

Reentrant Neural Pathways and the Theory-Ladenness of Perception

Introduction

The issue of the role of top-down, or reentrant, neural pathways that transmit signals from the higher cognitive centers to the perceptual modules is important to the philosophy of science, since it is the existence of these pathways that is used as one of the arguments for the cognitive penetrability, and thus for the theory-ladenness, of perception (Churchland, 1989), clearing the way for the relativistic theories of meaning and of scientific theories.

The argument consists in evidence concerning top-down links from higher cognitive centers to the peripheral systems. Findings from cell-staining techniques show that the ascending pathways from the retinal to the geniculate nucleus (*LGN*) and from there to the visual cortices and to other centers higher in the processing hierarchy are matched by descending pathways from the higher level of processing to even the earliest processing systems at the retina (Zeki, 1978; Van Essen, 1985). The function of these descending pathways is “centrifugal control”, that is, the modulation of neural activity in the periphery by the demands sent down from higher cognitive levels (Churchland, 1989).

In this paper I argue for the impenetrability of perception by undermining the argument from reentrant pathways. To do that I adduce psychological and neuropsychological evidence showing that (a) the early vision processing is not affected by our knowledge about specific objects and events and (b) that the role of the descending pathways is to enable the early-vision processing modules to participate in higher-level cognitive functions. Then I discuss evidence from connectionism showing the way the brain could implement a process in which peripheral systems are the locus of higher cortical functions. This same evidence suggests that the perceptual input modules are necessarily involved in higher

cognitive tasks. This is not a proof of the cognitive impenetrability of our perception (to do that one must account for illusions and other evidence suggesting the cognitive penetrability of perception) but it is a step towards it.

My thesis is that a part of observation, which I will call perception, is bottom-up and in some important sense theory neutral. Thus, it can play the role of common ground on the basis which a naturalized epistemology could be built and relativism be avoided. I will not elaborate here on the form of the product of perceptual processes except to say a few things at the end of the paper, addressing Churchland's (1989) claims that even if there is some theoretical neutrality at an early perceptual process, this "pure given" is useless and cannot be used for any "discursive judgment", since sensations are not semantically-contentful states. I will argue that though the theory-neutral given does not have, and should not have, a semantic content, it has a nonconceptual content in which cognition is grounded.

2. Observation and perception: an issue of terminology

I have spoken of perception and observation, or vision. These terms are not employed consistently in the literature. Sometimes "perception" purports to signify our phenomenological experience, and thus, includes the recognition and identification of objects and events. Since I do not use the terms the same way, I adopt Dretske (1985) and Shrager's (1990) usage, I will introduce some terminology to explicate my usage of the terms.

I call *sensation* all processes that lead to the formation of the retinal image. This image, which initially is cognitively useless, is gradually transformed along the visual pathways in increasingly structured representations that are more convenient for subsequent processing. I call these processes that transform sensation to a representation that can be processed by cognition *perception*. Perception includes both low-level and intermediate-level vision and is bottom-up. In Marr's (1982)

model of vision the *2 $\frac{1}{2}$ D sketch* is the final product of perception. All subsequent visual processes fall within *cognition*, and include both the post-sensory/semantic interface at which the object recognition units intervene as well as purely semantic processes, that lead to the identification of the array (high-level vision). At this level we have observation (Marr's *3D model*), which is a cognitive activity.

Before I proceed I would like to explain the qualification "knowledge about specific objects". Even if perception turns out to be bottom-up in character, still it is not insulated from knowledge. Knowledge intrudes on perception, since early vision is informed and constrained by some general world principles that reduce indeterminacies in information (mainly the underdetermination of the *2 $\frac{1}{2}$ D* structure from the *2D* retinal stimulation). They are general assumptions about the world constraining visual processing (Marr, 1982; Spelke 1990; Pylyshyn, 1999; Ulmann, 1979). These principles however are not the result of explicit knowledge acquisition about specific objects but are general reliable regularities about the optico-spatial properties of our world hardwired in our perceptual systems.

This knowledge is implicit, in that it is available only for the processing of the retinal image, whereas explicit knowledge is available for a wide range of cognitive applications. Implicit knowledge cannot be overridden. The general constraints hardwired in the visual system can be overridden only by other similar general constraints with which they happen to compete (although no one knows yet how the system "decides" which constraint to apply). Still, one cannot decide to substitute it with another body of knowledge, even if one knows that under certain conditions this implicit knowledge may lead to errors (as is the case with the visual illusions). This theoretical ladenness, therefore, cannot be used as an argument against the existence of a theory-neutral ground, because perception based on a shared theory is common ground.

3. The role of reentrant connections

Leopold and Logothetis (1996) studied in animals the activity of neurons in areas ranging from the primary visual cortex, where retinal signals first enter the brain, to the area called *IT*, which is the very end of one fork of visual processing. Their study showed that in the primary cortex only 18% of neurons changed their response according to the image perceived by the animal. In areas corresponding to the midway of visual processing about one half of the neurons changed their response. In the *IT* area almost all neurons did. These findings suggest that most of the neurons in early processing report information that can be extracted from the information recorded on the retina and are not influenced by the higher cognitive functions, that is by what the animal perceives the object to be. Other studies of visual perception by Perrett, *et. al.*, (1990) show that cell firing in the temporal cortex is not modulated by the significance of the event for the animal. But then what is the function of the descending pathways?

The answer comes from cognitive neuroscience. Research (Posner and Petersen, 1990; Posner and Carr, 1992; Kosslyn *et. al.*, 1993; Posner and Raichle, 1994; Heinze *et. al.*, 1994; Ziegler, *et. al.*, 1997) with positron emission topography (*PET*) and event-related potential (*ERP*) provides a spatio-temporal picture (literally) of the brain of subjects while they are performing (a) bottom up processes, such as passive visual tasks (viewing on a screen strings of consonants, words, and pseudowords), (b) processes that require some top-down influences, such as active attention-driven tasks (searching visual arrays for thickened letters), (c) processes that rely heavily on top-down semantic processing (generating a use in response to a visual word), and (d) processes that are purely top-down, such as imagery. This picture sheds light on the role of top-down pathways and supports the principle of modular design and the independence of low level visual processing from top-down influences.

In studies of passive visual tasks subjects were asked to fix their gazes on a point in the middle of a monitor, in which four kinds of complex stimuli were to appear: false fonts, letter strings, pseudowords and words. PET scans provided a pictures of the activation of visual areas in the brain during these tasks. The analysis of these pictures relied on the assumption that the visual stimuli consisted of four codes. First, the 'words' presented were complex collections of visual features, second, these features were aligned to form the letters of the English alphabet, third, some of the 'words' had forms that satisfied the rules of English language, that is, they were English words, and fourth, some of these words had meanings.

The responses observed were responses to some, or all, of the four codes. All four groups produced bilateral responses in multiple areas of the visual system. The subtraction of the PET images when the brain processes the visual features of the array from the PET images in semantic processing shows that only words and pseudowords produced characteristic responses in the inner surface of the left cerebral hemisphere, an area which is related to semantic processes. This suggests the existence of two levels of analysis in the visual system. The brain initially analyzes the visual features of the stimuli regardless of relationships to letters and words. At a second level, the brain analyzes the visual word forms. The fact that the subtraction of the PET images reveals an intense activity in the left hemisphere when semantic processing is taking place, however, does not mean that semantic processing is localized at that area only. The method of subtraction only highlights areas that are activated in the one task but not in the other. It does not reveal the entire area of semantic processing. In fact, we know that areas in both hemispheres are related to semantic activity (for a criticism of attempts to localize brain functions by using imagining techniques see Bechtel and Mundale, 1999).

More interesting were the PET images obtained in the active attention-driven visual tasks and in tasks of visual imagery. In the active attention-driven visual

tasks, subjects were presented with succession of images on a screen and were asked to react whenever some attributes (color, motion, form) were different from one image to another (focal attention groups). The passive control group were instructed to watch the screen without reacting. The divided attention group, finally, were instructed to react to any changes whatsoever in the images.

The PET images of the passive group showed activations of areas traditionally associated with registration of color, motion and form in the extrastriate cortex. The subtraction of the divided attention PET images from the focal attention PET images allows the isolation of the areas that compute the specific features of the focal attention groups. The results were clear. Attention enhances blood flow at the same areas which are activated during the passive tasks. Thus, the same areas that process information in the passive tasks, process information in the active attention tasks, only this time their activation is amplified. The subtraction of the PET images in the passive acts from the PET images in the focal attention tasks allows us to track those areas (if any) outside the visual areas that are also activated only during the focal attention tasks and not during the passive tasks. Indeed, there were found areas in the basal ganglia and the anterior cingulate gyrus (an area at the underside of the frontal lobe). These areas seem to be the sources of the amplification observed when attention is involved and it is likely that they constitute the attentional networks activated in the focal group conditions.

Similar results were obtained with the visual imagery tasks. Visual imagery activates the same brain areas as visual perception (Farah, 1984). Behavioral studies suggest that the processing of imagery and of visual perception share some mental operations. Studies with patients (Kosslyn, 1988) show, first, that the mental operations that support visual imagery are highly localized (are carried out in precise locations) and distributed in many different areas in the brain. Second, many of the neural systems at which mental images are generated are the same as those

activated during visual perception. Neuroimaging studies confirm these results. The subtraction of PET images during passive control tasks from the PET images in imagery tasks and from the PET images in visual perceptual tasks shows similar activations in imagery and perception, especially in posterior areas.

The PET studies were complemented by ERP studies of subjects who view words and consonant strings. Certain areas in the brain are activated about 100 ms after the word or the string is presented. Since these areas are activated irrespective of the stimuli, it can be surmised that they are activated by the features that words and consonant strings share, namely, visual physical features. Differences in the responses to words and consonant strings started about 150 ms after the stimuli appears. This means that the brain registers the word form 50 ms later than the visual features. What is important to note is that the *ERP* study shows that the distinction between words and consonant strings is not fed back by other higher processing areas but arises at the posterior sites of the cortex.

In other ERP studies, subjects were asked to search for a thickened letter in letter strings. This is clearly an attention-driven task, in which one would expect to find some top-down, task-driven, processes. Records of the electrical activity during the search show that this top-down activity involves the same processing areas that are involved in computing visual features. But the search for the thickened letter causes activity in these same areas only about 200 ms after the stimulus (recall that the activity recorded when these sites register the visual features takes place 100 ms after the stimulus). Thus, the computations involved in the top-down attention driven tasks take place in roughly the same brain areas— the same electrodes are activated, (predominately right posterior areas)— in which the bottom-up registration of visual features occurs, with a time delay of about 100 ms. Finally, similar studies of subjects performing semantic tasks, such as generation of the use of a noun, showed that word meaning is registered about 250 ms after presentation

of the stimuli, and some of the areas activated are the same with those areas activated when processing of visual physical features takes place.

Let me redraw the picture. 100 ms after the presentation of the stimulus (letter strings) an extensive part of our brain responds to the physical characteristics of the visual array. 150 ms after the stimulus these features fuse to a single form, and about 200 ms after it the voluntary task-driven search is registered in the same areas that process the visual features. Thus, the top-down effects of attention are delayed in time, involve the same anatomical areas as passive perception, except that attention amplifies the recordings in these areas. Finally, about 250 ms after the stimulus, some of the same areas participate in the semantic processing of the input.

What do these PET and ERP findings suggest for our discussion? Tasks that require top-down flow of information activate broadly the same areas that are needed to compute the purely bottom-up tasks. The active attention studies showed that when top-down processes occur, the activation of these areas is enhanced and the source of this amplification is higher areas in the brain. In order for the factors that cause this amplification to be transmitted to the lower areas certain descending pathways are required. The same conclusion can be drawn from the visual imagery studies. Visual imagery demands that activation originate from higher cognitive centers and descend to the visual cortex in which imagined images are formed.

How does the brain do this? The answer is found in the role of the *reentrant connections* or *mappings* among neurons in the brain, and explains the abundance of interconnections between neurons. These reentry connections map, as it were, the activity of any system onto the others and reciprocally, by allowing the transmission of information in all directions.

The point thus is that imagery and perception share some processing sites. The same conclusion can be drawn from the attention-driven tasks. The amplification implies that the “attention centers” modulate the processes that take place in the

relevant peripheral sites. This modulation, however, may be explained in two ways, both of which do not entail cognitive penetrability of perception. One is to say that these centers allocate attention to locations or properties of the upcoming visual scene prior to perception (and hence no cognitive penetrability of perception occurs) increasing sensitivity to a particular class of upcoming stimuli.

The other way concedes that the foregoing constitute evidence for task-dependent top-down processes that influence perceptual processes. But these are not influences by specific object knowledge. Active attention determines the kind of the task to be performed. In that respect Hildreth and Ulmann (1989) argue for the existence of an *intermediate level* of vision. At this level occur processes (such as the extraction of shape and of spatial relations) that cannot be purely bottom-up, but which do not require information from higher cognitive states. These tasks do not require recognition of objects, and thus any semantic processes that would have justified the cognitive penetrability of perception. They require the spatial analysis of shape and spatial relations among objects, as well as detection of motion, which are the most relevant physical properties that form the basis for focused attention (Egeth, *et al.*, 1984; McCleod *et. al.*, 1991). This analysis is task-dependent, since the processes involved may vary depending on the task being accomplished even when the same visual array is being viewed, but is not modulated by cognition.

4. The evidence from connectionism

I have argued that certain higher cognitive functions and the visual system share to a certain extent the same processing sites. Research with neural nets gives us an idea of how this can be implemented. Regier (1996) constructed a hybrid connectionist system which receives as input some simple geometrical figures in various static and moving spatial relations and the spatial relation terms for these relations (the term “on”, for instance, when a circle is on top of another figure). The

task for the network is to learn the spatial terms so that it can assign the proper concept for a novel spatial configuration.

Classical *PDP* networks could not handle the task. The hybrid model that learned the spatial concepts consisted of two parts. One was a *PDP* model that learned via back-propagation. The other was a network with a specific architecture (a structured network). Regier designed this second sub-network so that its architecture reflect that of the human visual system (the topographic maps of the visual field, the orientation sensitive cells, the center-surround receptive fields, top-down pathways).

The fact that a standard *PDP* network without a specific architecture could not learn the task and only the hybrid network could, indicates that the success of the latter may be due to the second sub-network, the one implementing the architecture of the actual visual system. This means that conceptual categories are created using the perceptual apparatus of vision, which implies that the higher concept-formation cognitive activity involves crucially a perceptual module. The same holds for imagery during which the visual system is employed to see things that are activated not from external input to the peripheral module but from top-down modulation. The qualification “crucially” means that the perceptual module is not merely a provider of input to, but it actively participates in, the cognitive processing. What connectionist research adds is the necessity of the peripheral module as a part of the higher cognitive process for learning to ensue.

Conclusion-Discussion

Consider the argument for cognitive penetrability. We know that there are many neural connections devoted to bringing information back to the sensory systems from higher cognitive centers. This constitutes evidence for the mediation of the output of the perceptual modules by information from higher cognitive states. But the descending pathways most likely have another role to play. The sensory systems

are fed back information from higher centers and signals from higher areas reenter the brain areas that had processed before the signals that were transformed by the higher centers to the reentrant new signals. The same areas that process in a bottom-up way sensory specific information are also involved in higher-level activities (voluntarily attention-driven search, imagery, concept formation), except that in the latter case they are reentered in a top-down manner.

The importance of the peripheral modules in the execution of higher cognitive functions not only undermines the descending pathways argument for the theory-ladenness of perception, but in addition shows that our conceptual systems are severely constrained by the architecture of the perceptual modules, since the cognitive processes that give rise to concepts involve in a significant way the perceptual processes. Perception does not serve only as the faculty that provides input to higher cognition and then comes on-line, after the cessation of the conceptual processing, in order to test empirically its outcome, but also constitutes an active participant of the conceptual processing itself.

Consequently, our discussion questions the standard distinction between conception and perception. According to this view, conception is thought of as a mental process, where as perception is deemed to be bodily in nature. The new picture emerging, drawing attention to the fact that perceptual processes are inextricably involved in higher cognitive processing, rejects this distinction and forces us to extend cognition to encompass the body, in so far as the perceptual bodily mechanisms do some conceptual work.

The perception-observation distinction runs along the distinction between *seeing something* and *seeing something as being such and such*. Perception corresponds to *seeing*, where as observation corresponds to *seeing as*, that is to identifying or recognizing visual patterns that are being *seeing*. Brown (1987), Churchland (1989), Gregory (1974), Hanson (1958), Kuhn (1962), and others argue that such a distinction

does not exist, and that all *seeing* is *seeing as*, since all stages of visual processing are cognitively penetrable. By undermining one of the arguments for this penetrability, and reinstating the old distinction back into life I add a step towards showing the cognitive impenetrability of some stages of vision. Consequently, the outcome of the perceptual processes can serve as the theory neutral (in the sense explained in the introduction) basis whose existence would undermine relativism.

I must note that the above distinction is somewhat simplistic, in that it leaves aside the issue of *vision for*, that is, the issue of the evidence regarding the existence of various (at least two) visual pathways with different outputs, each of which supplies different information to specific brain areas, such as the motor system and the system responsible for categorization (Goodale, 1995). Since, however, the controversy regarding the cognitive penetrability of vision concerns the extent to which the percepts used in cognition are informed from a top-down flow of information, I will restrict my discussion to the visual path that leads to the formation of these percepts.

Thus, Fodor's (1984) distinction between "fixation of appearances" or "observation", which is the result of the functioning of the perceptual modules, and "fixation of belief" is misguided. He seems to distinguish between the "sensory" and "cognitive" or "semantic" processes that are involved in the formation of observation statements and considers observation as a pre-cognitive activity whose output is processed by cognition giving rise to the observation statements. Philosophers would recognize here the distinction between what we see and how we perceive it to be, or how we interpret it. But this distinction is misleading because object recognition is a cognitive process, and observation involves object-recognition. The distinction Fodor wishes to draw between a bottom-up, theory-neutral, and a top-down, theory-laden, process should not be cut at the

observation/cognition interface, since such an interface does not exist, but at the perception/cognition interface.

Churchland (1989) claims that even if there is some rigidity and theoretical neutrality at an early perceptual process, this sensation, is useless in that it cannot be used for any “discursive judgment”, since sensations are not truth-valuable, or semantically-contentful, states. Only “observation judgments” can do that, because they have content, which is a function of a conceptual framework. I shall not argue here in detail whether this theory-neutral perceptual basis is semantically contentful and I will offer only a sketch of the kind of content of the theory-neutral perception.

This theory-neutral “given” does not consist in sensations without any content, but in perceptions. The perception of a visual scene is an image schema of this scene. This schema consists in some form of Marr’s $2\frac{1}{2}$ models of the objects present in the scene (Pylyshyn, 1999), that is, representations of shapes (structured representations of $2\frac{1}{2}D$ surfaces of objects), on the one hand, and on the other, in the rich kinaesthetic (Bickhard, 1993; Johnson 1987, Lakoff and Johnson, 1999) and positional (local) content-structure of the scene (Petitot, 1995). The positional content is not semantic (nonconceptual) and, along with the kinaesthetic structure, conveys information about nonvisual properties, such as causal relations (e.g., X “transfers” something to Y), and about functional properties of objects, what Gibson (1979) referred to as *affordances* of objects (for a similar conclusion see Pylyshyn 1999, 361). This content can be retrieved from the morphology of the scene and can be extracted from the visual array by the mechanisms of early vision (Petitot, 1995).

The nonconceptual content is a theme frequently discussed in philosophy and cognitive science. Consider the interaction of a frog with a fly (Bickhard and Campbell, 1996). The induced neural activity in the frog is not an internal representation of the fly. The “representational” content of the neural activity induced by the fly consists in the possibility of tongue flicking and eating on the

basis of indications about potentialities that are afforded by specific objects in the environment. Thus, representational content is about the potentialities, or possibilities of further interactions, that are afforded by the environment for the system's interactions with it. They implicitly predicate those interactive properties of the environment that could support the indicated interactions of the cognizer with it.

The mechanisms implementing these interactions with the environment are the hardwired, implicit, and inherent to the input systems, principles that constrain perceptual computations and coordinate perception with the motor systems. These principles are not explicitly represented by the system, as general beliefs about the world. Instead, they determine the *modus operandi* of the systems. As such, they account for reactions only to specific object tokens, not to object types, as it happens with our fly (for a similar analysis regarding the object-centered attentional indexing mechanisms see Scholl and Leslie, 1990). Findings regarding the different routes in vision support this analysis, suggesting that the content of the information transmitted via the route of vision for action to the motor systems is nonconceptual.

These interactive anticipations are representational for the system, being implicit predications about the environment. Cussins (1990) has called the content of such implicit representations "nonconceptual", by which he means those properties that are canonically characterized by means of concepts which are such that the organism need not have those concepts in order to satisfy the property. I am going to call such nonconceptual representations, pre-representations.

Thus, the perceptual outcomes are pre-representational, nonconceptual information-content bearers that constrain, and form the basis for, cognitive processing. Far from being without content or far from having a content that is a function of a conceptual framework, perception provides the background upon which conceptual frameworks are being build and is the locus in which meaning is infused into conceptual systems and the representations are grounded. Perception

must be pre-representational and not non-representational if its output is to be cognitively processed. The *pre-representational* character is due to its having a rich experiential structure; what makes perceptual schemata *pre-representational* is that they pick out the entities, actions, and events that fit our recurrent experiences; yet, they are not conceptual representations of these experiences.