1 Introduction

This chapter deals with the question of how the brain arrives at stable percepts that best represent the environment, based on the input it receives from the sensory organs. Perhaps it is best to introduce and motivate the material in this chapter with an analogy from logic. The analogy involves the ancient Greek "liar's paradox," attributed to Epimenides (Gardner 1982), a legendary philosopher who was born in Crete and moved to Athens as an old man, ca. 500 BC. The paradox is composed of two sentences:

Epimenides, the Cretan, said:

'All Cretans are liars.'

For the paradox to work, you need the assumptions that liars always lie, and that truth-tellers - let's call them knaves, following Smullyan's (1987) notation - always tell the truth. When we are first exposed to the sentence pair, we quickly come to the realization that there is an endless loop of alternating states, or hypotheses, or schemata, under these assumptions. The first state is that we accept the statement, "All Cretans are liars" as true. But then Epimenides must also be a liar, since he is a Cretan. However, if he is a liar, then his statement is a lie, that is, all Cretans must be knaves, which is the second state. Well, if this is so, Epimenides must also be a knave as a Cretan. As such, he must have spoken the truth when he said "All Cretans are liars." We are now back to the first state, where we started, accepting Epimenides' statement as true. Ultimately, we are led to a logical vicious circle, alternating...
between accepting either the statement itself ("All Cretans are liars") or its opposite ("All Cretans are knaves").

There are numerous such "puzzles" that result in multiple semistable states in logic (Dodds 1951; Gardner 1982; Hofstadter 1980; Smullyan 1978, 1987). Typically, puzzles are given as a set of propositions. When faced with such input, people attempt to make sense of it and to arrive at a solution. One common approach is to make a hypothesis and proceed to test it against the available evidence from the propositions at hand. With ordinary puzzles, there is a unique solution: that is, a stable hypothesis that satisfies all the propositions. However, multistable puzzles result in competing semistable solutions that refute each other ad infinitum, as illustrated above. At first glance, such puzzles may appear as contrived oddities to the average person, who may think that they have very little to do with reality. After all, logical systems have well-defined rules and axiomatic formulations that lead to a self-consistent structure. However, it turns out that such puzzles are quite important: the liar's paradox can be related to Gödel's famous theorem, which showed that any consistent axiomatic mathematic system includes undecidable propositions (just like the liar's paradox); in other words, such systems can never be proved to be consistent. In fact, Hofstadter (1980: 17) makes the point that "Gödel's discovery involves the translation of an ancient paradox in philosophy into mathematical terms. The paradox is the so-called Epimenides paradox." So, after all, it may not sound so strange that the philosopher Chrysippus (280-206 BC) wrote no fewer than six treatises on the liar's paradox or that the poet Philetas (fourth century BC) is said to have worried to an early grave over it (Gardner 1982).

The point of this chapter is that multistable stimuli, as well as related classes of stimuli, reveal a wealth of information on the perceptual process, just like multistable puzzles reveal a lot about logical systems. Rather than regarding them as oddities, we can look at them as useful tools in studying how the brain arrives at a stable percept based on the available sensory input. This chapter reviews evidence indicating that the sensory/perceptual component of the brain adopts an approach similar to that of the logical component of the brain: it arrives at a stable percept by constructing and testing alternative competing hypotheses, based on the input it receives from the sensory organs (Rock 1983). Gregory (1968, 1970, 1980, 1997) must be credited with formulating best this view of perceptions as hypotheses and the brain as the agent for generating and testing them. This chapter serves as one of two reports, in conjunction with chapter 9 by Papathomas, Kovacs, Feher, and Julesz, casting binocular rivalry as a special case of dealing with multistable stimuli.

The material is organized as follows: section 2 presents the view of the brain as an agent that constantly makes and tests alternative hypotheses and deals with the conflicting requirements for stability and plasticity in the brain. Section 3 presents evidence that supports this view from visual stimulation with multistable, "concealed," and "impossible" figures. Theoretical and computational models that deal with the interaction of data-driven and
concept-driven processes are briefly discussed in section 4. Finally, additional evidence and concluding remarks are given in section 5.

2 Theories of Perception as an Interactive Process

The brain's main sensory/perceptual function is to enable organisms to interact with their environment on a moment-to-moment basis for survival. Visual input plays a significant role in this function. However, the very nature of this input forces the visual system to work always with partial information in assessing the state of the environment. This is only partly due to occlusions, either of far objects by near objects or of rear surfaces by front surfaces of the same object (self-occlusions). Even under optimal circumstances, without occlusions, the input that is extracted by the visual system is but a small subset of the information that is available in the environment in the form of Adelson and Bergen's (1991) plenoptic function. This is due to limitations in the viewing angle and eye geometry and the limited sampling frequency of sensors that sample the input along the dimensions of space and wavelength.

Thus we only have the illusion of working with a complete input. One must conclude that the brain must go beyond the sensory data in providing us with an accurate representation of the environment. One view of how this is accomplished holds that the brain constantly constructs hypotheses based on the input it receives from the sensory organs; it then tests these alternative competing hypotheses, rejecting the weak ones and adopting the strongest one to arrive at a stable state. Gregory (1970) makes the point that the brain constructs hypotheses that go beyond available data, to formulate object hypotheses. Indeed, the point can be made that perception always involves the subject's perceptual set, in addition to the sensory input (Schiffman 1996: 187). The first modern account of this idea was given by Bruner (1957), who argued that the final percept is influenced by the perceiver's beliefs and state of mind.

The reason for emphasizing the word "constantly" in the previous paragraph is that there is evidence from multistable stimuli suggesting that this hypothesis-constructing-and-testing process does not stop after arriving at a stable state. The sensory input is continuously used to explore alternative hypotheses, in an endless quest to assess the state of the external world. Another way of putting it is that the brain goes continually through a "reality check," as the term so aptly suggests. In natural circumstances, when there is plenty of redundant sensory information, "reality" appears to be stable, and we are not aware of this continuous hypothesis-testing process. In these cases, the brain arrives at a dominant perceptual state and never departs from it, since there are only weak competing alternatives, if there are any at all.

However, there are circumstances, even in everyday life, when this influence of concept-driven processes becomes obvious. This happens, for instance,
when some unexpected coincidences create a stimulus that results in the adoption of a wrong hypothesis that lasts for several seconds. In such cases the observer has to invoke top-down processes to construct a cognitively correct hypothesis that is finally verified and adopted. An example is given in Sekuler and Blake (1994: 82): one of the authors woke up and mistook the pattern of light on his wife's hair, formed by a narrow beam of light through the window, for a sudden discoloration of her hair.

There are more extreme manifestations of the role of concept-driven processes, in which they almost override the input from the data-driven sensory organs. An example of this extreme is provided by Johnston and Hawley (1994: 69): one of the authors "saw" his cat on a couch for a moment, before realizing that he was looking at a stack of papers. I myself have also had examples of such extreme percepts: during the first night in a hotel room, I was woken by a thunderstorm, went to the window, and mistook the patterns formed by the tree-tops for puddles on the parking lot. I was immediately "transported" mentally to the ground floor, although I knew I had checked into a room on the third floor. My perceptual set influenced my mind-set. The combination of my sleepy state, the unfamiliar environment, and the stimulus's mild ambiguity contributed to the wrong hypothesis. Most of my similar extreme perceptual errors were obtained in hypnagogic or hypnopompic states, which are the states of drowsiness immediately preceding sleep and just before full awakening, respectively (notice that this was also the case in the example of Sekuler and Blake (1994: 82) in the previous paragraph). This can be taken as evidence that we need a "fully awake" concept-driven mechanism in the brain to arrive at accurate percepts of the environment, and it argues for the view of the brain as a hypothesis-constructing-and-testing agent.

It is reasonable to argue that the brain needs two conflicting characteristics in its effort to provide organisms with very accurate representations of the environment. Grossberg (1987) termed this conflict "the stability/plasticity dilemma": the brain needs stability to maintain continuity across time for expected, predictable inputs; at the same time, the brain must have plasticity and agility to deal with sudden changes and unexpected inputs. In a recent paper, Johnston and Hawley (1994) review evidence from their experiments which indicates the following: (1) Observers localize objects more easily in displays with exclusively familiar objects than in those with exclusively novel objects. They argue that this baseline effect is a manifestation of the brain's preference for stable, expected inputs. (2) Familiar objects are easier to localize in displays with exclusively familiar objects than in ones with one novel object among familiar ones, which they termed "familiar sink-in". (3) By contrast, novel objects are easier to localize when they are the only novel item in the display than when all objects are novel, the so-called novel-pop-out effect. The last two effects seem to argue that the brain is more attuned to novel, unexpected input, which is contrary to the tendency displayed by the baseline effect.
It appears that the brain is simultaneously biased toward familiar and unfamiliar inputs. Johnston and Hawley review evidence from a wide variety of sources, which provides additional support both for the perceptual inhibition of expected inputs and for the perceptual facilitation of novel inputs. They proposed the mismatch theory (Johnston and Hawley 1994), which explains these biases by postulating the following: The bias toward expected inputs is due to concept-driven processes which suppress data-driven processing of expected stimuli. The bias toward novel inputs is a consequence of this same suppression, which free up data-driven mechanisms to deal with unexpected stimuli. Their neural model implementing some aspects of the mismatch theory is dealt with in section 4.

Johnston and Hawley's mismatch theory is only one of many models that attempt to explain these characteristics; some of these models will be discussed in section 4. Similar paradigms regarding perception as a continuous interaction of data-driven units that provide the raw input and concept-driven processes that provide the hypotheses to be tested against the data were proposed a long time ago (e.g., Sokolov 1963; Gregory 1970). Sokolov (1963) formulated the dishabituation theory, based mainly on evidence from the orienting response, which is the spontaneous reaction of subjects to orient themselves toward novel or unexpected stimuli. One of the main tenets of his theory, shared with the mismatch theory, as well as Grossberg's (1976) adaptive resonance theory (ART), discussed in section 4, is that the output of data-driven processes is enhanced for unexpected inputs and inhibited for predictable ones. On the other hand, there is another group of theories, exemplified by McClelland and Rumelhart's (1981) interactive-activation model of perception. There is a substantial difference between the two groups of theories: the latter propose that data-driven processing is suppressed for unexpected inputs and enhanced for expected ones. Despite their differences, all these theories involve the formation of perceptual hypotheses, or schemata, based on environmental input from data-driven mechanisms and continuous interaction of the schema-driven and data-driven mechanisms for the formation of the final percept.

3 Evidence for Interaction of Data-driven and Concept-driven Processes

As mentioned above, the interaction of data-driven and concept-driven processes is not evident in everyday life; instead, most of us have the illusion of immediate, stable percepts all the time. In the previous section, examples were presented of cases in which this interaction is manifested in everyday life. Additional examples are provided in the next three subsections for three special classes of visually ambiguous stimuli: multistable, concealed, and impossible figures.
Figure 7.1. Classical multistable figures: (a) Necker’s cube; (b) Mach’s book; (c) Rubin’s vase/faces.

3.1 Multistable figures

In multistable figures, the physical stimulus is fixed, but it can give rise to a multiplicity of almost equi-probably stable percepts. Stadler and Kruse (1995) have compiled a list of categories of multistable figures that include (i) fluctuations of complex patterns, (ii) figure - ground reversals (see fig. 7.1c), (iii) figures with multiple symmetry axes, (iv) two-dimensional figures that give rise to multiple three-dimensional interpretations (see figures 7.1a and b), (v) actual three-dimensional objects that allow multiple interpretations, (vi) moving stimuli with multistable direction percepts, and (vii) drawings with multiple meanings (see figure 7.4). Examples of multistable percepts from fixed stimuli also exist in audition (Deutsch 1975).

Some early examples of multistable visual stimuli that have been established as classics are shown in figure 7.1: figure 7.1a shows the Necker cube (Necker 1832), which can be perceived in two stable 3-D states; figure 7.1b is the two-dimensional version of “Mach’s book” (cited in Stadler and Kruse 1995), which can be viewed as a book with its spine toward or away from the viewer; figure 7.1c is Rubin’s (1921) "vase/faces" reversing figure-ground image, in which the alternating percepts are the center vase and the two faces. A stimulus which has characteristics of both types (i) and (ii) above appears in Papathomas (1995). Factors that influence which parts of an image are most likely to be considered as figure and which parts as ground are reviewed by Kanizsa and Luccio (1995).

When presented with multistable stimuli for long intervals, observers experience spontaneous reversals among the stable percepts: one percept is adopted for a while, suppressing the alternative interpretation(s), then is supplanted by another competing percept, and the random alternation continues forever. It is as if the brain is simultaneously accepting a hypothesis and also getting ready to reject it as soon as contrary evidence warrants it, reminiscent of the seemingly conflicting simultaneous biases of the brain toward expected and unexpected inputs (Johnston and Hawley 1994). The theory that views the brain as a hypothesis-constructing-and-testing agent is one possible explanation of this continuous alternation: namely, that this
alternation occurs, even though there is one fixed stimulus, because the brain
keeps constructing and adopting different hypotheses that are consistent with
the visual input. Some researchers have suggested that this alternation could
be explained by selective adaptation and fatiguing mechanisms. This view
holds that different mechanisms are responsible for different stable percepts,
and they become fatigued when their preferred percept dominates for some
interval; at that point an antagonistic mechanism takes over, and the process
continues *ad infinitum* (von Griinau et al. 1984; Long et al. 1992). The
advantage of well-designed multistable figures is that they allow the
researcher to conduct experiments in which the time course of the putative
adoption of alternative perceptual hypotheses, or schemata, can be observed
and measured.

Mach's book of figure 7.1b exhibits bi-stability even in its 3-D implementa-
tion (category (v) of Stadler and Kruse 1995), which is obtained by folding a
rectangular piece of cardboard in half along its long dimension, as suggested in
the figure, and setting it upright on a flat surface such as a desk. Suppose that
the "book" is set with its "spine" away from you. If you now close one eye, so
as to suppress stereoscopic vision, and stare at the object, your percept will
eventually change spontaneously from the veridical one to a "false" stable state
with the spine toward you. The fact that this perceptual change is accompanied
by changes in apparent brightness of surfaces prompted Gregory (1997) to
suggest a top-down signal modifying primary processing and to introduce the
idea of "sideways" rules, as distinguished from schema-generating knowledge.
According to Gregory (personal communication, 5 Feb. 1998) both "sideways"
rules and knowledge are top-down and cognitive, but they are different, rather
like syntax and semantics. The two states of the Mach book will continue to
alternate as long as you view the object. What is remarkable is that, if you move
your head left and right while you perceive the false state, the object will appear
to move, as if suspended in the air, as long as you maintain the false state; the
moment you obtain the veridical percept the object will appear stationary, as
indeed it is. Explanations of this latter phenomenon have been provided by

Another example of an actual 3-D multistable figure, also belonging to a
category (v) of Stadler and Kruse (1995), is obtained by viewing the inside of
a hollow mask of a human face, painted on the inside (concave) surface with
natural human features. The hollow mask is seen as a natural convex face (the
*false* state), even when viewed with both eyes, as long as the viewing distance
is reasonably large, say two meters or so. It appears in its false state even more
easily, and from shorter distances (one meter or less), when viewed with only
one eye. What is remarkable in this case is that, as a rule, people cannot
perceive the hollow mask unless they come very close to it, and even after they
perceive it as a hollow mask, once they move far enough away, a stable percept
of a regular face is restored again. This is a clear case of a top-down process
(experience, knowledge) overcoming bottom-up signals to form the percept.
The reason why the false state is more difficult to perceive with binocular than
with monocular vision is most likely that stereopsis provides an additional bottom-down signal to be overcome by knowledge. As with Mach’s book, when the viewer moves while maintaining the false perceptual state, the "face" appears to turn toward the observer (Gregory 1970: 128).

An interesting tri-stable ambiguous figure (Fisher 1968), together with sketches of its possible interpretations, is shown in figure 7.2. This is one of the figures used by Stark and Ellis (1981) to test the role of fixational eye movements in the perception of ambiguous stimuli. While observers viewed the top figure, the experimenters recorded their eye movements, as well as the observers' reports about which of the three percepts of the bottom panel they were experiencing. By analyzing the data, Stark and Ellis (1981) concluded that the fixational pattern of the observers during the time of any given percept was well correlated to the corresponding perceptual hypothesis. They obtained similar results with Necker-cube stimuli. Gale and Findlay (1983) conducted a similar study with the so-called "daughter/mother" bi-stable figure that was introduced to the psychology literature by Boring (1930). They observed that the most important factor in the perception of multistable figures is selective visual attention, and that eye movements are important because they play a significant role in the deployment of attention. They also report that the eye fixational location pattern correlated closely with the dominant percept, but that percept alternation does not necessarily occur right after a major change in fixation location. The overall conclusion of Gale and Findlay (1983) is that saccadic eye movements are not exclusively determined
by the perceptual hypothesis, or by the stimulus characteristics, but by the interaction of the two factors.

It should be mentioned that the reversals of perceptual states cannot be attributed solely to the effect of eye movements, because such reversals occur even with paralyzed eye muscles (see Zinchenko and Vergiles 1972). However, one may argue that the image appears to move when a viewer whose eye muscles have been temporarily paralyzed attempts an eye movement. It turns out that the reversals still occur when retinal after-images are employed (Magnussen 1970). Since this method produces stabilized retinal images by construction, it rules out eye movements as the exclusive source of the alternation of percepts. Finally, we note that the cyclopean versions of these ambiguous figures also result in reversals of percepts (Julesz 1971: 40). This is to be expected, because the “cyclopean retina” (Julesz 1971) is very early in the visual pathway (Poggio 1995), well before the high-level site(s) of the brain where perceptual hypotheses are putatively generated.

3.2 Ambiguous concealed figures

Typical well-known examples of such figures are the “concealed Dalmatian dog” (Sekuler and Blake 1994: 137) and the “concealed cow” (Dallenback 1951; also in Schiffinan 1996: 190-2 and in Sekuler and Blake 1994: 15-16). Some of these figures are inherently ambiguous by coincidences in the natural world. Others are obtained by degrading originally unambiguous images; this can be done by adding visual noise, or by “thresholding” (assigning black or white to all pixels below or above a certain intensity, respectively), or by fragmenting their parts, etc. Figure 7.3 is meant to illustrate how extremely ambiguous...
difficult these figures are to identify/recognize. The semantic content of figure 7.3 is revealed in the next paragraph, to give the reader who sees it for the first time the opportunity to appreciate the process of arriving at a meaningful percept. Upon first seeing this figure, you catch yourself making cognitive/perceptual hypotheses (Is it a landscape? Is it a face? Is it an animal?) and testing them against the visual input. The time required for recognition varies widely across individuals. In anecdotal tests, the author has found some people who identified the figure in a matter of a few seconds and others who gave up after repeated attempts, each lasting several minutes. People in the latter group achieved recognition only after shown a visual sketch of the actual concealed figure.

Stark and Ellis (1981) used figure 7.3 to study saccadic eye movements before and after identification of the concealed figure. Observers viewed the figure, and were asked to signal the moment they achieved recognition. By the way, this figure shows a man with a mustache, beard, and long hair; he is shown from the chest up, centered in the picture, with half of his forehead clipped off at the top border of the picture (Porter 1954). The results obtained by Stark and Ellis (1981) show a significant change in eye movements after recognition. They start as a nearly random pattern of eye movements during the 75 seconds preceding identification, and they change to a well-recognized scanpath that visits the salient features of the face during the 75 seconds following identification. Stark and Ellis (1981) interpret these results as evidence that cognitive, schema-driven models direct eye movements. Figure 7.4 is another example that has elements of both bi-stable and concealed figures. It has an obvious stable percept of an apple core seen against a background composed of tree branches and a blue sky. The concealed figures are easier to find if the viewer knows the title of the painting. Again, in the interest of allowing the viewer time to appreciate the perceptual process, the title is given in the caption of figure 7.5.

It’s worth mentioning that similar effects are also obtained in audition; for example, Deutsch (1972) has shown that a familiar tune is not recognized if each note in the sequence is selected randomly from one of three adjacent octaves. However, when observers are primed about the tune, they find it much easier to recognize it.

3.3 Impossible figures

Another class of figures with which one can "catch" the brain in the hypothesis-constructing- and-testing mode is the class of impossible figures (Simon 1967). These are images of "objects" that cannot physically exist, because they violate global constraints of physical objects, even though they are locally coherent. Two classical examples are shown in figure 7.5: Figure 7.5a displays the so-called devil’s fork (Ernst 1992); and the impossible triangle of Penrose and Penrose (1958) is shown in figure 7.5b. A remarkably rich collection of
impossible figures was compiled by Ernst (1992, 1996). The work of M. C. Escher (1967; see also Ernst 1976) contains numerous examples of ingeniously designed impossible figures. When one stares at impossible objects, one is aware of being in an endless loop of constantly constructing hypotheses based on local features, in an attempt to resolve global inconsistencies and arrive at a percept of a physically realizable object; however, one constantly
finds her/himself rejecting each global hypothesis, to give way to alternative hypotheses that are also rejected in turn.

In addition to phenomenological differences, there is a fundamental difference between ambiguous figures of the types discussed in sections 3.1 and 3.2 on the one hand, and impossible figures, on the other. The former cause perceptual alternations by virtue of the image features that the viewer attends to, but the ensuing percept accounts for the entire image, with parts of the scene perceived possibly as background. By contrast, the perception of impossible figures depends critically on the focal point. For example, when you look at the right part of figure 7.5a, you see a coherent solid object bounded by planar surfaces, and two poles of square cross sections extending to the left; at the same time, the left part of the figure, seen in the periphery, appears indistinct. As soon as you stare at the left of the picture, you perceive three cylindrical rods extending to the right. What is a material object for the three rods is empty space for the two poles, and vice versa. The impossibility arrives when one tries to integrate the two contradictory percepts. The conflict is obvious even in the above verbal description. You cannot "bind" two square-cross-section poles with three cylindrical rods.

4 Neural models

In addition to occupying vision researchers, the perception of ambiguous multistable figures has attracted the attention of theoreticians and physicists who deal with multistable phenomena in physics (Haken 1977) or chemistry (Plath and Stadler 1994). A computational model based on synergetic systems was developed by Ditzinger and Haken (1989) to account for the temporal aspects of the alternation with multistable figures. It obtains respectable fits to empirical data by mimicking the effect of saturation of attention. Johnston and Hawley (1994) also developed a simple neural model to implement some basic characteristics of their mismatch theory. The neural model assumes just two layers of nodes: a data-driven layer of iconic nodes, which are activated by excitatory bottom-up connections but inhibit each other through lateral connections. The next layer contains two classes of nodes: location and identity nodes, which are activated by the outputs of the iconic nodes and in turn send inhibitory signals to the iconic nodes via a feedback pathway. Additional excitatory and inhibitory connections are assumed by the model, and the simulations are in close agreement with the experimental results.

Grossberg's (1976) adaptive resonance theory (ART) was generalized by Carpenter, Grossberg and Rosen (1991) using fuzzy set theory to handle analog, in addition to binary, input patterns. The model is based on the premise that the input received from early data-driven sensory organs, appropriately coded, triggers top-down processes, which in turn constantly construct hypotheses, or schemata; an interaction ensues between top-down and bottom-
up processes, to test these alternative competing hypotheses, rejecting the weak ones and adopting the strongest one to arrive at a stable state. Accordingly, in *Fuzzy ART*, as the authors call the new model, the sensory input $I$ generates an activity pattern and transmits it to a first-layer (mostly data-driven) mechanism $F_1$, which in turn transforms it and provides input $T$ to the top-layer (concept-driven) mechanism $F_2$. Activation of $F_2$ may be viewed as "making a hypothesis" about the input. In turn, $F_2$ generates its own top-down signal $U$, based on $T$, which it transmits back to $F_1$; this can be viewed as "testing the hypothesis." This two-way interaction between $F_1$ and $F_2$ continues until the signal received by $F_1$ and $F_2$ (expected input, according to the hypothesis of $F_2$) matches the actual input $I$, in which case the input is classified as per the previous training of the network. If a match cannot be found, it means that this is a novel input schema, and it must therefore be added to the repertoire of the available hypotheses.

Fuzzy ART was further extended by Aguilar and Ross's (1994) *Incremental ART* model, which incorporates ideas for efficient classification of input patterns based on partial feature information. The algorithm proceeds in three basic steps: (1) A partial set of features is extracted from the image, and supplied as input to the $F_1$ layer. (2) The interaction of $F_1$ and $F_2$, as outlined in the paragraph above for ART, tries to classify the image on the basis of the partial input; if it succeeds in converging, the process is over when a match is found. (3) If no convergence is possible, Incremental ART determines which feature to extract next from the image and repeats the process. The classification of features is accomplished by representing each feature by an ON channel, which is active (1) when the feature is present, and an OFF channel, which is active (1) when the feature is entirely absent from the image. In Fuzzy ART, all features are extracted first, and the absence of a feature from the input stimulus is coded as 0 in the ON channel and as a 1 in the OFF channel. However, Incremental ART can work with a partial list of features, which means that a feature can be absent because it has not *yet* been extracted from the image; in this case the states of both the ON and OFF channels are 0. Table 7.1 summarizes the three possibilities.

The major advantage of Incremental ART over Fuzzy ART is computational efficiency. Convergence may be reached with a partial feature vector, which means that the time required for extraction of the rest of the features has been saved. After all, humans also often achieve recognition based on partial input,
giving this model biological relevance, as well. In addition, the decision on which features to extract next, when the image cannot be classified on the basis of the presently available feature vector, offers the possibility of modeling saccadic scanning patterns. It is this last property of the model that allows it to mimic alternations between perceptual states in multistable figures, including the role of fixational eye movements in influencing this alternation process (Aguilar and Ross 1997).

5 Discussion and Conclusions

This chapter has reviewed evidence that demonstrates the plausibility of complex interactions between concept-driven and data-driven processes in the perceptual act. In particular, the brain's property of simultaneously showing a preference for both expected and unexpected stimuli can be explained by concept-driven processes exerting an inhibitory effect on the processing of expected inputs. We also reviewed studies with three main classes of visual stimuli, which support the theory that concept-driven components in the brain constantly construct perceptual hypotheses which are tested against the outputs of data-driven early mechanisms. Finally, we discussed computational models that attempt to emulate the view of the brain as a perceptual hypothesis-testing agent, with results that are in adequately close agreement with empirical data.

Another area that can offer support for this theory of the brain as a "hypothesis-constructing-and-testing engine" is what happens in dreams. During sleep, the brain has no sensory visual stimulation to drive its construction of hypotheses. Thus, it can roam free in an open-loop mode and can construct its own "reality" which is devoid of the real-world visual feedback, constrained by physical laws, that would be provided by the eyes in a fully awake state. This may well explain the fantastic quality of most dreams. But how about the other senses that can still provide physical stimulation even in sleep? For example, how about auditory input? It turns out that there are cases where, in accordance with the hypothesis-constructing-and-testing theory, dreams are "molded" to account for sounds experienced during sleep. As a striking example, I once dreamt about being in a street where a worker was operating a loud pneumatic drill, only to wake up and realize that a woodpecker was pecking on my home's wooden shingles. Apparently, the brain constructed a "reality" that incorporated the physical sound, driven by the auditory input. Parenthetically, a view has been proposed that such dreams are designed to protect us from waking up when we are disturbed by strong stimuli during sleep (Horne 1988).

The theme of top-down assumptions and expectations influencing early mechanisms is common. For example, Pinker (1997: 213) attributes visual illusions to conditions that violate the visual system's assumptions. He also
suggests that the visual system may compare the probabilities of different hypotheses/percepts that are consistent with the raw data (ibid., 243). Adelson (1997) has identified some top-down processes that affect the apparent lightness of surfaces. These processes are not necessarily driven by schemas or hypotheses, but originate instead at some intermediate level, by computing gray-level statistics over some local area. He found evidence that these mechanisms compute their statistics over areas that are consistent with natural "atmospheric" boundaries, as signaled by T-, X-, and psi-junctions, and use this input to estimate the illumination level, haze, and properties of interposed filters (Adelson 1997).

Finally, we note that stimulation by binocularly rivalrous stimuli has several important similarities to stimulation by ambiguous multistable figures, but the difference is that, in rivalry, the two eyes are presented with different visual inputs. Chapter 9 below, by Papathomas, Kovacs, Feher, and Julesz, deals with how the brain reacts when stimulated by binocularly rivalrous visual inputs. It offers a view of binocular rivalry which is compatible with the theory that the brain uses high-order processes to construct and test possible interpretations of the available stimulus in an attempt to arrive at meaningful percepts.

It should be noted that the issue of whether concept-driven processes can affect data-driven mechanisms in vision is still under debate. This chapter has presented a partial review of the literature that interprets the evidence in favor of an affirmative answer. The other view has been argued forcefully through the years; an excellent review that offers strong arguments for the opposite view was given recently by Pylyshyn (1999), who concludes that visual perception is cognitively impenetrable (see also Fodor 1983). At this point in time, it seems that the evidence is equivocal, and the debate is far from over. Additional work is needed to investigate this important issue in perception/cognition.

Acknowledgments

I am truly indebted to Mario Aguilar for enriching this chapter with invaluable references on a wide range of research in multistable figures and for introducing me to his computational model. I was fortunate that Richard Gregory and Diana Deutsch were kind enough to read an early draft and to provide me with expert advice. I wish to thank Bela Julesz, Ilona Kovacs, and Akos Feher for useful discussions and ideas. My good friend Charalambos (Harry) Constantinides, who appreciates the classics more than anyone I know and has quite a broad knowledge of the classic literature, provided me with valuable references on the liar’s paradox. Tiffany Conway helped me to clarify many points in the chapter. Many thanks to the McDonnell Foundation for supporting the research at the Laboratory of Vision Research through grant 95-60.
Notes

1 The sophisticated reader is asked to excuse the oversimplification in partitioning the brain into sensory, perceptual, and logical "components."

2 The notation of Johnston and Hawley (1994) will be used here. The terms data-driven and concept-driven will refer to the origin of the information; the former will indicate early sensory mechanisms that receive input from the external world, the latter, mechanisms that are provided input by higher-order cognitive processes. The terms bottom-up and top-down will be used to signify the direction of the information flow, the former going from early, low-level units to later stages, the latter in the opposite direction.

References


