of spatial frequency-based information in natural scenes, Vis. Res. **50**, 2158–2162 (2010)
26.31 C. Peyrin, C.M. Michel, S. Schwartz, G. Thut, M. Seghier, T. Landis, C. Marendaz, P. Vuilleumier: The neural processes and timing of top-down processes during coarse-to-fine categorization of visual scenes: A combined fMRI and ERP study, J. Cogn. Neurosci. **22**, 2678–2780 (2010)
26.32 A. Delorme, G.A. Rousselet, M.J.-M. Macé, M. Fabre-Thorpe: Interaction of top-down and bottom up processing in the fast visual analysis of natural scenes, Cogn. Brain Res. **19**, 103–113 (2004)
26.33 L. Chelazzi, E. Miller, J. Duncan, R. Desimone: A neural basis for visual search in inferior temporal cortex, Nature **363**, 345–347 (1993)
26.34 P.R. Roelfsema, V.A.F. Lamme, H. Spekreijse: Object-based attention in the primary visual cortex of the macaque monkey, Nature **395**, 376–381 (1998)
26.35 S.M. Kosslyn: *Image and Brain* (MIT Press, Cambridge 1994)
26.36 A. Raftopoulos: *Cognition and Perception: How Do Psychology and the Neural Sciences Inform Philosophy?* (MIT Press, Cambridge 2009)
26.37 T. Burge: *Origins of Objectivity* (Clarendon Press, Oxford 2010)
26.38 P. Cavanagh: Visual cognition, Vis. Res. **51**, 1538–1551 (2011)
26.39 R. Gregory: *Concepts and Mechanisms of Perception* (Charles Scribners and Sons, New York 1974)
26.40 K. Grill-Spector, T. Kushnir, T. Hendler, S. Edelman, Y. Itzchak, R. Malach: A sequence of object-processing stages revealed by fMRI in the Human occipital lobe, Human Brain Mapping **6**, 316–328 (1998)
26.41 H. Liu, Y. Agam, J.R. Madsen, G. Krelman: Timing, timing, timing: Fast decoding of object information from intracranial field potentials in human visual cortex, Neuron **62**, 281–290 (2009)

26.42 M. Peterson: Overlapping partial configurations in object memory. In: *Perception of Faces, Objects, and Scenes: Analytic and Holistic Processes*, ed. by M. Peterson, G. Rhodes (Oxford Univ. Press, New York 2003)

26.43 M. Peterson, J. Enns: The edge complex: Implicit memory for figure assignment in shape perception, Percept. Psychophys. **67**(4), 727–740 (2005)

26.44 M. Fabre-Thorpe, A. Delorme, C. Marlot, S. Thorpe: A limit to the speed of processing in ultra-rapid visual categorization of novel natural scenes, J. Cogn. Neurosci. **13**(2), 171–180 (2001)

26.45 J.S. Johnson, B.A. Olshausen: The earliest EEG signatures of object recognition in a cued-target task are postesensory, J. Vis. **5**, 299–312 (2005)

26.46 S. Thorpe, D. Fize, C. Marlot: Speed of processing in the human visual system, Nature **381**, 520–522 (1996)

26.47 S.M. Crouzet, H. Kirchner, S.J. Thorpe: Fast saccades toward faces: Face detection in just 100 ms, J. Vis. **10**(4), 1–17 (2010)

26.48 H. Kirchner, S.J. Thorpe: Ultra-rapid object detection with saccadic movements: Visual processing speed revisited, Vis. Res. **46**, 1762–1776 (2006)

26.49 K. Grill-Spector, R. Henson, A. Martin: Repetition and the brain: Neural models of stimulus-specific effects, Trends Cogn. Sci. **10**, 14–23 (2006)

26.50 M. Chaumon, V. Drouet, C. Tallon-Baudry: Unconscious associative memory affects visual processing before 100 ms, J. Vis. **8**(3), 1–10 (2008)

26.51 S. Ullman, M. Vidal-Naquet, E. Sali: Visual features of intermediate complexity and their use in classification, Nat. Neurosci. **5**(7), 682–687 (2002)

26.52 V.A.F. Lamme, H. Super, R. Landman, P.R. Roelfsema, H. Spekreijse: The role of primary visual cortex (V1) in visual awareness, Vis. Res. **40**(10–12), 1507–1521 (2000)

26.53 R. VanRullen, S.J. Thorpe: The time course of visual processing: From early perception to decision making, J. Cogn. Neurosci. **13**, 454–461 (2001)

26.54 A. Torralba, A. Oliva: Statistics of natural image categories, Network **14**, 391–412 (2013)

26.55 R. Jackendoff: *Consciousness and the Computational Mind* (MIT Press, Cambridge 1989)

26.56 F. Jackson: *Perception: A Representative Theory* (Cambridge Univ. Press, Cambridge 1977)

26.57 F. Dretske: Conscious experience, Mind **102**, 263–283 (1993)

26.58 F. Dretske: *Naturalizing the Mind* (MIT Press, Cambridge 1995)

26.59 S.E. Palmer: *Vision Science: Photons to Phenomenology* (MIT Press, Cambridge 1999)

26.60 J. McDowell: *Mind and World* (Harvard Univ. Press, Cambridge 2004)

26.61 A. Treisman: How the deployment of attention determines what we see, Vis. Cogn. **14**, 411–443 (2006)

26.62 A. Treisman, N.G. Kanwisher: Perceiving visually presented objects: Recognition, awareness, and modularity, Curr. Opin. Neurobiol. **8**, 218–226 (1998)

26.63 J. Perry: *Knowledge, Possibility, and Consciousness*, 2nd edn. (MIT Press, Cambridge 2001)

26.64 R.C. Stalnaker: *Our Knowledge of the Internal World* (Clarendon Press, Oxford 2008)

26.65 L. Magnani: *Abductive Cognition. The Epistemological and Eco-Cognitive Dimensions of Hypothetical Reasoning* (Springer, Berlin 2009)

26.66 C.S. Peirce: Perceptual judgments (1902. In: *Philosophical Writings of Peirce*, ed. by J. Buchler (Dover, New York 1955)

26.67 C.S. Peirce, N. Houser: *The Essential Peirce: Selected Philosophical Writings*, Vol. 2 (Indiana Univ. Press, Bloomington 1998)

26.68 S. Toulmin: *The Uses of Argument* (Cambridge Univ. Press, Cambridge 1958)

26.69 D.I. Perrett, M.W. Oram, J.K. Hietanen, P.J. Benson: Issues of representation in object vision. In: *The Neuropsychology of Higher Vision: Collated Tutorial Essays*, ed. by M.J. Farah, G. Ratcliff (Lawrence Erlbaum, Hillsdale 1994)

26.70 E.S. Spelke, R. Kestenbaum, D.J. Simons, D. Wein: Spatio-temporal continuity, smoothness of motion and object identity in infancy, Br. J. Dev. Psychol. **13**, 113–142 (1995)

26.71 E.S. Spelke: Principles of object perception, Cogn. Sci. **14**, 29–56 (1990)

26.72 A. Karmiloff-Smith: *Beyond Modularity: A Developmental Perspective on Cognitive Science* (MIT Press, Cambridge 1992)

26.73 G.F. Poggio, W.H. Talbot: Mechanisms of static and dynamic stereopsis in foveal cortex of the rhesus monkey, J. Physiol. **315**, 469–492 (1981)

26.74 D. Ferster: A comparison of binocular depth mechanisms in areas 17 and 18 of the cat visual cortex, J. Physiol. **311**, 623–655 (1981)

26.75 J.F.W. Mayhew, J.P. Frisby: Psychophysical and computational studies towards a theory of human stereopsis, Artif. Intell. **17**, 349–385 (1981)