Conscious and unconscious vision

Fahrenfort, J.J.

Citation for published version (APA):
Main thesis
This dissertation is about conscious and unconscious vision. When somebody walks into an elevator, the electronic eye at the entrance of the door takes notice and prevents the door from closing. However, no one will seriously claim that the sensation of an elevator as somebody walks in is the same as your own sensation when you see somebody entering an elevator. What makes us conscious of the things we see? And what processes take place outside consciousness and can be regarded as more similar to the way electronic eyes process visual information? The aim of this thesis is to formulate some answers to questions like these.

![Figure 1.1 Illustration of image complexity that the visual system needs to disentangle. a) The human subjects in this image show complex figure-background relations that are notoriously difficult to disentangle by computer algorithms. b) Even under completely altered conditions of light and color the human brain is able to effortlessly disentangle such complex relations.]

In order to make sense of our world, the visual system has to perform functions by which complex visual scenes can be disentangled, a problem that has proven to be notoriously difficult to
Chapter 1

tackle in computer vision (see Figure 1.1). An important step in this process is figure-background segregation, by which surfaces belonging to objects are labeled as distinct from their surrounding background. In another step, objects need to be detected and recognized. At some point during these processing routines, a conscious representation emerges. One of the key functions of consciousness is often considered (sometimes by definition) to be the overt detection and recognition of objects (e.g. Dehaene, Changeux, Naccache, Sackur, & Sergent, 2006). In many theories of visual perception, figure-background segregation is thought to be the first step in vision, preceding object recognition and conscious vision (e.g. Marr, 1982; Nakayama, Zijiang, & Shinsuke, 1995; Rubin, 1958). In such a theoretical framework, one may presume the neural correlates of consciousness to be more closely related to the neural correlates of object detection and categorization than to correlates of figure-ground segregation. Following this line of reasoning, the steps needed for the brain to reach a conscious state go from (1) figure-ground segregation to (2) object detection / categorization and consciousness.

In this thesis I attempt to demonstrate that this view is incorrect. It will be shown that the brain starts out to quickly and automatically detect and classify objects, even before a scene has been segregated into objects and background, and even when one is not conscious of the objects at hand. Correlates of conscious visual experience and figure-ground segregation, on the other hand, appear after this initial response, through the reactivation of early visual areas as a consequence of recurrent processing. This suggests that conscious vision may share its neural correlate with figure-ground segregation rather than with object detection and categorization. Consequently, in the proposed scheme the steps required to reach consciousness are (1) object detection / categorization to (2) figure-ground segregation and consciousness. If accepted, these findings require a revision of the way many scientists think about how the brain executes visual routines and reaches consciousness.

In the following sections a brief background on the coarse architecture and function of the visual system is given, followed by some discussion on the role of recurrent processing therein. The introduction ends with an overview of the contents of this thesis and how these relate to its central position.
**Architecture and function of the visual system**

After light has been transformed to nerve impulses in the retina, these impulses are relayed by the optic nerve fibers through the lateral geniculate nucleus (LGN) of the thalamus and onwards to visual area V1 in the occipital lobe of the cortex. From there cortical pathways continue on to other visual areas through two broad routes: the dorsal route to parietal cortex and the ventral route to inferotemporal cortex (Ungerleider & Mishkin, 1982, also see Figure 1.2). Functionally, the dorsal route has been suggested to be engaged in processing object location and movement. It is therefore also referred to as the “where-pathway”. Later it has been suggested to be important for visually guided action and was redubbed the action-pathway or “how-pathway” by Milner and Goodale (1995). It is generally considered to process information outside consciousness (Fang & He, 2005). The ventral route is often referred to as the “what-pathway” and is thought to be responsible for shape perception, object categorization, object detection and consciousness (Bar, et al., 2001; e.g. Grill-Spector, 2003; Gross, Rocha-Miranda, & Bender, 1972; but see Konen & Kastner, 2008).

![Diagram of visual information processing](image)

**Figure 1.2** Coarse schematic of visual information processing in the brain. Light enters the eye and is converted to nerve impulses. From there on it is relayed to visual area V1 and onwards to other visual areas. Routing follows two major pathways: the dorsal stream and the ventral stream. The dorsal stream is engaged more in processing positional information whereas the ventral stream is engaged more in processing object identity.
The dorsal – ventral distinction very much inspired the idea that visual information is processed in functionally different ways, and that these functions are supported by different pathways. This modular way of thinking about the visual brain is also present on a finer scale. Within the two pathways, information passes through a hierarchy of visual areas which have been identified with a combination of microelectrode mapping, tracer injections, histological stains and functional mapping (Felleman & Van Essen, 1991; Sereno, et al., 1995). It is thought that each area performs a cortical algorithm specific to that area and that it builds on computations performed by the previous areas, increasing complexity downstream. This idea is supported by a wealth of neurophysiological studies performed over the years. Neurons in early visual areas such as V1 have small receptive fields that respond to small line elements (Hubel & Wiesel, 1962, 1968). As one moves up in the hierarchy, the size of the receptive fields increases and the tuning properties of these neurons become progressively more complex. Eventually, neurons responding to complex shapes and specific object categories can be found in inferotemporal cortex (Epstein & Kanwisher, 1998; Gross, et al., 1972; Kanwisher, McDermott, & Chun, 1997). It is only a small leap from there to infer that objects are recognized by building up complexity, akin to a cortical assembly line (see Figure 1.3).

![Figure 1.3 Tuning characteristics of ventral stream neurons. Early visual areas respond to small parts of the visual fields to simple stimulus characteristics such as oriented bars. Areas higher up in the cortex have larger receptive fields and respond to stimulus conjunctions. The highest areas have even larger receptive fields and respond to complex shapes and complete objects.](image-url)
Quite a bit of evidence supports the idea that the visual system is able to recognize objects in such a fast incremental feedforward manner (Serre, Oliva, & Poggio, 2007; Thorpe, Fize, & Marlot, 1996; VanRullen & Koch, 2003). Moreover, in a study by Grill-Spector and Kanwisher (2005) it is shown that the presentation time required to detect an object is right on par with the presentation time required to categorize it. They conclude that as soon as you are able to detect an object, you know what it is, suggesting that neural mechanisms of conscious detection and object categorization are closely related. However, there are also studies that show that mere activation of object selective neurons in the inferotemporal cortex is insufficient to generate visual awareness (Kouider, Eger, Dolan, & Henson, 2008; G. Kovacs, Vogels, & Orban, 1995; Marois, Yi, & Chun, 2004). Those findings question the notion that conscious perception and object detection / categorization are related. If object categorization may proceed outside awareness, the question is what neural mechanism differentiates conscious from unconscious object categorization, and what the function of consciousness is if it is not to recognize objects.

Recurrent processing

A starting point for finding a neural mechanism that differentiates conscious from unconscious object categorization may be found by looking at the spatiotemporal dynamics of cortical processing. A meta-analysis of response latencies in the macaque brain during visual information processing has shown that an initial feedforward volley of neuronal activity proceeds hierarchically through all areas of the brain, activating its highest areas within about 120 ms (Lamme & Roelfsema, 2000). This initial sweep of information processing is often referred to as the fast feedforward sweep (FFS). The majority of the feedforward connections enabling this FFS have been shown to be reciprocal, with matching connections projecting downwards from higher to lower visual areas (Felleman & Van Essen, 1991, also see Figure 1.4). These back projections serve to reactivate earlier areas, a process known as feedback, re-entrant or recurrent processing (RP) following the initial FFS. Many theories regarding the role of RP in visual information processing have been proposed, some of the most notable of which will be briefly discussed below.
Figure 1.4 Visual processing pathways do not only contain feedforward connections to higher visual areas, but also include massive back projections from high to low visual areas, enabling recurrent processing (RP).

**Figure-ground segregation, object based attention and perceptual grouping**

One of the first major findings concerning the function of RP was its role in figure-ground segregation. Neurons in visual area V1 have traditionally been found to be sensitive to oriented line elements within only a fraction of the visual field, now known as the classical receptive field (Hubel & Wiesel, 1962; 1968, also see Figure 1.3). However, a series of studies conducted in the 1990’s have shown that these neurons start responding differently depending on the context in which they are stimulated, even if this context is located far beyond their receptive field horizon. For example, after about 100 ms., a V1 neuron responds more strongly to a texture defined figure than to a homogenous texture, even if the stimulated region within the classical receptive field of that neuron is identical in both cases (Lamme, 1995; Zipser, Lamme, & Schiller, 1996, also see Figure 1.5). It seems that this differential signal, also known as contextual modulation, plays a role in figure-ground segregation, as it selectively comes into play when a figure region perceptually pops out (Lamme, Zipser, & Spekreijse, 1998; Lamme, Zipser, & Spekreijse, 2002; Super, Spekreijse, & Lamme, 2001). Contextual modulation seems to originate from visual areas with larger receptive fields higher up in the
Introduction

cortex that are able to encapsulate the differential figure-ground context (Lamme, Super, & Spekreijse, 1998). Information regarding this context is used to effect figure-ground segregation in lower visual areas through RP. For a biologically inspired neural network model illustrating how this could work see Roelfsema et al. (2002).

Figure 1.5 After about 100 ms. a V1 neuron starts firing more strongly when it is located on a figure region than when located on a background region, even though its classical receptive field is too small to encapsulate this context. This signal is also referred to as contextual modulation. It is likely that contextual modulation is caused by recurrent interactions with high visual areas that have receptive fields that are large enough to encapsulate this context.

RP has also been suggested to play a role in object based attention (Roelfsema, Lamme, & Spekreijse, 1998). In a mental curve tracing task, it was shown that V1 neurons with receptive fields corresponding to regions of a traced curve show increased firing rates compared to neurons with receptive fields covering distractor curves that are not being traced. This is explained by RP enabling the spread of rate enhancements in V1 neurons corresponding to traced objects. The idea is that these rate enhancements serve as a label that incrementally group image elements belonging to the attended object into a coherent representation. This and earlier work on the role of RP in figure-ground segregation, led to the formulation of the incremental grouping theory (Roelfsema, 2006; Roelfsema, Lamme, &
Spekreijse, 2000). According to this theory, objects and shapes are initially quickly classified in elementary base groupings in the FFS. Subsequently, elements belonging to the same object are bound in perception in a more time consuming incremental grouping phase that depends on RP, akin to the proposed role of RP in figure-ground segregation. This incremental grouping phase allows the grouping of image elements that cannot be represented in base groupings, but require such an analysis due to task demands. In doing so, it enables the binding of different features of the same object together, thus solving what is often referred to as the ‘binding problem’. Although a neurophysiological theory at heart, it does seem to claim a role for attentional processes during incremental grouping. This is where it probably diverges most strongly from interpretations of RP as a correlate of figure-ground segregation only, such as studies showing that figure-ground signals reflecting RP are also present during states of complete inattention (e.g. Scholte, Witteveen, Spekreijse, & Lamme, 2006). More recently, it was proposed that RP serves distinct roles at different times during visual processing, implementing figure-ground segregation early in time and serving as an object based attentional mechanism later on (Roelfsema, Tolboom, & Khayat, 2007).

**Predictive coding and perceptual hypothesis testing**

A different strand of theories assign a role for RP in either predicting stimulus input on the basis of higher level interpretations or testing higher level interpretations against stimulus input. In a paper by Rao and Ballard (1999), a computational model is put forward in which feedback connections from higher to lower visual areas carry predictions of lower-level neural activities, while feedforward connections carry the residual errors between the predictions and the lower-level input. When inputting natural images into the model network, neurons responsible for carrying residual errors developed extra-classical receptive field properties resembling those of the early visual system, including an effect similar to that found by Lamme and others (Lamme, 1995; Zipser, et al., 1996). The larger idea behind their proposal seems to be that calculating predictions and feeding prediction errors back into the system may be a particular efficient coding strategy for dealing with change in the environment. If the
system learns the statistical regularities of the natural world while signaling deviations from such regularities to higher order visual areas, redundancy is removed by removing the predictable and focusing on prediction errors (the change). They suggest that predictive coding may be a general function expressed in many other areas of the brain, such as dopaminergic neurons carrying reward-prediction errors when projecting to the cortex and striatum (Schultz, 2001).

A somewhat related model but with different functional characteristics was proposed by Di Lollo and Enns (Di Lollo, Enns, & Rensink, 2000) to explain the effectiveness of masking. In their model, high level interpretations of an incoming input pattern are fed back to a lower level working space with the same or highly similar characteristics as the input layer (see Figure 1.6).

![Figure 1.6 The reentrant processing hypothesis by Di Lollo, Enns and Rensink (2000). An incoming stimulus at the input layer (I) is fed-forward to a higher level pattern recognition area (P). The interpretation at this level is subsequently fed back to a lower working space area (W) that has a similar neuronal representation as the input layer. Only if the comparison process at I and W constitutes a match, the representation reaches consciousness. Image reproduced from Di Lollo et al. (2000) with permission.]

Here, the working space prediction is compared to the input layer. If the two match, the system is locked into this interpretation, by which the stimulus becomes conscious. If the comparison process fails, the interpretation does not reach consciousness. When this happens, the higher level interpretation is eventually replaced by the new representation originating from the lower level input, making way for a new high level interpretation. The comparison process is suggested to serve a role in resolving ambiguity (i.e. when more than one higher
level interpretation becomes activated or when the lower level input is incomplete). Second, it is suggested that the comparison serves to retain spatial precision as the high level interpretation is represented in visual areas with large receptive fields that otherwise have little sensitivity to location, whereas the lower level input layer has small receptive fields that do have this sensitivity. The model was put forward by Di Lollo and Enns to explain the masking phenomenon. In masking, a stimulus is shortly followed by a second stimulus (the mask), rendering the first stimulus invisible. In their model, this happens because the mask replaces the stimulus representation at the input level even before a match can be made between the working space and the input, thus preventing the stimulus from ever reaching consciousness.

The reverse hierarchy theory of visual perception
Reverse hierarchy theory (RHT) by Hochstein and Ahissar (2002) asserts that conscious vision advances in reverse hierarchical direction. In this theory complex representations such as required for object categorization are built up in the FFS (see Figure 1.3), but the buildup process itself is implicit, and not available to consciousness. When reaching the highest visual areas, representations become conscious in “vision at a glance”, but at this point no details are explicitly represented in the representation. RP is required to enable “vision with scrutiny” in which object details are made available to consciousness by reactivating low level neurons that carry this level of description (see Figure 1.7).

RHT aims to explain a number of properties of the visual system, such as its ability to extremely rapidly detect the gist of a scene through vision at a glance and its inability to detect major changes in such a scene as observed in change blindness. The idea is that details are implicitly detected by low level areas and used for the buildup of the gist of a scene in the FFS, but these details themselves are not represented in the high-level receptive fields in which this gist is represented. As details are not explicitly bound to the high-level representation during vision at a glance, change blindness occurs. Binding at this point is not explicit, but done on the basis of prior knowledge, and may in some unusual cases even lead to false
conjunctions. Subsequent vision with scrutiny through RP unbinds illusory conjunctions and revises vision at a glance to enable the explicit detection of image details.

**Figure 1.7** Reverse hierarchy theory (RHT) by Hochstein and Abissar (2002). The feedforward hierarchy quickly builds up a representation of a scene, the details of which are not available to consciousness. Only by explicitly going back through the hierarchy are details explicitly bound to the higher level representation, enabling detailed conscious vision. Image reproduced in edited form from Abissar and Hochstein (2004) with permission.

RHT also claims to account for the phenomenon of pop-out, in which odd elements pop-out when they differ substantially from the distractors in one feature, but require serial search when they differ in a conjunction of features or when the difference is small (Treisman & Gelade, 1980). Pop-out and feature search were initially thought to occur in early visual areas. However, this association has been questioned in a number of ways: low-level neurons have precise position information, while pop-out does not (Atkinson & Braddick, 1989); they discriminate fine orientation and color differences that do not pop-out, and their receptive fields are too small for lateral inhibition among distractors to be the source of pop-out. Therefore, RHT asserts that pop-out is vision at a glance, with its neural substrate in high visual areas, whereas the neural correlate of visual search is vision with scrutiny, iteratively going back through the reverse hierarchy of Figure 1.7 to identify the odd element. In this
framework, the reason that some features pop-out whereas others do not is that these features have properties that are relevant for high-level categorization in the FFS.

**RP and visual consciousness**

Although RP seems to have different functional roles in these theories, they do not have to be mutually exclusive. Different functions may be served by the same physiological process. The theories also show important overlap. For example, both incremental grouping theory by Roelfsema and reverse hierarchy theory by Hochstein and Ahissar, point out the possibility that RP serves as mechanism by which to achieve binding in the visual system. Both Reverse hierarchy theory and Di Lollo and Enns’ model suggest that RP may serve to resolve perceptual ambiguity at high cortical levels, and that RP may serve to recapture spatial detail. A unifying description of the role of RP on the basis of these theories could be that RP is important for **perceptual** or **phenomenal** aspects of vision, or what some may call consciousness. For example, on the one hand RP is suggested to play a role in binding features together while on the other hand the unity or oneness of perceptual experience is often referred to as an important feature of consciousness (e.g. LaRock, 2007). Another example is that RP works to resolve perceptual ambiguity, while perceptual switching has frequently been presented as an expression of visual consciousness (e.g. Tong, Nakayama, Vaughan, & Kanwisher, 1998). In addition, these theories seem to largely agree on the idea that the FFS is an automatic process the contents of which are not available to consciousness, whereas visibility of stimuli or stimulus features is uniquely identified with RP.

The idea that RP is important for consciousness is supported by a large number of studies showing the co-occurrence of consciousness and RP. On the one hand, RP disappears when abolishing consciousness through classical methods, such as during general anesthesia (Lamme, Zipser, et al., 1998) and visual masking (Fahrenfort, Scholte, & Lamme, 2007; Lamme, et al., 2002 as well as chapter 2 of this thesis). On the other hand, consciousness disappears when disrupting RP experimentally as can be achieved with TMS (Pascual-Leone & Walsh, 2001) or extra-striate lesioning (Lamme, Super, et al., 1998). Finally, when recording from V1 while monkeys
are freely viewing figure-ground stimuli, ‘Seen’ responses are associated with RP, while ‘Unseen’ responses coincide with absence of RP (Super, et al., 2001). These and other arguments (Lamme, 2003, 2004) regarding the role of RP in conscious vision has even led to the suggestion to redefine consciousness in terms of RP altogether (Lamme, 2006).

**Overview of this thesis**

Summarizing, the architecture of the visual system seems to allow for objects to be detected and categorized in the FFS, while consciousness and figure-ground segregation are supported by RP only. If true, this would mean that (1) object detection and categorization happen quickly, automatically and unconsciously in the FFS while (2) conscious object perception and figure-ground segregation require RP and happen only after objects have been unconsciously detected and categorized. This would shed new light on the function of consciousness and the order in which important functions like figure-ground segregation and object detection/categorization are established in the brain. This thesis focuses on testing these hypotheses by looking at the neural mechanisms (FFS/RP) subserving object detection, object categorization, visual consciousness and figure-ground segregation.

In chapter 2 of this thesis it is shown using a masking paradigm that objects are detected in the FFS, even when subjects themselves are unable to detect those objects above chance levels. This shows that object detection by the brain proceeds quickly and automatically, and that the FFS enabling detection proceeds outside consciousness. It is also shown that conscious detection is associated with reactivation of early visual areas through RP. This reactivation is absent when subjects cannot see the object.

Chapter 3 builds on these findings by showing that object detection in the FFS does not correlate with one’s ability to consciously detect the object. Put differently: the brain always detects the object in the FFS in the same way, regardless of how good or bad one is at consciously detecting the object. The first signals that correlate with one’s capacity to consciously detect the object emerges
later in time as a consequence of recurrent interactions within visual cortex.

These two chapters give a good impression of the temporal course of information processing during object detection and visual perception. However, the use of EEG does not allow one to determine which visual areas are involved in these processes. In chapter 4 it is shown, with the help of fMRI that the largest difference between consciously and unconsciously detected objects can be found in V1. This shows that reactivation as a result of recurrent interactions as found in chapter 2 and 3, reaches back all the way to the earliest visual area V1, and that this reactivation can be linked to consciously detecting an object. The BOLD response in dorsal area V7 on the other hand proves to be identical for consciously and unconsciously detected objects, in line with an FFS that automatically and indiscriminately detects objects.

These first three chapters establish a clear description of the neural substrates of conscious and unconscious object detection in the brain. Objects are detected in the FFS, at least up to area V7, and this process unrolls regardless of whether the object reaches consciousness. RP selectively emerges when somebody consciously detects the object, and the degree to which this happens correlates with the strength of the recurrent signal, which reaches back to V1. But what about object categorization? Is the brain able to categorize an object as a face or a house without its owner becoming conscious of this categorization?

In chapter 5 is shown that even object categorization can proceed outside of consciousness, plausibly in the FFS. Cortical areas that respond to faces and houses are selectively activated by these stimuli, even if they cannot be consciously seen. Moreover, the difference between consciously and unconsciously categorized objects is expressed in recurrent interactions within visual cortex rather than by object selective activity. Objects that are consciously seen show a clear signature of RP / figure-ground segregation in V1, V2 and LOC, whereas objects that are not consciously seen show no RP related signals whatsoever. Hence, objects are categorized by the brain even when subjects are unconscious of them, and only consciously seen objects are characterized by RP.
Together, the studies in this thesis show that the brain initially detects and categorizes objects as in a reflex: quickly, automatically and unconsciously. The recurrent processes that follow, mediate figure-ground segregation and visual consciousness. Although counterintuitive, this places visual consciousness in closer proximity to processes of figure-ground segregation and perceptual grouping than to the processes that achieve object detection and categorization.