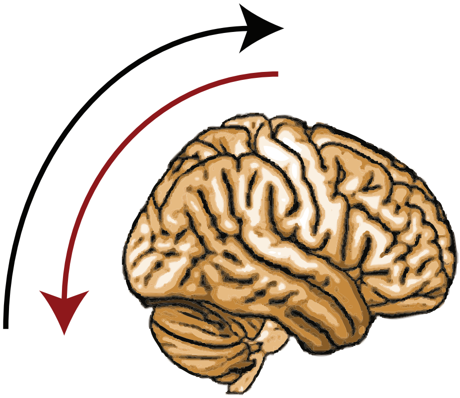
Conscious and Unconscious Vision



Johannes Fahrenfort

CONSCIOUS AND UNCONSCIOUS VISION

Printed by Ridderprint Offsetdrukkerij B.V., Ridderkerk
ISBN: 978-90-5335-211-3

Cover design by J.J.Fahrenfort. based on an FSL rendered brain.

Printing of this thesis was supported by:
BioSemi

CONCIOUS AND UNCONSCIOUS VISION

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de universiteit van Amsterdam
op gezag van de Rector Magnificus
prof. dr. D.C. van den Boom
ten overstaan van een door het college voor
promoties ingestelde commissie,
in het openbaar te verdedigen in de Agnietenkapel
op dinsdag 3 november 2009, te 12:00 uur
door

Johannes Jacobus Fahrenfort

geboren te Amsterdam

Promotiecommissie

Promotor : prof. dr. V.A.F. Lamme
Co-promotor : dr. H.S. Scholte

Overige leden: prof. dr. C.M.A. Pennartz
prof. dr. K.R. Ridderinkhof
prof. dr. P. Roelfsema
prof. dr. F. Verstraten
prof. dr. B.L.M.F. De Gelder

Faculteit der Maatschappij- en Gedragswetenschappen

Ik draag dit proefschrift op aan mijn
lieve vader Joop die er helaas niet meer is.

Table of contents

Chapter 1.

Introduction

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.

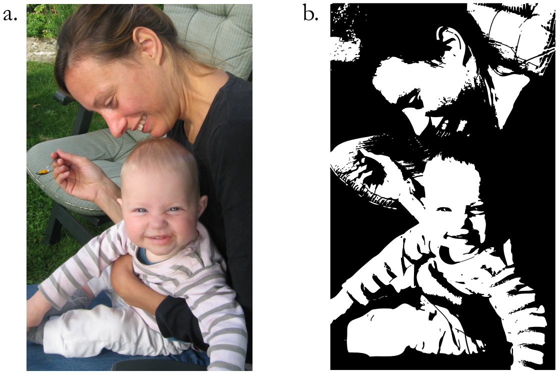


Figure 1.1 *Illustration of image complexity that the visual system needs to disentangle. a) The human subjects in this image show complex figure-background relations that are notoriously difficult to disentangle by computer algorithms. b) Even under completely altered conditions of light and color the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

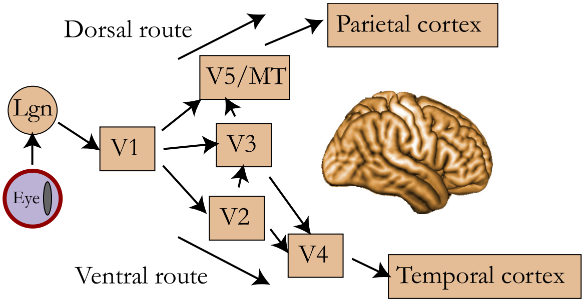


Figure 1.2 *Coarse schematic of visual information processing in the brain. Light enters the eye and is converted to nerve impulses. From there on it is relayed to visual area V1 and onwards to other visual areas. Routing follows two major pathways: the dorsal stream and the ventral stream. The dorsal 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).

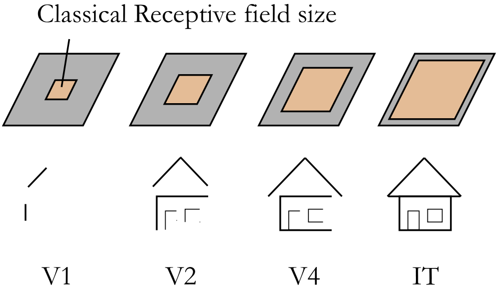


Figure 1.3 *Tuning characteristics of ventral stream neurons. Early visual areas respond to small parts of the visual fields to simple stimulus characteristics such as oriented bars. Areas higher up in the cortex have larger receptive fields and respond to stimulus conjunctions. The highest areas have even larger receptive fields and respond to complex shapes and complete objects.*

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.

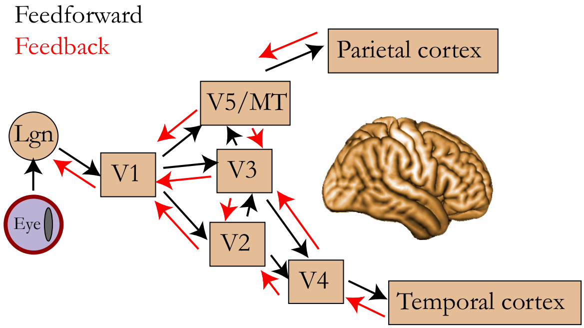


Figure 1.4 *Visual processing pathways do not only contain feedforward connections to higher visual areas, but also include massive back projections from high to low visual areas, enabling recurrent processing (RP).*

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 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).

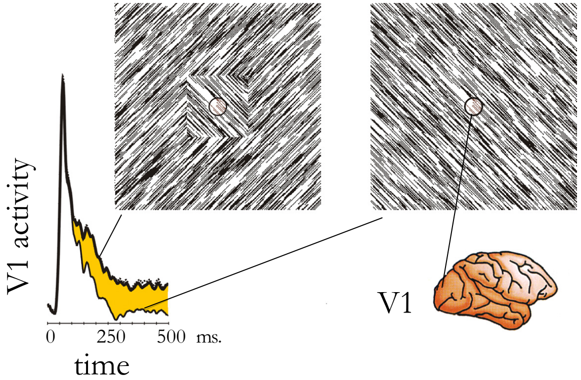


Figure 1.5 *After about 100 ms. a V1 neuron starts firing more strongly when it is located on a figure region than when located on a background region, even though its classical receptive field is too small to encapsulate this context. This signal is also referred to as contextual modulation. It is likely 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).

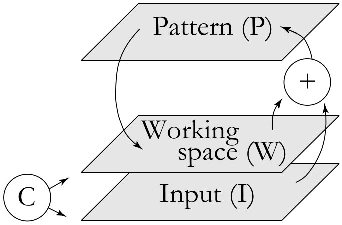


Figure 1.6 *The reentrant processing hypothesis by Di Lollo, Enns and Rensink* (2000). *An incoming stimulus at the input layer (I) is fed-forward to a higher level pattern recognition area (P). The interpretation at this level is subsequently fed back to a lower working space area (W) that has a similar 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.

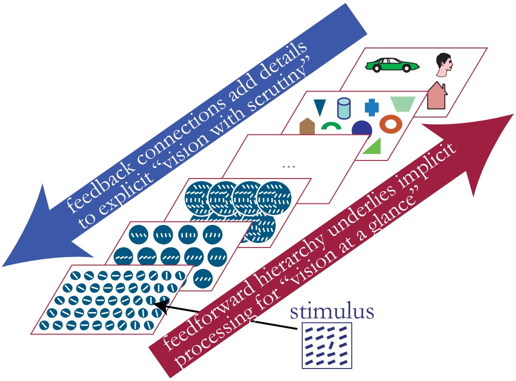


Figure 1.7 *Reverse hierarchy theory (RHT) by Hochstein and Ahissar (2002). The feedforward hierarchy quickly builds up a representation of a scene, the details of which are not available to consciousness. Only by explicitly going back through the hierarchy are details explicitly bound to the higher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

Chapter 2.

Masking disrupts recurrent processing in human visual cortex

Abstract

In masking, a stimulus is rendered invisible through the presentation of a second stimulus shortly after the first. Over the years, authors have typically explained masking by postulating some early disruption process. In these feed forward type explanations, the mask somehow “catches up” with the target stimulus, disrupting its processing either through lateral or interchannel inhibition. However, studies from recent years indicate that visual perception - and most notably visual awareness itself – may depend strongly on cortico-cortical feedback connections from higher to lower visual areas. This has led some researchers to propose that masking derives its effectiveness from selectively interrupting these recurrent processes. In this experiment, we used EEG measurements to determine what happens in human visual cortex during detection of a texture defined square under non-masked (seen) and masked (unseen) conditions. EEG derivatives that are typically associated with recurrent processing turn out to be absent in the masked condition. Moreover, extrastriate visual areas are still activated early on by both seen and unseen stimuli, as shown by scalp surface Laplacian current source density maps. This conclusively shows that feedforward processing is preserved, even when subject performance is at chance as determined by objective measures. From these results we conclude that masking derives its effectiveness, at least partly, from disrupting recurrent processing, thereby interfering with the neural mechanisms of figure-ground segregation and visual awareness itself.

Fahrenfort, J. J., Scholte, H. S., & Lamme, V. A. F. (2007). Masking disrupts reentrant processing in human visual cortex. *Journal of Cognitive Neuroscience, 19*(9), 1488-1497.

Introduction

In backward masking, a target stimulus is rendered less perceptible or even invisible through the presentation of a second stimulus, the mask. By now, a respectable number of studies have investigated the neural basis of visual masking, both in its own right, and as a corollary to studying visual perception and awareness. In explaining a variety of masking phenomena, some have emphasized low level lateral inhibitory mechanisms (Francis, 1997; Macknik & Livingstone, 1998) , while others have postulated target disruption at the pre-categorical level through interchannel inhibition (Breitmeyer, Ro, & Ogmen, 2004; Ogmen, Breitmeyer, & Melvin, 2003).

However, early studies as well as more recent imaging studies on masked priming show that processing of masked targets may continue well beyond the pre-categorical level (Dehaene, et al., 2001; Dehaene, et al., 1998; Eimer & Schlaghecken, 1998; Greenwald, Klinger, & Liu, 1989; Naccache, et al., 2005). These results suggest that masking does not disrupt target processing at an early stage. This is in line with various neurophysiological reports in which early transient responses to undetected masked stimuli were measured in high visual areas such as IT, and even in visual neurons of the frontal eye field (G. Kovacs, et al., 1995; Rolls & Tovee, 1994; Thompson & Schall, 1999).

This has resulted in a number of researchers proposing that masking disrupts reentrant cortical signals, while largely leaving signals related to feedforward processing intact (Bridgeman, 1980; Di Lollo, et al., 2000; Lamme, et al., 2002; Ro, Breitmeyer, Burton, Singhal, & Lane, 2003). These proposals build on the notion that besides feedforward connections from low to higher visual areas, roughly equal numbers of recurrent or reentrant connections exist going back from higher to lower visual areas (Felleman & Van Essen, 1991). Such feedback connections have been suggested to play an integral role in a range of processes, such as figure-ground segregation, attention, awareness, predictive coding, as well as perception of visual detail (Hochstein & Ahissar, 2002; Kastner & Ungerleider, 2000; Lamme, 1995; Lamme & Roelfsema, 2000; Luck, Chelazzi, Hillyard, & Desimone, 1997; Rao & Ballard, 1999; Spratling & Johnson, 2004). It is not unlikely that masking derives its effectiveness (at least in part) from disrupting recurrent signals, while leaving feedforward signals intact, thereby interfering with mechanisms thought to mediate visual awareness itself (Lamme & Roelfsema, 2000).

However, many masking studies do not show a relation between recurrent processing and masking. This may be caused by the fact that many such studies make use of luminance or color contrast targets and metacontrast masks. In such metacontrast masking paradigms, the mask tightly fits the preceding target and the common contour has a contrast of opposite polarity (Becker & Anstis, 2004). It is quite likely that low level lateral inhibitory mechanisms come into play when processing these types of stimuli, at least at sufficiently short SOA’s (Macknik & Livingstone, 1998; Macknik & Martinez-Conde, 2004).

Contrast sensitivity is fundamental to vision and contrast is greatly enhanced by lateral inhibition at many levels of visual processing, starting in the retina (Hartline, 1949). Therefore, the use of (meta)contrast stimuli without appropriately controlling for the influence of contrast on the signal exaggerates the influence of low level lateral inhibition, while obscuring the influence of other higher level processes, such as recurrent processing. The stimuli used in this experiment - in combination with a subtraction method (detailed below) – eliminate the influence of luminance contrast on the signal, while at the same time controlling for other low level influences.

Aside from the contrast issue, visibility of masked targets is often assessed using an identification task. However, a detection task is more fundamentally connected to the notion of visual awareness, as many studies demonstrate that residual identification in the absence of awareness exists. This is evident not only from research into the phenomenon of blindsight (Weiskrantz, Barbur, & Sahraie, 1995; Weiskrantz, Warrington, Sanders, & Marshall, 1974), but selective behavior in the absence of awareness has also been shown in normal subjects (Boyer, Harrison, & Ro, 2005; VanRullen & Koch, 2003) as well as in studies using masked priming (e.g. Eimer & Schlaghecken, 1998; Vorberg, Mattler, Heinecke, Schmidt, & Schwarzbach, 2003). Selectivity thus seems an inappropriate indicator of awareness. Therefore, the present experiment employs an objective detection task to assess visibility.

We investigated the effect of pattern masking on the detection of an orientation defined square in a texture (Figure 2.1), while recording EEG in human subjects. Figure and No Figure trials were either strongly masked by a pattern mask leading to chance detection performance (Masked condition), or weakly masked by a uniform grey screen leading to near perfect detection performance (Unmasked condition). We used a subtraction procedure to control for various confounding influences on the EEG.

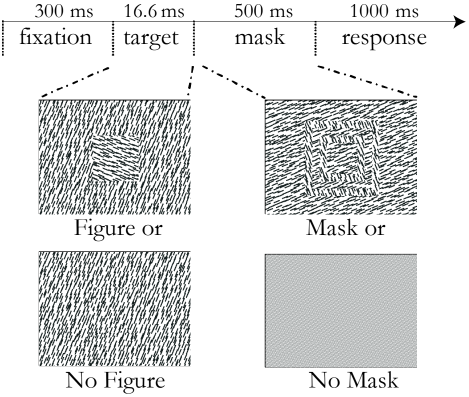


Figure 2.1 *Schematic description of a trial. Subjects were required to indicate whether a figure was present in a texture or not. Half of the trials contained a figure. Half of the trials only contained a homogenous texture. A trial could either be strongly masked (masked) or weakly masked (unmasked). If subjects could not see anything due to masking they were instructed to guess (forced choice detection).*

By subtracting the Figure trials from the No Figure trials (see Figure 2.2), two goals were achieved: (1) because both Figure and No Figure trials were followed by a mask in the masked condition, any direct influence of the mask on the EEG was subtracted out. The same holds in the unmasked condition. (2) Because both Figure and No Figure trials were created using the same set of oriented line elements, Figure and No Figure trials contained exactly equal amounts of contrast and orientation. Thus, any direct influence of low level processes such as contrast detection and orientation tuning is subtracted out (Zipser, et al., 1996). This subtraction procedure would not work for metacontrast target-mask combinations as used in other studies, as the net result of the subtraction of target-present and target-absent trials would inherently leave a net result of contrast.

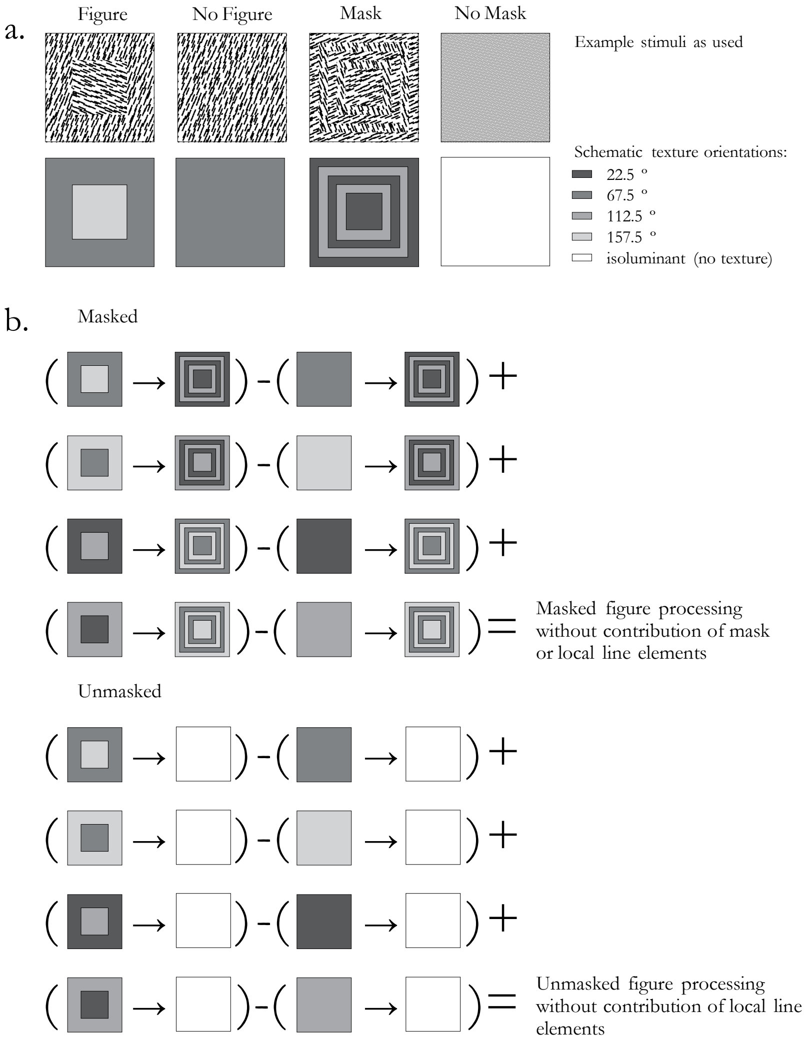


Figure 2.2 *(a) Examples of stimuli and the schematic orientations scheme. Each grey value represents an orientation, white is isoluminant grey. (b) Representation of subtractions leading to correlates of global processing. Influences of local line elements as well as direct mask contributions are canceled out by ensuring each orientation is present equally often at each side of the minus sign. Color scheme same as in (a).*

The spatiotemporal dynamics of the target-present and target-absent subtraction signal was studied at an extremely short SOA (16.6 ms) where targets were either easily perceived or remained totally undetected as determined by objective measures. We observed marked activation in extrastriate regions in both the visible and invisible condition as a result of feedforward processing. Later and more posterior activation resulting from recurrent processing was observed only in the visible condition. From the results we infer that recurrent processing is abolished as detection performance drops to chance, even though feedforward processing is entirely preserved.

Methods

Participants

Twenty one Psychology students took part in the experiment in partial fulfillment of first year course requirements. All subjects (mean age 21.4, ±1.9) had normal or corrected-to-normal vision. 71% was right handed. Each subject provided written informed consent before the start of the experiment. All procedures were approved by the ethical committee of the psychology department of the University of Amsterdam. Two subjects scored more than two standard deviations below the mean in the unmasked task, indicating difficulty performing the task properly. These were excluded from further analyses. Three subjects scored significantly above chance in the masked condition as ascertained by binomial tests ($p<.05$), indicating an ability (however slight) to detect the masked figure at this masking interval. These were excluded because it cannot be ruled out that they were still consciously processing these stimuli. In total five subjects were excluded.

Stimuli and Task

Without prior training, subjects had to detect the presence or absence of an orientation defined square in a texture pattern, which was either followed by a mask or by an isoluminant grey screen (all stimuli had a luminance of 72.7 cd/m^2). Each of the 4 resulting trial types was presented an equal number of 300 times in random order throughout the experiment. A trial started with 300 ms fixation, followed by a centrally located target stimulus for 16.6 ms, immediately followed by either a mask or an isoluminant grey screen for 500 ms (Figure 2.1). Target stimuli and mask consisted of a texture of oriented line elements (approx. 0.07° thick and 0.4° long in visual angle). All line elements in a target either had the same orientation (homogenous trials), or contained a centrally located square with line elements orthogonal to the background (Figure trials). The square subtended 2.73° of visual angle. In the masked condition, targets were immediately followed by a texture defined pattern mask (4.54° visual angle) using orientations not used in the preceding target. In the unmasked condition, targets were followed by an isoluminant grey screen (no texture). A total of four orientations of line segments were used for creating figure, homogenous and mask stimuli (22.5°, 67.5°, 112.5° and 157.5°, also see Figure 2.2), so that different orientations for target and mask could always be used within any one trial. All orientations were completely counterbalanced over trials so figure and homogenous trials did not differ from each other with respect to local stimulation (for a similar procedure see Caputo & Casco, 1999; Lamme, Van Dijk, & Spekreijse, 1992; Scholte, et al., 2006). Stimuli were created using Matlab (The MathWorks, Inc., Natick, MA, USA). Stimuli were presented using Presentation (Neurobehavioral Systems, Inc., Albany, CA, USA).

Subjects were instructed to fixate throughout the experiment. The fixation dot would turn green 500 ms after target offset, to indicate a response was required, after which the subject had a maximum of 1000 ms to respond before the experiment would continue. On half of the trials subjects were required to press the left button to indicate square detection and the right button for a homogenous trial. In the other half of the trials subjects had to switch so that a square was associated with right and homogenous with left button presses. This response scheme was counterbalanced across subjects, half of the subjects starting out square detection with their right hand, and half of the subjects starting out using their left hand. Switching hands posed no difficulty as the high percentages correct on unmasked trials shows. Subjects were encouraged to guess if they were unable to see if a figure was presented or not, effectively turning the task into a forced choice task.

Measurements and Analyses

ERP's were recorded from the scalp using a BioSemi ActiveTwo 48 channel active EEG system (BioSemi, Amsterdam, the Netherlands) at 256 Hz. 48 scalp electrodes were measured (referenced to Fz), as well as two electrodes for horizontal and two for vertical eye movements (each referenced to its opposite counterpart). The data was filtered using a high pass filter of 1 Hz and a low pass filter of 20 Hz, as well as a 50 Hz notch filter. Before ocular correction, automatic artefact rejection was applied by removing segments containing voltage steps of more than 50 µV, removing any segments falling outside the -200 µV to 200 µV range as well as removing those segments containing larger than 300 µV differences within the segment. Ocular correction was applied on the basis of the horizontal and vertical electro-oculograms (Gratton, Coles, & Donchin, 1983). After ocular correction, artefact rejection was applied again by removing all segments outside the -75 µV to 75 µV range. Linear local DC detrending was applied. This is a procedure to remove current drift by subtracting a linear function from each segment. Baseline correction was applied in the -300 ms to 0 ms interval. All pre-processing steps were done using Brain Vision Analyzer (Brain Products GmbH, Munich, Germany).

All analyses were done on difference waves. These were obtained by subtracting average No Figure trials from Figure trials, separately for the Masked and Unmasked condition. Spline Laplacian distribution maps were calculated by interpolating difference waves using spherical splines and approximating current source densities (Perrin, Pernier, Bertrand, & Echallier, 1989). The resulting maps are spatial second order derivatives of the scalp potentials lending greater weight to local contributions of cortical generators, filtering out deep sources, as well as being reference free (Nunez & Srinivasan, 2006).

In order to test at which time points the differences between Figure and No Figure trials for the Masked and Unmasked condition significantly deflect from chance, as well as test the difference between masked and unmasked trials, a random effects analysis was performed by employing a paired two-tailed t-test at each time point, treating the average of each subject at that time point as an observation. Correction for multiple comparisons with respect to the number of time points being tested was done by limiting the False Discovery Rate (FDR), a method in which the p-value at which significance is evaluated is corrected for the number of tests being performed (Benjamini & Hochberg, 1995). The method fixes the expected proportion of false positives (the FDR) in contrast to Bonferroni correction, which controls the chance of any false positive among all tests. The FDR is an often used method in a wide variety of scientific fields. For an explanation of how it is used in the field of neuroimaging see Genovese, Lazar & Nichols (2002). Statistical analyses and visualization of the timecourses were done using Matlab (The MathWorks, Inc., Natick, MA, USA).

Results

The difference between Figure and No Figure trials was easily detected when no mask was present, as expressed in an average detection rate of 96.5% (±2.5). For the masked trials the detection rate was at chance at 51.4% (±4.1). Detection rates are expressed as the average perfect observer score, which is an objective and subject bias free measure of a subjects' ability to detect a signal (Wickens, 2002). None of the included subjects scored significantly above chance in the masked condition (binomial tests, significance evaluated at .05).

Visual Evoked Potential (VEP) averages were calculated from the EEG data for each condition. Difference waves of these averages were computed by subtracting No Figure from Figure trials in order to isolate activity related to processing of global organization of the target stimulus and discard processing related to processing of local line elements (Figure 2.2, see also Caputo & Casco, 1999; Lamme, et al., 1992; Scholte, et al., 2006). This was done separately for the Masked and Unmasked condition, as to subtract out direct contributions of the mask to the VEP's and enable comparison of masked and unmasked trials. All VEP difference waves were converted to spline Laplacian's (see method section) to be able to better isolate local generators of any observed effects and filter out deep sources (Nunez & Srinivasan, 2006).

A split half procedure of odd and even trials was used to evaluate the data (Figure 2.3). *Even* trials were used for a visualization of instructive time segments using spline Laplacian maps. These were used to select relevant electrodes for pooling. *Odd* trials were used to evaluate significant deflections of the average current source density at these electrodes over time. Using *odd* trials for the spatial map and *even* trials for the time course values ensures complete independence of data used for inspection and data used for evaluation of significance. This gives an enormous boost to the reliability of the effects that were found, while precluding coincidental significance of sources that were identified visually. It is important to note that even though for brevity not all time points in the 0-305 ms range have been visualized using scalp distribution maps in Figure 2.3, electrode pooling was done on the basis of all visually identifiable sources in this domain, while none were opportunistically left out. Pooling the data to a limited number of electrodes in this way circumvents the multiple comparisons problem of having to evaluate large numbers of electrodes. Pooling relevant electrodes furthermore increases the signal to noise ratio.

Figure 2.3a shows the course of neural processing in the case of unmasked trials. Three stages can be distinguished:

1. A pre-110 ms stage with bilateral anterior occipitotemporal neural generators consistent with feedforward processing towards extrastriate visual areas and beyond (Foxe & Simpson, 2002).
2. A post-110 ms stage with a clear posterior occipital generator consistent with recurrent processing towards early visual areas, possibly V1.
3. A strong recurring bilateral generator in the 200-300 ms range, plausibly extrastriate and beyond.

All generators show significant deflections as ascertained by paired t-tests between Figure and No Figure trials. These were performed on each time point in the time segment, subsequently corrected for multiple comparisons by fixing the False Discovery Rate (FDR) at .05 (see method section).

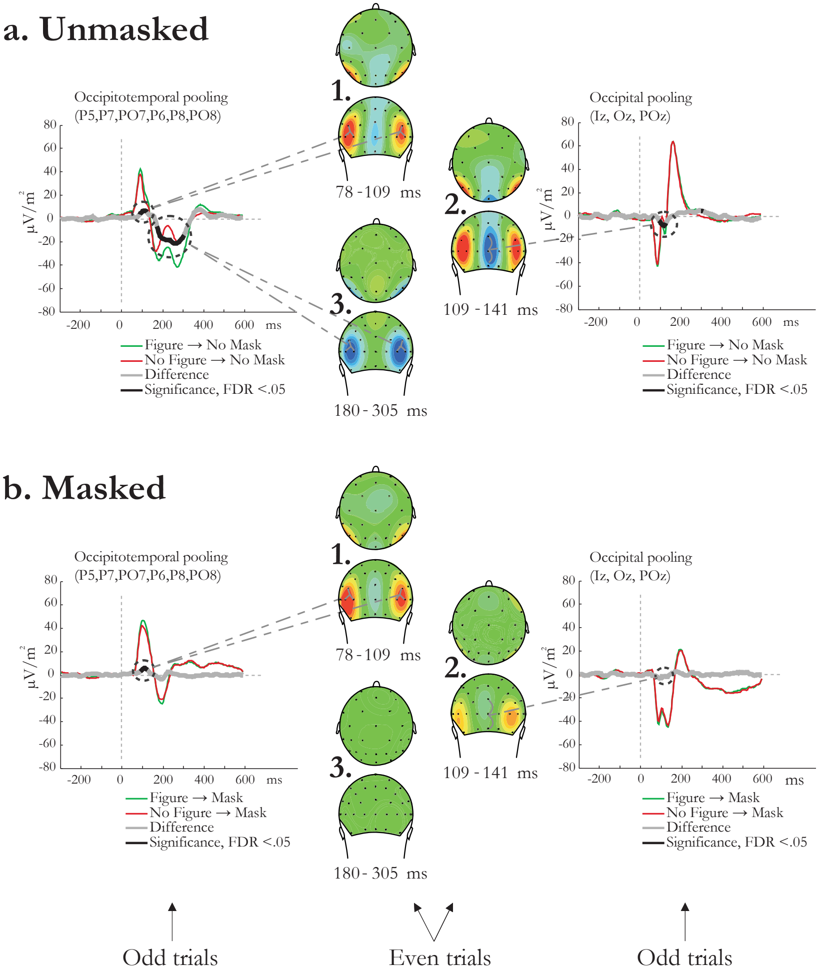


Figure 2.3 *Results. Spline Laplacian maps over three time segments: (1) 78-109 ms, (2) 109-141 ms and (3) 180-305 ms and event related averages of relevant electrodes (pooled occipitally and occipito-temporally). Even trials were used for data inspection (the maps) to determine sensible electrode poolings. Odd trials were used for the averages and evaluation of statistical significance. (a) Unmasked condition (b) Masked condition. Both masked and unmasked show strong significant temporal bilateral activation reflecting feedforward activation in stage 1. More posterior occipital activation reflecting recurrent processing can only be seen in the unmasked condition in stage 2, but is abolished by masking.*

Figure 2.3b shows the same three stages for the masked trials. Stage (1) again shows a significant bilateral extrastriate neural generator. Stage (2) and (3) are largely absent and show no significant deflections. A direct comparison of the masked and unmasked difference waves was also done (see Figure 2.4). In accordance with the other results, these showed significant differences in stage (2) and stage (3), but not in stage (1). Taken together, these results are consistent with an interpretation where masking results in an intact initial feedforward activation of extrastriate areas, followed by a disruption of recurrent processing, possibly affecting both striate and extrastriate visual cortex.

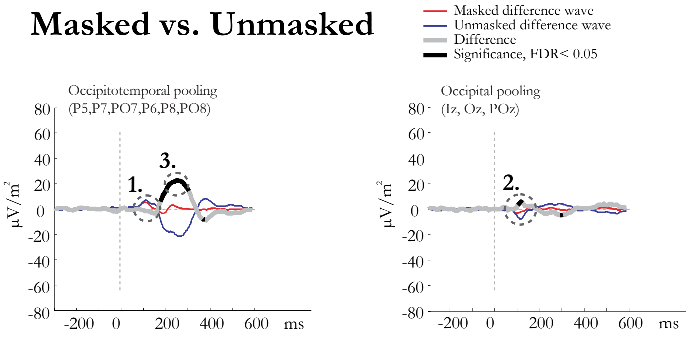


Figure 2.4 *A difference wave of the difference waves outlined in Figure 2.3 was calculated so as to directly test at which intervals the masked and the unmasked figures differ from each other. This figure clearly shows that masked and unmasked trials do not significantly differ from each other in the early 78-109 ms occipitotemporal stage related to feedforward processing (1), but do differ significantly from each other in the 109-141 ms occipital stage (2) which is plausibly related to recurrent processing as well as the later 180-305 ms occipitotemporal stage (3).*

Because the resolution of ERP's is rather limited, some may object to the idea that recurrent processing can be pinpointed using ERP's. Yet in our view, only a relatively simple claim needs to be substantiated: (1) there is early activation found anterior on the scalp (2) there is distinct later activation posterior on the scalp. In combination with current source densities, and provided that the generators are clearly distinct, this simple anterior/posterior claim seems sufficient to claim recurrent processing,

Discussion

The most salient aspect of these results is that even when subjects are unable to perceive anything by objective measures, widespread bilateral occipitotemporal activation still takes place, showing a clear ability of the brain to pick up differences that the subject cannot express. This early activation is almost indistinguishable from early activity in the visible situation and is insufficient to generate visual awareness. Some additional process appears necessary in order to achieve conscious perception and detection. The earliest difference between the visible and non-visible condition arises in the >110 ms timeframe, and occurs more posterior than the early bilateral activation. This is most consistent with an interpretation where masking disrupts recurrent processing between high and low visual areas, while leaving feedforward activation intact.

It is unclear how masking accounts resting solely on lateral inhibition could explain these data. The strongest evidence for the influence of lateral inhibition in masking comes from studies employing first order metacontrast stimuli (e.g. Macknik & Livingstone, 1998). First order (luminance defined) stimuli are obvious candidates for spatiotemporal lateral inhibitory influences, since one can see how a metacontrasting stimulus laterally inhibits a first order contrast stimulus of opposite polarity. However, the present study makes use of second order target stimuli and pattern masking. As target and mask do not bear a metacontrasting relation to each other, low level lateral inhibitory influences are less likely. Moreover, any remains of low level inhibitory influences that may still exist are subtracted out using the Figure minus No Figure subtraction.

However, one could postulate that high level lateral inhibitory mechanisms operate between complex representations – as suggested by Macknik and Martinez-Conde (2004). If this were the case, one would predict the pre-110 ms activation to be disrupted in the masked scenario as a result of these inhibitory influences. But the locus of interruption in this experiment is later in time and more posterior. This is not to say that lateral inhibitory influences do not play a role in many types of masking, only that the view of ubiquitous lateral inhibition as an explanation for our results is not tenable.

Interchannel inhibition accounts of masking presume the existence of two channels in visual processing, a fast and a slow one. Masking is presumably caused by the mask in a fast acting channel inhibited by the processing of the target in a slow acting channel. Such accounts cannot be ruled out entirely by these data, as such theories do not specify at which level inhibition takes place. It could be argued that the feedforward activity found in this study is caused by target activity in a fast acting channel, and that the posterior >110 ms activity which we call recurrent activity is caused by a slow acting channel. In the masked scenario this activity would be inhibited by the mask through subsequent activation of the fast acting channel (which would have to be projecting downwards to inhibit the slow acting channel). However, aside from the fact that the posterior activity we observed is not very "sustained" as such models propose, there is little neurophysiological evidence that interchannel inhibition actually exists (also see Enns & Di Lollo, 2000). This makes an explanation built on proven notions of recurrent processing much more parsimonious in this context.

Aside from the present study, a number of other studies have inferred the influence of masking on recurrent processing in human subjects, most notably studies employing object substitution masking (Di Lollo, et al., 2000; Weidner, Shah, & Fink, 2006; Woodman & Luck, 2003). In object substitution masking (also referred to as four-dot masking), a four dot mask is presented surrounding a non-attended target. The subsequent discrimination of this target is unimpaired when target and mask co-terminate, but strongly impaired when the dots remain visible after target offset. This is explained by postulating that the four dots (after offset of the target) create a mismatch between the cortical representation of the four dots in early visual areas and the cortical representation of the target in higher areas. Although we believe the results of these studies have significant implications for the influence of recurrent processing on perception, they differ from our results in important ways.

Four dot masking depends on attention being directed elsewhere and increases with increasing set size (Di Lollo, et al., 2000; Enns, 2004). In our pattern masking study, attention is fully directed on the target, and set size is just one. Moreover, uninterrupted processing of masked targets using object substitution continues well into the 200-300 ms domain (Woodman & Luck, 2003), whereas our results show that processing is disrupted at 110 ms. Indeed, a recent fMRI study comparing pattern and object substitution masking shows that differences between the two can be found in the left middle frontal gyrus, in the precentral gyrus of the right hemisphere and in the medial superior frontal gyrus (Weidner, et al., 2006), while our study shows strictly occipital disruption due to pattern masking.

What could account for these differences? What happens during object substitution masking seems to be more akin to what happens during other attentional manipulations, such as the attentional blink. Many attentional blink studies have shown that processing of unseen target stimuli goes almost completely uninterrupted up to 300 ms and that differences between reported and unreported targets are to be found in frontal and parietal cortices (Kranczioch, Debener, Schwarzbach, Goebel, & Engel, 2005; Luck, Vogel, & Shapiro, 1996; Sergent, Baillet, & Dehaene, 2005). This is reminiscent of the response profile found in object substitution masking. As other authors have noted, the failure of subjects to report stimuli in such instances may reflect a loss of information at a post-perceptual stage, where information is replaced before it can be transferred to working memory and accessible awareness (e.g. Luck, et al., 1996; Woodman & Luck, 2003).

It may still be recurrent processing that is interrupted, but at a later stage and between visual and higher brain areas (possibly frontal and parietal), while early recurrent processing within visual areas goes uninterrupted. Recent models of consciousness proposed by Dehaene, Changeux, Naccache, Sackur and Sergent (2006) and Lamme (2003, 2006) outline similar ideas in more detail. The notion of several reentrant or recurrent loops at multiple levels is highly likely given the temporal dynamics of cortical processing (e.g. Foxe & Simpson, 2002). Because of the +110 ms timing and location observed in this experiment, we conclude that recurrent processing was disrupted at perceptual levels early on, comparable to what has been shown in other studies using different stimuli (Breitmeyer, et al., 2004; Haynes, Driver, & Rees, 2005; Ro, et al., 2003).

The function of recurrent processing within occipital cortex is a somewhat open question, though strong evidence exists that it is of importance for figure-ground segmentation (e.g. Hupe, et al., 1998;

Lamme, 1995; Lamme, Rodriguez-Rodriguez, & Spekreijse, 1999). In such studies it is typically shown that a neuron in V1 starts responding differentially after about 100 ms when its classical receptive field (CRF) is located on the center of a figure compared to when it is on a background, despite local stimulation being exactly equal (cf. Figure 2.2). These extra-classical receptive field effects are usually referred to as contextual modulation, as they are modulated by context information *outside* their CRF. The presence and timing (>100 ms) of contextual modulation strongly suggests that recurrent neural processing takes place going back from higher areas containing larger receptive fields to lower areas such as V1. Also, when extrastriate regions are lesioned, contextual modulation disappears, while leaving receptive field tuning properties intact, conclusively showing the role of recurrent processing in contextual modulation (Lamme, Super, et al., 1998).

A study in which the effect of target-mask SOA on contextual modulation in macaque V1 was investigated showed a strong relation between SOA, contextual modulation, and the monkey's ability to detect a texture defined figure (Lamme, et al., 2002). Moreover, any influence of SOA on the CRF tuning properties of these neurons was almost completely absent, showing feedforward activation was largely preserved. However, since measurements were done only in V1, it was unclear to what extent feedforward activation would continue to higher areas and to what extent the same would hold for humans.

The present study clearly demonstrates that in humans, undetected and strongly masked texture stimuli are processed far beyond striate cortex (see Figure 2.3), showing that feedforward processing is preserved, while recurrent processing is selectively interrupted. A neural network model of figure-ground segregation by Roelfsema, Lamme, Spekreijse and Bosch (2002) in which texture segregation in macaques is modeled in terms of feedforward and recurrent processing serves extremely well in explaining the spatio-temporal profile of such results.

Related models propose that recurrent processing is more intimately related to visual awareness (e.g. Lamme, 2001; Ro, et al., 2003). These views on recurrent processing are not necessarily contradicting, as visual awareness might simply be the way surface segregation and related processes express themselves phenomenologically. The view that recurrent activity in visual cortex correlates with visual awareness is now supported by converging evidence from monkey physiology (e.g. Lamme, Super, Landman, Roelfsema, & Spekreijse, 2000), EEG (the present study), TMS (Pascual-Leone & Walsh, 2001) and fMRI (e.g. Haynes, et al., 2005).

Conclusion

We argue and found evidence for at least three stages in early visual processing: (1) a completely unconscious pre-110 ms feedforward stage which is probably involved in boundary and object detection but goes uninterrupted by masking (2) a second stage, which operates in the 110-140 ms range reflecting recurrent processing in visual cortex, plausibly encapsulating processes such as figure-ground segmentation and phenomenal awareness and (3) a stage in the 200-300 ms range which is contingent upon earlier stages. Specifically the second stage was interrupted by backward masking, confirming that pre-110 ms activation proceeds outside awareness, whereas more posterior >110 ms activity is (at least) a necessary condition for awareness to arise. Finally, we reject explanations of backward masking as a unitary phenomenon, in agreement with other authors such as Bachmann et al. (2005). Any explanation of the masking phenomenon should carefully take into account physical stimulus characteristics and how these are processed by the various components of the visual system.

Chapter 3.

The spatiotemporal profile of cortical processing leading up to visual perception

Abstract

Much controversy exists around the locus of conscious visual perception in human cortex. Some authors have proposed that its neural correlates correspond with recurrent processing within visual cortex, whereas others have argued they are located in a frontoparietal network. The present experiment aims to bring together these competing viewpoints. We recorded EEG from human subjects that were engaged in detecting masked visual targets. From this, we obtained a spatiotemporal profile of neural activity selectively related to the processing of the targets, which we correlated with subjects' ability to detect those targets. This made it possible to distinguish between those stages of visual processing that correlate with human perception and those that do not. The results show that target induced extra-striate feedforward activity peaking at 121 ms does not correlate with perception, whereas more posterior recurrent activity peaking at 160 ms does. Several subsequent stages show an alternating pattern of frontoparietal and occipital activity, all of which correlate highly with perception. This shows that perception emerges early on, but only after an initial feedforward volley, and suggests that multiple reentrant loops are involved in propagating this signal to frontoparietal areas.

Fahrenfort, J. J., Scholte, H. S., & Lamme, V. A. F. (2008). The spatiotemporal profile of cortical processing leading up to visual perception. *Journal of Vision, 8*(1), -.

Introduction

Over the past 15 years, our view of the way information is processed in primate visual cortex has dramatically changed. It has become increasingly clear that information is not only processed hierarchically from bottom to top, but that interactions between cortical areas play a crucial role in the way the brain extracts information from its retinal input (Bar, et al., 2006; Hupe, et al., 1998; Lamme, Van Dijk, & Spekreijse, 1993). An ever growing body of literature shows extra-classical receptive field effects - also known as contextual modulation - occurring ~100 ms after stimulus onset in V1 and up (Lamme, 1995; Lamme, et al., 1999; Lamme & Roelfsema, 2000; Lamme, et al., 1993). These type of effects point to the influence of higher on lower visual areas and have been widely recognized to result from stimulus related recurrent processing within visual cortex (Dehaene, et al., 2006; Lamme, 2006; Roelfsema, et al., 2002). Early on, such findings were restricted to recordings in macaque V1, but recently EEG, fMRI and TMS studies have confirmed the importance of recurrent cortico-cortical interactions in human perception (Fahrenfort, et al., 2007; Haynes, et al., 2005; Murray, Boyaci, & Kersten, 2006; Pascual-Leone & Walsh, 2001).

However, it is difficult to uncover the spatiotemporal profile of such interactions. The temporal resolution of fMRI is too limited. In neurophysiology, it is difficult to simultaneously locate cells in high and low visual areas covering the same retinotopic area. Thus, when speaking about "recurrent processing" or "feedback processing", researchers often posit multiple iterations or loops, without being clear on the number or timing of these loops. For that matter, not everything is known about the function of recurrent processing either, although there are strong indications that it plays a crucial role in the function and phenomenology of figure-ground segregation and visual awareness (Fahrenfort, et al., 2007; Hupe, et al., 1998; Lamme, 1995; Lamme & Roelfsema, 2000; Lamme, et al., 1993).

In this experiment, we sought to identify the spatiotemporal profile of cortical processing in human visual cortex using EEG, and find out which activations correlate with perception and which ones do not. Although it is impossible to resolve the origin of EEG generators in the brain in an unconstrained manner (Nunez &

Srinivasan, 2006), a claim with respect to relative anteriority or posteriority of such generators can be made much more easily (Fahrenfort, et al., 2007; Foxe & Simpson, 2002). We calculated spherical surface Spline LaPlacian distribution maps of the EEG (Perrin, et al., 1989) and subsequently pooled electrodes in an anterior-posterior fashion, making claims only about the relative position of generators with respect to the front-back dimension. In combination with EEG millisecond timing information this method makes it possible to draw conclusions about the temporal order in which consecutive brain areas become active and about the modes of processing that subserve these activations (Fahrenfort, et al., 2007; Foxe & Simpson, 2002).

Subjects were asked to identify masked texture defined Figure and No Figure stimuli while ERP's (Event Related Potentials) were recorded (see Figure 3.1 and Figure 3.2). We calculated an ERP Figure minus No Figure difference wave in order to isolate figure induced activity, and correlated this with behavioral scores reflecting the ability of subjects to discriminate between Figure and No Figure trials. This was done for the entire spatiotemporal profile of cortical processing, allowing inferences about those aspects of processing that do and those that do not correlate with visual perception.

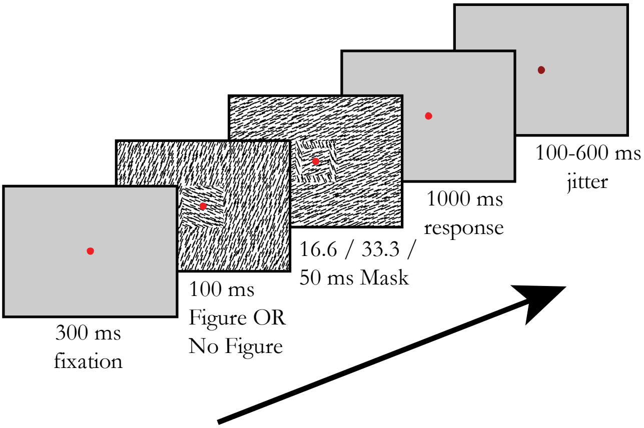
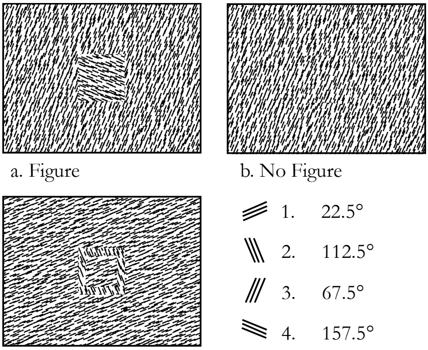


Figure 3.1 *Schematic time course of a trial. Trial started with 300 ms fixation, followed by a 100 ms target, followed by a 16.6, 33.3 or 50 ms mask. Subjects were given 1000 ms to indicate whether a figure was presented or not. The inter trial interval was jittered between 100 and 600 ms.*



a. Figure b. No Figure c. Mask d. Orientations used

Figure 3.2 *Examples of stimuli used in the experiment. (a) Figure stimulus (b) No Figure stimulus (c) Mask (d) Orientations used to configure target and mask stimuli. In any given trial, the orientation(s) used in the target stimulus would not be used again for the mask stimulus. Each condition consisted, on average, of the exact same sets of oriented textures.*

The behavior-EEG correlation shows an early bilateral feedforward signal which does not correlate with subjects' ability to distinguish Figure from No Figure targets even though there is a difference in the signal generated by Figure and No Figure stimuli. Slightly later in time, an extremely consistent posterior occipital generator shows a strong correlation with perception. In all likelihood, this generator is due to stimulus-related recurrent processing within visual cortex, as its timing and location are highly consistent with the effects of contextual modulation in early visual areas of the macaque (Lamme, et al., 2002, a very similar effect using human EEG has also been shown in Fahrenfort et al. 2007 and chapter 2 of this thesis). Due to its correlation with perception and its precession to parietal and frontal activations, this generator seems to act as the primary seed of perception. Within 200 ms, bilateral parietal and centrofrontal regions become active. These are followed at approximately 250-300 ms by concurrent frontal and occipital generators both of which correlate highly with perception, uncovering a distributed network of frontoparietal and occipital areas, plausibly involved in the transition of visual perception to a reportable stage (Dehaene, Sergent, & Changeux, 2003; Lamme, 2003, 2006; Lumer & Rees, 1999; Sergent, et al., 2005).

Methods

Participants

Nineteen Psychology students took part in the experiment in partial fulfillment of first year course requirements. All subjects (mean age 23.4, ±6.8) had normal or corrected-to-normal vision. 84% were right handed. 63% were female. Each subject provided written informed consent. All procedures were approved by the ethical committee of the University of Amsterdam.

Stimulation

Trial time course and example stimuli can be found in Figure 3.1 and Figure 3.2. At the start of each trial, a fixation dot on a grey background turned from dark into bright red, followed after 300 ms by texture target with a duration of 100 ms. In half of the trials this target contained an orientation defined square (Figure trials), in the other half the target was a homogenous texture (No Figure trials). Each target was followed by a 16.6, 33.3 or 50 ms pattern mask containing an orientation defined square annulus (Figure 3.2).

Backward masking, in which a target stimulus is followed shortly by a second stimulus, degrades the visibility of the target stimulus (Breitmeyer, 1984). There are many types of masking, with different explanations for their effectiveness (e.g. Fahrenfort, et al., 2007; Lamme, et al., 2002; Macknik & Livingstone, 1998). For reviews see (Breitmeyer & Ogmen, 2000; Enns & Di Lollo, 2000). The mask used in the present experiment was determined in a pilot phase to fit in well with the general phenomenology of the task and to maximally elicit variable responses between subjects. The aim here was to exploit inter individual differences in masking effectiveness, so visibility scores could be correlated with differences in the EEG signal over time. The different mask durations had no function other than making the graded response described below meaningful, as a single mask duration would not have resulted in differential responses.

Subjects were given 1000 ms to respond, after which the fixation dot turned dark red until the start of the next trial. This intertrial interval varied between 100 and 600 ms. All conditions were randomized and evenly distributed. A total of 6 blocks of 192 trials were recorded per subject.

Textures consisted of black (0.9 cd/m$^{2)}$ on white (104 cd/m$^{2)}$ line elements, spanning 0.06° and 0.37° of visual angle. Line elements could have four possible orientations: 22.5°, 67.5°, 112.5° and 157.5°. A No Figure target contained a single orientation; a Figure target consisted of two orthogonal orientations. The square in the figure trial subtended 2.47° of visual angle in the center of the screen. Masks consisted of an orientation defined square annulus of the same size as, and in the same central location as the Figure target, and consisted of orientations not used in the preceding target. Border thickness of the mask annulus was 0.39° of visual angle. All stimuli were isoluminant at 66.8 cd/m^2. Stimuli were created using Matlab (The MathWorks, Inc., Natick, MA, USA) and were presented using Presentation (Neurobehavioral Systems, Inc., Albany, CA, USA).

All texture orientations used in a trial were completely counterbalanced over conditions in such a way that Figure and No Figure trials were equal with respect to local stimulation. This was done in order to be able to carry out the EEG subtraction procedure detailed in the results section (also see Figure 3.5). For a similar procedure in the context of figure ground segregation and EEG/MEG see (Caputo & Casco, 1999; Fahrenfort, et al., 2007; Lamme, et al., 1992; Scholte, et al., 2006) or for other examples of subtraction procedures in EEG see (Hillyard, Hink, Schwent, & Picton, 1973; Thorpe, et al., 1996).

Task and behavioural measure

Approximately one week prior to the EEG session, all subjects were given a 30 minute training session to become acquainted with the task. After training they took part in an EEG session in which they carried out the same task. Subjects were instructed to distinguish between Figure and No Figure trials. With their right hand, they pressed a single button if they perceived a No Figure target (Not-Seen response), or one of three buttons (3-point scale) if they perceived a Figure target, depending on perceptual strength. The 3-point scale ensured subjects based their responses on their phenomenology, and not on guessing. Later, the 3-point scale was collapsed into a single response category (Seen response) and not used any further. For each subject, a perfect observer score was calculated based on Seen and Not Seen responses, reflecting a subjects' ability to distinguish between Figure and No Figure trials. The perfect observer score is a linear and subject bias' free measure derived from d' (Wickens, 2002).

EEG measurements and pre-processing

EEG was recorded from the scalp using a BioSemi ActiveTwo 48 channel active EEG system (BioSemi, Amsterdam, the Netherlands) at 256 Hz. 48 scalp electrodes were measured (referenced to the ears), as well as two electrodes for horizontal and two for vertical eye movements (each referenced to its opposite counterpart). The data were filtered (high pass > 0.5 Hz, low pass < 20 Hz, 50 Hz notch) and automatic artefact rejection was applied by removing segments containing voltage steps of more than 50 µV, segments falling outside the -200 µV to 200 µV range, as well as segments containing larger than 300 µV differences within the segment. Ocular correction was applied on the basis of the horizontal and vertical electro-oculograms (Gratton, et al., 1983). After ocular correction, artefact rejection was applied again by removing all segments outside the -50 µV to 50 µV range. Baseline correction was applied using the signal in the -300 ms to 0 ms interval. All EEG processing was done using Brain Vision Analyzer (Brain Products GmbH, Munich, Germany).

We converted all ERP signals using a surface Laplacian procedure by interpolating ERP signals to approximate scalp current densities (SCD's) (Perrin, et al., 1989) using spherical splines. SCD's are spatial second order derivatives of the potentials measured on the scalp. SCD's are reference free, and act as a band-pass spatial filter isolating signals due to sources localized in superficial cortex. The spatial resolution of SCD's is typically in the order of 2 to 3 cm in surface tangential directions (Nunez & Srinivasan, 2006).

EEG spatiotemporal maps

In order to construct spatiotemporal maps of the EEG, electrodes were pooled in a posterior-anterior fashion, starting with occipital electrodes and moving towards the front around both sides of the scalp (separately for the left and the right hemisphere). Electrode poolings were chosen in such a way that they were separated from each other by approximately 3 cm (see Figure 3.3). Because SCD's have a spatial resolution of around 2 to 3 cm, this ensured that the scalp current densities were accurate with respect to the front-back dimension. Separate poolings were made for the right and the left hemisphere. 2D spatiotemporal maps were constructed on the basis of these poolings (see Figure 3.4 for an example). Three electrodes (Fz, Cz and Pz) were left out of the poolings because they were located centrally and could therefore not be assigned to either left or right hemispheric poolings. Two electrodes were left out because they violated the 3 cm requirement (P1 and P2). Visualization and statistical analyses of spatiotemporal maps and other time courses were done with Matlab (The MathWorks, Inc., Natick, MA, USA) using custom code.

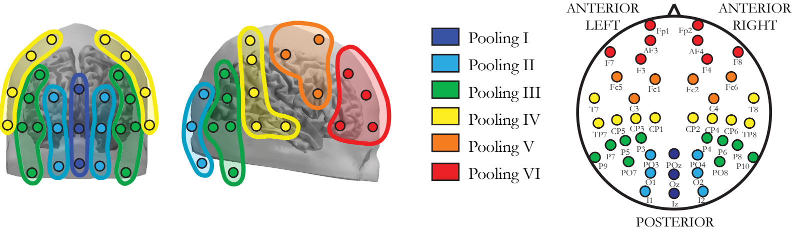


Figure 3.3 *Layout of the electrode poolings used in the experiment. The left half of the figure shows the electrodes displayed on the scalp. The right half shows the same electrodes on a flattened head view and the color scheme associated with the poolings.*

Statistical testing

Differences between experimental conditions were ascertained using paired t-tests. Corrections for multiple comparisons were made by limiting the FDR (False Discovery Rate), a method by which the expected proportion of falsely rejected hypotheses is controlled (Benjamini & Hochberg, 1995).

Formally, the FDR is given by $E\left[\frac{V}{V+S}\right]$ in which V is the number of false positives and S is the number of true positives. By applying the FDR correction, this value is kept below the threshold *q*, often set at .05 (although higher values can be acceptable depending on the research question). The formula used for finding the p-values at which this is true for a series of ordered p-values from small to large, is

$P(i) \leq \frac{i}{L} q$, in which *L(i)* is the temporal or spatiotemporal position corresponding to p-value *P(i)*, thus *L* is the number of temporal or spatiotemporal positions (i.e. number of tests being performed). Independence or positive correlation of tests is assumed. For an explanation of how the FDR is used in the field of neuroimaging see Genovese, Lazar, & Nichols (2002), for an application of the FDR in EEG see Fahrenfort et al. (2007).

Results

Figure 3.4 shows the spatiotemporal profile of cortical processing of Figure and No Figure trials on 2D spatiotemporal maps (4a and 4b respectively). Color represents the scalp current density at each spatiotemporal location. Electrode poolings are represented at each tick mark on the x-axis, occipital in the middle, left-frontal on the left and right-frontal on the right. Values in between poolings were interpolated using Spline interpolation. Time is represented on the y-axis.

The Figure and No Figure spatiotemporal maps in Figure 3.4 are remarkably similar to each other. Both maps show a strong occipital generator in the 100-150 ms timeframe, flanked by equally strong bilateral parietal generators, followed by two occipital generators later in time. Not surprisingly, this shows that the strongest cortical response due to these texture stimuli are early on and within (or close to) visual cortex. Because of the strength of these responses, it is difficult to infer small differences in cortical processing between Figure and No Figure trials from these raw spatiotemporal maps.

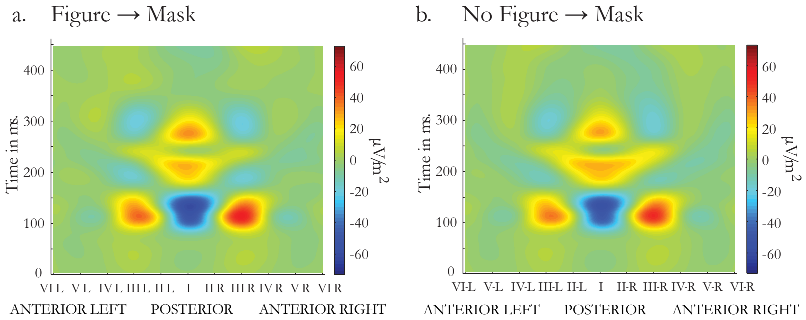


Figure 3.4 *2D spatiotemporal maps of (a) Figure followed by Mask and (b) No Figure followed by Mask processing. The y-axis represents time. Electrode poolings are at each tick mark of the x-axis, locations in between poolings have been spline interpolated. Color indicates the strength of the scalp current density at each spatiotemporal location.*

Therefore, all subsequent analyses were done on SCD difference waves. These were obtained by subtracting averaged No Figure trials from averaged Figure trials. An example of this procedure is shown in Figure 3.5. This figure shows the time course of the SCD for the occipital pooling for both Figure and No Figure trials, as well as for the difference between the two. It also shows where the difference is significant according to a False Discovery Rate of .05 (FDR, see the statistical testing paragraph in the methods section). Subtracting No Figure from Figure trials has two other major advantages:

1. The Figure minus No Figure subtraction isolates activity related to the processing of the figure. As Figure and No Figure targets are, on average, made up of the exact same sets of oriented textures (see Figure 3.2), any influence of local stimulation on cortical processing, such as caused by the line elements of the textures themselves, is subtracted out. The only signal left is related to the processing of differences in figure-ground organization between Figure and No Figure trials (for other examples on the topic of texture segregation see Caputo & Casco, 1999; Fahrenfort, et al., 2007; Lamme, et al., 1992; Scholte, et al., 2006).
2. By the same token, as the rest of the stimulus sequence is exactly equal between Figure and No Figure trials, any direct

contribution of other stimuli in the sequence, such as fixation dots and masks, is subtracted out as well. Masks of different durations were evenly distributed over Figure and No Figure trials, and were thus subtracted out.

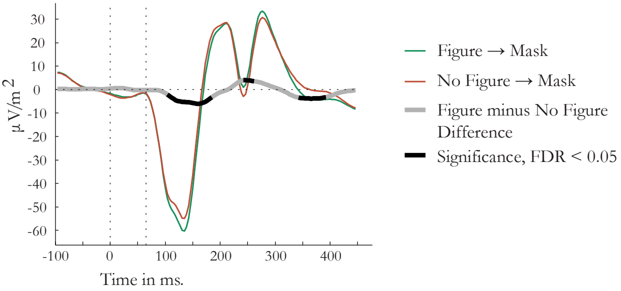


Figure 3.5 *Example of the Figure minus No Figure subtraction procedure on posterior pooling I. The SCD time course due to a Figure and a No Figure trial is shown for the occipital pooling. The difference between Figure and No Figure is also shown, in blue. A paired t-test between Figure and No Figure trials was performed at each time point in the time series. Significant differences between Figure and No Figure according to an FDR of .05 are shown in black.*

The spatiotemporal profile of the Figure minus No Figure subtraction is shown in Figure 3.6a. In this 2D map, color represents the difference in scalp current density between Figure and No Figure trials. The axes are the same as in Figure 3.4. To evaluate differences in cortical processing between Figure and No Figure trials, a random effects analysis was performed by employing a paired two-tailed t-test between Figure and No Figure averages at each space-time point in the spatiotemporal map in Figure 3.6a, treating the average of each subject at that space-time point as an observation. The correction for multiple comparisons with respect to the number of tests was done by limiting the False Discovery Rate (FDR, see the statistical testing paragraph in the methods section), a method by which the p-value at which significance is evaluated is corrected for the number of tests being performed (Benjamini & Hochberg, 1995). The spatiotemporal locations for which the difference between Figure and

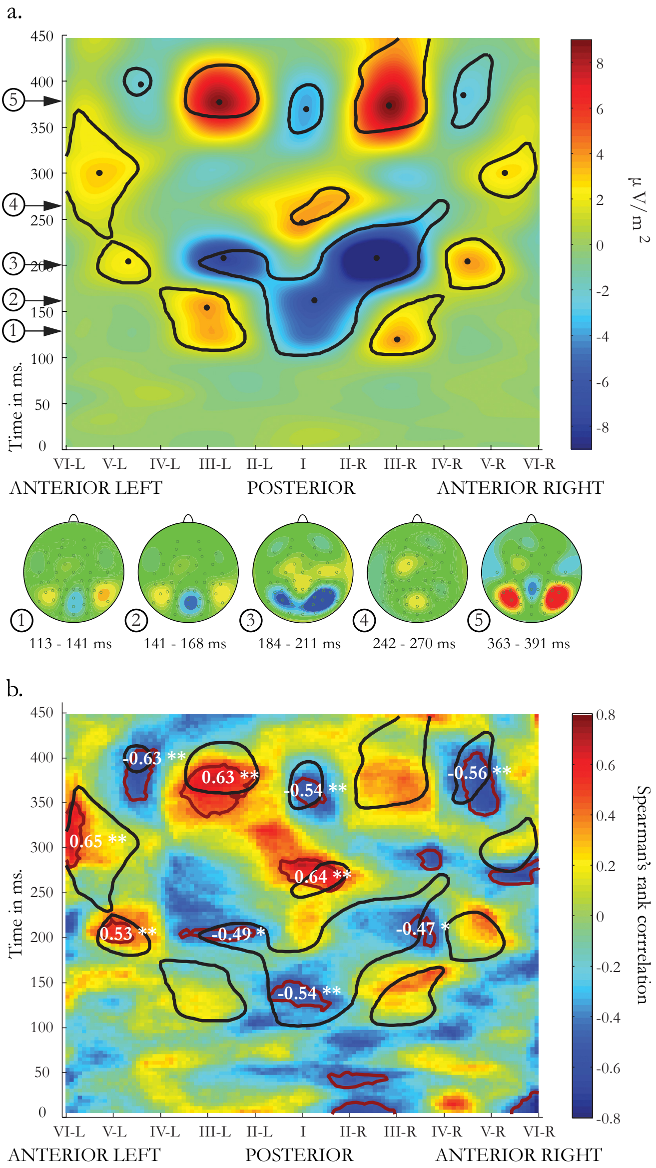


Figure 3.6 *Legend on next page*

Figure 3.6 *(a) Figure minus No Figure activity. The top shows the 2D spatiotemporal map for the difference between Figure and No Figure processing (effectively the difference between Figure 3.4a and Figure 3.4b). The y-axis represents time. Electrode poolings are at each tick mark of the x-axis such that the middle is the most posterior (occipital) location and the left and right of the axis are the most anterior (frontal) left and right positions. Color indicates the strength of the difference between Figure and No Figure trials in scalp current density at each spatiotemporal location. A paired t-test between Figure and No Figure trials was performed at each spatiotemporal location in the map. The solid black lines show the areas within which these differences were significant, corrected for multiple comparisons at an FDR of .05. Solid dots indicate local maxima and minima. On the left of the figure, five relevant stages (see results section) are indicated by encircled numbers with pointing arrows. For each of these stages, a topographic map is shown at the bottom. (b) Figure minus No Figure correlation with detection. The lower figure shows the correlation between subject's ability to discriminate between Figure and No Figure trials and the Figure minus No Figure difference for each spatiotemporal location in the map. Color represents the strength of Spearman's rank correlation. The solid black lines are the same as in the top figure; the solid red lines represent areas within which the correlations are significant at the .05 level. In white the correlations are given for the areas that are enclosed by both black and red lines. Level of significance is indicated by asterisks: *$p<0.05$, **$p<0.01$ (one-tailed). Spurious activations of significant clusters smaller than 25 spatiotemporal locations were not reported in either figure.*

No Figure was significant is encapsulated by solid black lines, corrected for multiple comparisons at an FDR of .05 ($q = .05$). The solid dots in Figure 3.6a indicate local minima and maxima of the SCD difference. On the bottom of Figure 3.6a topographic plots of the critical time windows are shown to provide an unambiguous description of the spatial distribution of the effects, thus confirming the overall picture.

Figure 3.6a clearly shows that processing does not occur hierarchically from bottom to top (i.e. from center to edges in Figure 3.6a) in a feedforward fashion, but that massive activation of early visual areas occurs up to at least 400 ms after stimulus presentation, long after more frontal areas have been recruited. These activations are likely to reflect both sustained local processing (Foxe & Simpson, 2002), recurrent interactions within visual areas (Fahrenfort, et al., 2007; Foxe & Simpson, 2002; Lamme, 1995), as well as long range interactions between frontal and visual areas (Lumer & Rees, 1999; Rodriguez, et al., 1999). From Figure 3.6a we can infer a number of stages. Each stage is indicated by an encircled number on the left of the figure, for each of which the scalp topographic flat map is shown on the bottom:

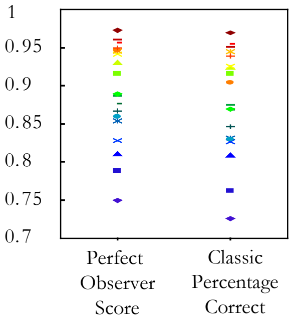
1. A bilateral parietal generator peaking at 121 ms (collapsed average: 121 ms, right: 117 ms / left: 152 ms). A meta-analysis of studies employing macaque intracranial recordings (Lamme & Roelfsema, 2000) has shown average response times in early visual areas of 72 ms (V1), 84 ms (V2) and 77 ms (V3). Dorsally these continue to 129 ms (V7a), 92ms (V7ip) and ventrally to 106 ms (V4). Given the fact that the earliest Figure - No Figure differences here peak parietally at 121 ms, it is not unlikely that they reflect sustained activity resulting from feedforward processing, although it cannot be ruled out that some feedback is already incorporated at this interval (Foxe & Simpson, 2002). Note that we report peak ERP latencies and average response latencies, not onset latencies, which are considerably shorter.
2. A more posterior occipital generator peaking at 160 ms which, due to the fact that it is later in time *and* more posterior than the bilateral 121 ms generator is probably due to a combination of feedforward and feedback activity within early visual cortex (Fahrenfort, et al., 2007; Foxe & Simpson, 2002). Incidentally, this interpretation is in excellent agreement with a large number of studies showing contextual modulation due to recurrent processing in this time frame in V1 and up using highly comparable stimuli (Lamme, 1995; Lamme, et al., 1993; Super, et al., 2001). Several other authors have identified EEG correlates of conscious vision around this time window, some earlier peaking around 100 ms (Fahrenfort, et al., 2007; Pins & ffytche, 2003) and some later starting at 130 ms and peaking later on in time (Koivisto, Revonsuo, & Lehtonen, 2006; Koivisto, Revonsuo, & Salminen, 2005).
3. Occipitoparietal and centrofrontal regions peak at around 200 ms. A negativity starting around 200 ms is typically reported as part of the N2 family of components, of which the most notable in this context is the N2pc (N2 posterior-contralateral). This component is largest at posterior scalp sites and is observed over the hemisphere contralateral to the location of an attended object (given that the target stimulus is not located centrally). It has been suggested to reflect perceptual-level attentional selection, for instance to zoom in on a target within an array of distractors

(Luck, Girelli, McDermott, & Ford, 1997). This component has been shown to occur virtually unimpaired even when a stimulus is unreportable due to object substitution masking (Woodman & Luck, 2003).

4. A more posterior occipital generator peaks at 246 ms, with a concurring frontal generator. Given their timing and approximate concurrence these may be engaged in long range coordinated recurrent activity enabling conscious access (Dehaene, et al., 2006; Lamme, 2006; Lumer & Rees, 1999; Rodriguez, et al., 1999).
5. Strong recurring occipitoparietal generators appear at 350-400 ms, flanked by centrofrontal generators, which may well reflect a third iteration of recurrent processing within and/or between these areas. A posterior-parietal component in this time window is classically reported as the P3 or P300, referring to a third positivity (or a positivity around or after 300 ms) in the ERP waveform. The P3 has been associated with a number of psychological variables, the most prominent of which are working memory and attention (Donchin & Coles, 1988; Kok, 2001). More recently, it has been suggested to be a signal related to the outcome of internal decision making processes (Nieuwenhuis, Aston-Jones, & Cohen, 2005). Accordingly, activity in this and the previous time window has generally been observed to be attenuated by attentional manipulations such as the attentional blink (e.g. Koivisto, et al., 2006; Kranczioch, Debener, & Engel, 2003; Sergent, et al., 2005).

In Figure 3.6b we show which of these stages correlate with perception and which ones do not. For each subject the perfect observer score was calculated, reflecting his or her ability to detect masked figures. Perfect observer scores and classic percent correct are shown in Figure 3.7 for each subject. We calculated the correlation between these perfect observer scores and the Figure minus No Figure difference for the entire spatiotemporal profile in Figure 3.6a. Thus, for each space-time point in the spatiotemporal map from Figure 3.6a, Spearman’s rank correlation was computed between subjects’ average Figure minus No Figure difference SCD and subjects’ perfect observer score at discriminating Figure from No Figure trials. The result is shown in Fig 3.6b, the strength of the correlations in color. The black lines enclose the spatiotemporal locations where the Figure minus No Figure difference is significant, as redrawn from Figure 3.6a.

Figure 3.7 *For each subject, the perfect observer score for the Figure detection task as well as the corresponding classic percentage correct score is shown. Subjects are uniquely identifiable by both shape and color of the markers.*



The dark red lines enclose the spatiotemporal locations within which the correlations between detection accuracy and the SCD difference wave are significant at the .05 level. In white, the correlations are given for each of the areas where both the difference and the correlations are significant (i.e. those areas within which the black lines and the dark red lines overlap). Only correlations in spatiotemporal locations where Figure and No Figure significantly differ were reported, so as to exclude correlations that occurred outside of the periods of neural activity related to the processing of figure from ground. Note for example, that we also found significant correlations (i.e. dark red circles in Figure 3.6b) at about stimulus onset (0 ms) in the right parietal and frontal regions. These might reflect attentional set being higher at trials in which detection is successful, and will not be directly related to the processing of the figure stimulus per se (also see Super, van der Togt, Spekreijse, & Lamme, 2003). Their location is consistent with right hemispheric dominance for attention for the entire visual field (e.g. Heilman & Abell, 1980; Mesulam, 1999). Alternatively, they may reflect spurious correlations, as calculating such large numbers of correlations may produce significant results even when fitting noise.

The map shows that the first bilateral parietal generators due to feedforward processing (stage 1 above, also see Figure 3.8a and

3.8b) do not correlate with subjects' ability to detect a figure, whereas the later occipital generator due to recurrent processing (stage 2 above, also see Figure 3.8c) does. As this occipital activation is the first one to show a strong correlation with perception, and almost all ensuing correlations are highly significant, it seems to act as a seed for further correlations.

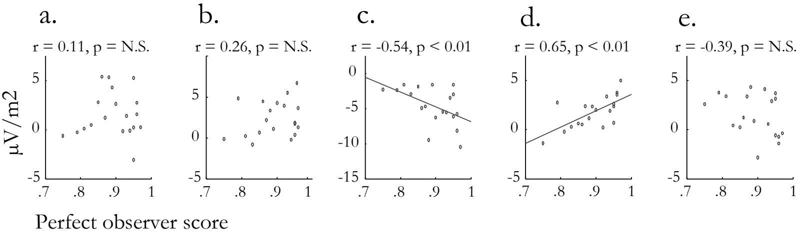


Figure 3.8 *Graphs showing data points and their fit for 5 selected relevant generators. For significant correlations these represent clusters of spatiotemporal locations within the generator where the correlation had a p-value < .05. For non-significant correlations these represent the 50% contiguous spatiotemporal locations having correlations with the lowest p-values within that generator. (a) the generator peaking at 152 ms at pooling III-L attributed to feedforward processing. (b) the generator peaking at 117 ms at pooling III-R attributed to feedforward processing. (c) the generator peaking at 160 ms at occipital pooling I attributed to feedback processing. (d) the generator peaking around 300 ms on the anterior left (e) the generator peaking around 300 ms on the anterior right.*

Further evidence of the importance of this generator in visual processing comes from the fact that it is by far the most consistent difference between Figure and No Figure. When reducing the FDR is to 0.0001, the only significant generator surviving this overly strict threshold is this occipital one, reflecting the fact that it represents a highly consistent difference between Figure and No Figure processing.

Later generators show an alternating pattern of anterior and posterior activity, most of which correlate with perception, although anterior correlations appear more in the left (also see Figure 3.8d) than in the right hemisphere (also see Figure 3.8e). This left-hemispheric dominance may be caused by the fact that subjects had to report Figure presence with their right hand, although correlations with activations due to motor preparation and response are unlikely because a response was always required and always given with the right hand (and should thus be subtracted out of the Figure minus No Figure difference). Also, language is known to be predominantly left hemispheric (Vigneau, et al., 2006). Left hemispheric dominance of correlations could therefore partly be due to the fact that subjects had to give an appraisal of stimulus strength that may have been verbalized mentally.

We would like to stress that the descriptions of generators and correlations found are not exhaustive in terms of the neural activity that underlies them. EEG activity is caused by coordinated postsynaptic activity of huge cell assemblies producing dynamic patterns of electric potential on the scalp. Aside from the inverse problem, the skull is also beset by problems of volume conduction, leaving us with a very coarse reflection of neural activity (Nunez & Srinivasan, 2006). Thus, multiple coherent neural events may show up as a single generator, or may not show up at all. A single coherent neural event may even show up as multiple distinct generators as the polarity of the difference between experimental conditions shifts over time. Therefore the only aim in this experiment is to embed the generators and correlations that were identified in a coherent picture of cortical processing given knowledge we have from other sources such as monkey physiology, and not to give a comprehensive description of all cortical processing.

Discussion

With the employed subtraction paradigm we isolated signals that discriminate between Figure and No Figure stimuli. Thus, we obtained a spatiotemporal profile of the EEG activity that is induced by a texture figure. We could discern 5 stages in this signal, starting with a signal resulting from initial feedforward activation, followed by signals reflecting recurrent loops of feedback and feedforward activity within visual cortex, and later between more distant cortical areas. Correlating this activity with behavior, i.e. the ability to detect figures when masked, shows that the early feedforward activation of the parietal cortex does not correlate with perception. This is in line with many studies showing that feedforward processing goes uninterrupted in visual cortex even when a subject is fully unconscious of the stimulus (Fahrenfort, et al., 2007; Lamme, Zipser, et al., 1998; Lamme, et al., 2002).

More posterior in the brain, and later in time, an occipital generator plausibly due to recurrent processing correlates highly with perception. This suggests that the correlates of perception and visual awareness start to emerge at around 100 ms due to recurrent processing, and are propagated further along the system through multiple recurrent loops (also see Dehaene, et al., 2006; Lamme, 2006). This view is consistent with many studies showing the importance of recurrent processing in figure-ground segregation and visual awareness (Fahrenfort, et al., 2007; Haynes, et al., 2005; Hupe, et al., 1998; Lamme & Roelfsema, 2000; Lamme, et al., 1993; Scholte, et al., 2006; Silvanto, Cowey, Lavie, & Walsh, 2005).

As have others, the present study thus stresses the importance of the 100+ ms activation of visual cortex in human perception (Fahrenfort, et al., 2007; Koivisto, et al., 2006; Koivisto, et al., 2005; Pins & ffytche, 2003; Wilenius-Emet, Revonsuo, & Ojanen, 2004). Although the exact timing differs somewhat between these studies - which may relate to differences in stimulus complexity and on screen duration - they are all consistent with the idea that visual awareness emerges as a result of recurrent activity in visual cortex.

Confirmation of this idea can also be found in studies on the topic of high-level perceptual decision making. A number of studies have employed linear regression techniques to create single trial predictions about subjects' performance in perceptual decision tasks, such as discriminating between a face and car (Philiastides, Ratcliff, & Sajda, 2006; Philiastides & Sajda, 2006; M. L. Smith, Gosselin, & Schyns, 2004). These studies suggest that EEG components reflecting early visual perception can be found occipitally at ~170 ms (plausibly incorporating feedback mechanisms), whereas a late ~300 ms component reflects a postsensory/decision stage.

This is not to say that stimulus categorization cannot be triggered by non-perceptual (feedforward) events. For instance, face tuning starts in inferotemporal cortex by ~100 ms (Liu, Harris, & Kanwisher, 2002; Oram & Perrett, 1992) and Thorpe (1996) has shown that frontal cortex starts detecting the presence of an animal in a natural scene by ~150 ms. VanRullen and Koch (2003) have shown that subjects can perform such stimulus categorizations even when highly confident that they did not see the (masked) test stimulus. Crucially, the earliest time at which subjects start categorizing unperceived scenes with above chance performance matches exactly the time at which they do so for consciously perceived scenes. This leads these researchers to conclude that both conscious and unconscious categorization is initially triggered by the same (unconscious) feedforward process. In such a scheme, "perceptual" or high-confidence decisions logically depend on subsequent recurrent activity (for a similar argument see Jolij, Scholte, Van Gaal, & Lamme, in press).

A large number of physiological studies show correlates of figure-ground segregation in V1 in the 100+ ms time window due to recurrent processing, which fits well with the 100+ ms timing found in the present experiment (e.g. Lamme, et al., 1993) and which has been argued to be a correlate of visual awareness. Although the exact nature of cortical recurrent processing may be more dynamic than can be uncovered with EEG, the present study suggests that in humans too, the phenomenology of figure ground segregation may originate from early recurrent processing in visual cortex.

Conclusions

The first correlates of perception emerge right after an initial (automatic) feedforward sweep, which does not correlate with subjects' ability to detect a figure from a background. All recurrent and feedforward activity after this seed correlates with the ability of subjects to perceive a figure. There are multiple recurrent loops involved in visual perception spanning the entire human cortex in the 100-450 ms timeframe.

Chapter 4.

Tunnel vision in visual cortex: Suppression of cortical activity in the periphery of figure-ground segregation

Abstract

Although there is strong evidence that extra-classical receptive field effects play an important role in figure ground segregation, the exact mechanisms which underlie it remain incompletely understood. In order to characterize these mechanisms more clearly, we set out to determine the distribution of cortical responses to lightly masked figure-ground stimuli in early visual areas using fMRI. We found two effects in striate and extrastriate cortex, which together constitute what could be construed as cortical tunnel vision: enhanced cortical processing at the location of the figure surface and strong background suppression at increasingly peripheral locations around the figure surface, even though local physical stimulation is – on average – the same. These effects extend way beyond the classical receptive field of neurons in V1. Similar effects have previously been observed in the literature on endogenous top down attention, but not as a sensory-driven exogenous effect. When splitting up the data in seen and unseen figure trials, suppressive effects remained intact for both seen and unseen trials, whereas enhanced processing at the location of the figure surface was selectively abolished in unseen trials. From this we conclude two things: (1) Suppression of the (far) surround occurs in V1 and up, and is not exclusively a top down attentional phenomenon but may also result from sensory-driven processing within visual cortex due to the global organization of the stimulus. (2) Enhanced activity at retinotopic locations in V1 corresponding to the figure surface is only present for seen trials and not for unseen trials.

Fahrenfort, J. J., Scholte, H. S., van Gaal, S. & Lamme, V. A. F. (Submitted) Tunnel vision in early visual cortex: Suppression of cortical activity in the periphery of figure-ground segregation.

Introduction

In top down spatial attention, a type of biasing seems to take place in retinotopic cortex with respect to location. Facilitation takes place for attended locations (Gandhi, Heeger, & Boynton, 1999; Martinez, et al., 1999) and suppression occurs for unattended locations (Slotnick, Schwarzbach, & Yantis, 2003; A. T. Smith, Singh, & Greenlee, 2000; Tootell, et al., 1998). The co-occurrence of these mechanisms can be likened to the neural equivalent of what is often termed "the spotlight of attention" (Brefczynski & DeYoe, 1999); a sort of cortical tunnel vision in which processing at attended locations is enhanced at the expense of locations that are not attended. Notably, these effects have been shown in the earliest human visual area, V1, which points to a role for feedback mechanisms. Such top down influences are commonly thought to originate from frontoparietal cortex (Desimone & Duncan, 1995; Kastner & Ungerleider, 2000), although it is difficult to ascertain causality and origin with certainty (Miller & D'Esposito, 2005; Sirotin & Das, 2009).

Many physiological studies have shown that neurons in V1 can also be influenced by context information due to physical stimulation outside their classical receptive field (cRF) (Cavanaugh, Bair, & Movshon, 2002; Knierim & Van Essen, 1992). In studies of figure ground segregation, enhancements caused by stimulation outside the cRF have been dubbed contextual modulation (CM). These are probably caused by top down feedback from higher areas within visual cortex (Lamme, Super, et al., 1998), and seem to play an important role in segregating a figure from its background (Lamme, 1995; Roelfsema, et al., 2002; Zipser, et al., 1996). CM in the context of figure ground segregation has also been shown in humans (Skiera, Petersen, Skalej, & Fahle, 2000), where it has been demonstrated to continue to operate even during complete inattention (Scholte, et al., 2006) and independently of attention (Heinrich, Andrés, & Bach, 2007).

Thus, unattended texture segregation seems to enhance cortical responses to the segregating surface in V1 and up (Scholte, et al., 2006), while voluntary attention in the absence of visual stimulation, has been shown to do the same thing; enhancing neural responses in early visual areas which retinotopically correspond to the attended region (Kastner, Pinsk, De Weerd, Desimone, & Ungerleider, 1999). This suggests that even in cases where spatial attention and figure-ground segregation have been functionally dissociated, they both employ neural feedback mechanisms. Because they seem to employ similar mechanisms, some have hinted that they may recruit the same neural pathways (Roelfsema, 2006; Spratling & Johnson, 2004), although the fact that they can be dissociated shows that they are functionally different.

In this experiment, we explore whether "cortical tunnel vision" in early visual areas - as found in studies of endogenous spatial attention - may also play a role in a more stimulus driven context, such as texture segregation. We used fMRI to measure brain activity while showing lightly masked figure-ground stimuli. In order to exclude the possibility that observed effects are caused by the recruitment of endogenous top down attention, we separately analyzed seen and unseen trials, the latter of which should not draw endogenous attention. Cortical suppression of the surround was found in both seen and unseen trials. Enhancement of cortical responses to the figure surface was only present for seen trials, and strongest in V1.

Materials and Methods

Participants

Nineteen subjects were paid to participate in the experiment. All subjects (mean age 21.5, ±3.6) had normal or corrected-to-normal vision. 79% were female. All subjects provided written informed consent and the study was approved by the ethical committee of the University of Amsterdam.

Stimulation

Stimuli were presented using a DLP beamer which projected onto a projector screen in the scanner room. Subjects viewed the projector screen via a mirror attached to the head coil. Projection size was 16.9° by 12.7° visual angle. The time course of stimulation during a typical trial is shown in Figure 4.1. Example stimuli are shown in Figure 4.2. At the start of each trial, a fixation dot on a grey background would turn from dark into bright red, followed after 300 ms by texture target with a duration of 100 ms. In half of the trials this target contained an orientation defined square (Figure trials), in the other half the target was a homogenous texture (No Figure trials). All targets (both Figure and No Figure trials) were followed by a pattern mask containing an orientation defined square annulus (Figure 4.2) covering the same area as the target.

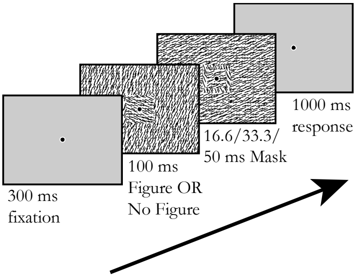


Figure 4.1 *Example of a typical trial. The fixation dot lights up to bright red (shown here in black with white outline) 300 ms prior to target presentation. The target, which is either a Figure or a homogenous No Figure stimulus, is on screen for 100 ms, followed by a 16.7, 33.3 or 50 ms mask. The subject has 1000 ms to respond, after which the fixation turns back to dark red. The intertrial interval was 1800 ms.*

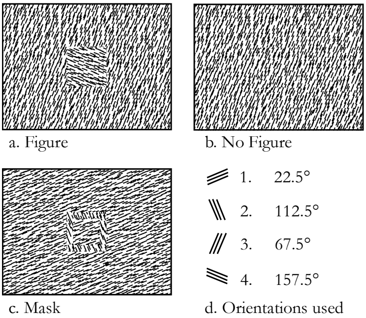


Figure 4.2 *Example of stimuli. The target stimulus consisted of either a texture defined square (a. Figure) or a homogonous texture (b. No Figure). All targets were followed by a pattern mask (c). The texture mask contained an orientation defined annulus of which the outer borders coincided with the outer borders of the Figure target. Four orientations were used (d), which were balanced out over conditions so that on average, local physical stimulation for the Figure and No Figure trials was exactly equal.*

Mask duration was varied between 16.7, 33.3 or 50 ms. for no other reason than to make the graded response scheme meaningful as a single mask duration would not have resulted in differential responses (see 'Task and behavioral measure' below). Subjects were given 1000 ms to respond, after which the fixation dot turned dark red until the start of the next trial. The intertrial interval was 1800 ms. All conditions were randomized and evenly distributed within each run. Textures consisted of black (0.2 cd/m$^{2)}$ on white (46.4 cd/m$^{2)}$ line elements, spanning approximately 0.05° and 0.30° of visual angle, which were spaced apart by an average of 0.20° with a jitter of 0.13° visual angle. Line elements could have four possible orientations: 22.5°, 67.5°, 112.5° and 157.5°. A No Figure target contained a single orientation; a Figure target consisted of two orthogonal orientations. The square in the figure trial subtended 2.0° of visual angle in the center of the screen. Masks consisted of an orientation defined square annulus of the same size as, and in the same central location as the Figure target, and consisted of orientations not used in the preceding target. Border thickness of the mask annulus was 0.33° of visual angle. Stimuli were created using Matlab (The MathWorks, Inc., Natick, MA, USA) and were presented using Presentation (Neurobehavioral Systems, Inc., Albany, CA, USA).

All texture orientations used in a trial were counterbalanced over conditions in such a way that Figure and No Figure trials were equal with respect to local stimulation. This was done in order to be able to carry out the fMRI subtraction procedure detailed later on.

Task and behavioral measure

Approximately one week prior to the fMRI session, all subjects were given a 30 minute training session to become acquainted with the task. After training they took part in one or two fMRI sessions in which they carried out the same task. Subjects were instructed to distinguish between Figure and No Figure trials. A response was always required. With their right hand, they pressed a single button if they were absolutely sure they did not perceive a Figure target (Unseen response), or one of three buttons (3-point scale) if they perceived a Figure target - however faint - depending on perceptual strength. This graded response scheme ensured subjects based their responses on phenomenology and not on guessing, making sure that Unseen responses were truly not seen. It biases the response data towards a conservative measure for invisible trials, rather than diluting the Unseen category with visible trials. The 3-point 'Seen' scale was collapsed into a single response category (Seen response) and not used any further. Separate analyses were carried out on Seen and Unseen trials. To calculate overall visibility for each subject, a perfect observer score was calculated based on Seen and Unseen responses, reflecting a subjects' ability to distinguish between Figure and No Figure trials. The perfect observer score is a linear and subject bias' free measure derived from d' (Wickens, 2002). It runs from 0 for chance performance to 1 for perfect performance.

fMRI measurements and pre-processing

Recordings were done on a Philips 3T Intera scanner. A total of 32 sessions were scanned using 19 subjects with an average of 5 runs per session. A run started and ended with 16 seconds rest and lasted approximately 10 minutes. The functional recordings were acquired using a T*-weighted sequence (TR 2261 ms, TE 28 ms, 35 slices, slice thickness 2.5 mm; slice gap 0.3 mm, FOV 200 mm). Each run consisted of 288 trials, of which 96 trials were dummy trials intended to improve deconvolution and counteract BOLD saturation in visual areas.

Analyses were done using BrainVoyager QX (Brain Innovation B.V., Maastricht, the Netherlands) and Matlab (The MathWorks, Inc., Natick, MA, USA). The functional images were motion corrected and slice time aligned. No spatial or temporal smoothing was applied. Images were then aligned to a structural image of the subject and transformed to Talairach space.

Regions of interest

For each subject, we extracted the activity of several visual areas. These areas were functionally defined using standard correlational mapping procedures (Wandell, 1999) using runs in which we presented an expanding/contracting wedge (excentricity mapper) and a rotating wedge (polar mapper) which were obtained in a prior experiment. For each subject we used the averages of two eccentricity and two polar mappers. The following areas were mapped: V1, V2, V3, V3a, V3b, V4, V7 (Brewer, Liu, Wade, & Wandell, 2005; Wandell, Brewer, & Dougherty, 2005). Furthermore, we subdivided visual areas V1, V2 and V3 in three regions of differing eccentricity using the eccentricity mapper, in degrees of visual angle: 0° - 1.9° (center region), 1.9° - 5.9° (near surround region) and 5.9° – 12.7° (far surround region), see Figure 4.3. The center region lies within the retinotopic region covered by the Figure stimulus; the near region is the region directly surrounding the Figure stimulus, while the far surround region represents the far surround. The other visual areas V3a, V3b, V4 and V7 were subdivided in just two eccentricities due to difficulties in accurately mapping three regions in these higher areas: 0° - 1.9° (center region) and 1.9° - 12.7° (surround region).

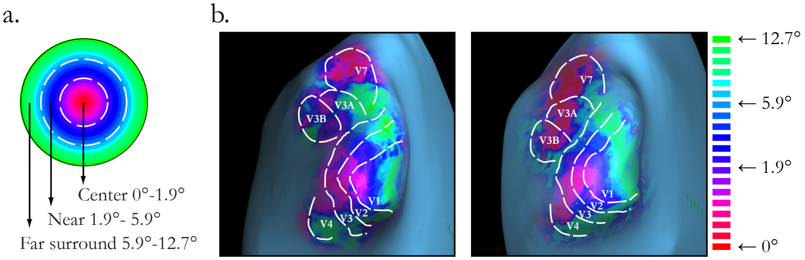


Figure 4.3 *a) Color coding of eccentricity phase across the visual field as well as the spatiotopic subdivision in eccentricity bands of early visual areas V1, V2 and V3. The center region covered an area of approximately 0°-1.9° visual angle, the near surround covered an area of 1.9°-5.9° and the far surround an area of 5.9° to 12.7°. Note that the cortical magnification factor was applied to the physical stimulation in the eccentricity mapper so that degrees across the visual field are not linearly represented in this schematic. b) Left hemispheric eccentricity phase data of two arbitrarily chosen subjects projected on the occipital pole of their respective inflated brains. Retinotopic areas are demarcated by white dashed lines. Eccentricity is color coded, legend on the right. The 0°-1.9° eccentricity band is coded in red, the 1.9°-5.9° band in blue and the 5.9° to 12.7° band in green. For V3A, V3B, V4 and V7 the outer eccentricity band regions of interest were collapsed into a single surround band.*

Analyses

The functional data was normalized towards percent signal change. Percent signal change over time was estimated for each session, region of interest and condition by running an event related Finite Impulse Response (FIR) deconvolution analysis (Glover, 1999). For each condition, eight predictors were used to code eight consecutive single delay responses from stimulus onset onwards. By fitting a GLM using these predictors, an eight-point haemodynamic response was deconvolved covering a time window of approximately

18 seconds. To give an unbiased depiction of deconvolved BOLD activity, no post-hoc baseline corrections were applied to the deconvolved data. The deconvolved data for each session, region of interest and condition were exported to Matlab for statistical testing and visualization. The No Figure condition was subtracted from the Figure condition; differences between the two conditions for were tested using paired t-tests separately for each time point. Subtracting No Figure from Figure trials has three major advantages:

1. The Figure minus No Figure subtraction isolates activity related to the processing of the figure. As Figure and No Figure targets are made up of the same sets of oriented textures (see Figure 4.2), the influence of local stimulation on cortical processing, such as caused by the line elements in the textures themselves, is subtracted out. The signal left is related to processing of differences in boundaries and figure-ground organization between Figure and No Figure trials. For other examples of this subtraction procedure using figure-ground stimuli see Caputo & Casco (1999), Fahrenfort, Scholte , & Lamme (2007) and Scholte et al. (2006).
2. By the same token, as the rest of the stimulus sequence is exactly equal between Figure and No Figure trials, any direct contribution of other stimuli in the sequence, such as fixation dots and masks, is subtracted out as well. Masks of different durations were balanced between Figure and No Figure trials, and were thus subtracted out.
3. Influences of top-down attention are subtracted out: subjects do not know in advance whether a Figure or a No Figure stimulus will be presented. Therefore the amount of (pre-stimulus) anticipatory top down attention for Figure and No Figure stimuli will – on average - be equal.

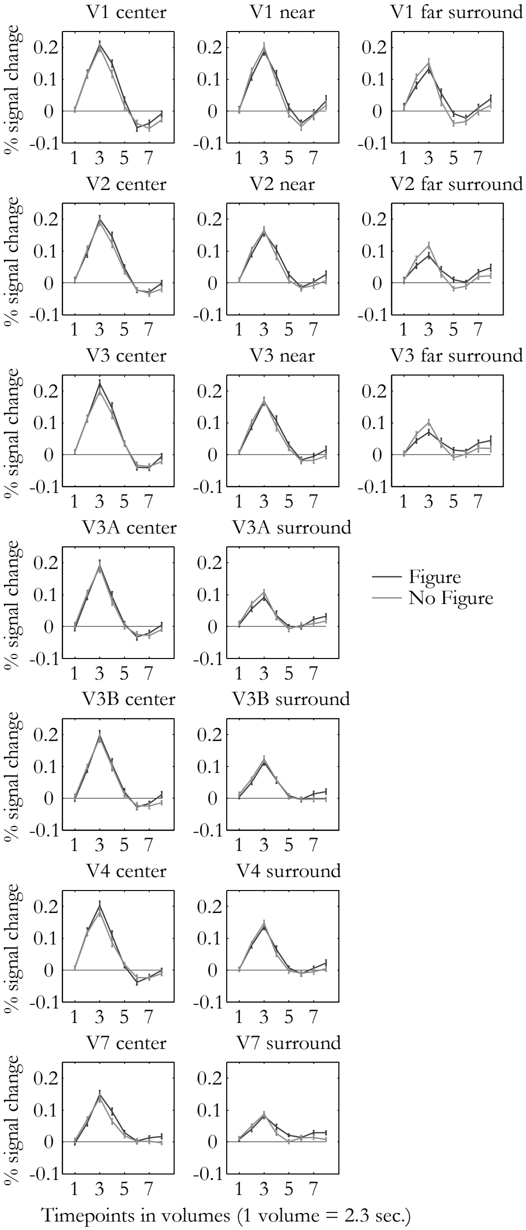
The Figure minus No Figure subtraction was also done separately for Seen and for Unseen responses. For this analysis, trials were selected in which subjects either indicated they saw a Figure (Seen), or positively did not see a Figure (Unseen). For both response types, post-hoc counterbalancing was done within each session, such that on average all local orientations and mask durations were equal for Figure and No Figure trials. Figure or No Figure trials that could not be used because they contained an orientation and/or mask duration that could not be matched to their counterpart within that response type were modeled separately and later on discarded. Consequently, local stimulation was balanced out between the Seen Figure and No Figure trials on the one hand, and between the Not Seen Figure vs. No Figure trials on the other hand. This ensured that the Figure minus No Figure subtraction procedure detailed above would retain its validity with respect to physical stimulation. If no balanced design for both response categories could be obtained for a particular session, that session would be discarded. 23 sessions were used for the Seen and Unseen analysis.

Results

The average perfect observer score for discriminating between Figure and No Figure trials for all three mask durations was 0.81 (±0.11), indicating that masking had an effect but that many Figure trials were detected. Figure 4.4 shows the average deconvolved haemodynamic responses to cortical processing of Figure and No Figure trials in early visual areas, subdivided into two and three eccentricities (see Figure 4.3 and methods section), and collapsed over mask duration. Error bars indicate ± one standard error of the mean. A typical haemodynamic response can be observed for all areas and conditions, with a peak around volume 3 and undershoot around volume 6. Naturally, all visual areas responded strongly to these texture stimuli. Because the relative contribution of the processing of high contrast textures is much larger than the contribution of global texture organization, the differences between Figure and No Figure stimuli is rather small. To highlight this difference, we subtracted Figure from No Figure trials (see Figure 4.5).

Subtracting Figure from No Figure trials has a number of advantages, more extensively highlighted in the Methods section under Analyses: (1) Activity related to the processing of the figure is isolated as influences of local contrast and orientation processing are subtracted out, (2) direct contributions of other stimuli in the sequence are subtracted out and (3) the amount of pre-stimulus top-

Figure 4.4 *Shows the estimation of the haemodynamic response function for visual areas V1, V2, V3, V4, V3A, V3B and V7 for Figure and No Figure trials at different eccentricities (see Figure 4.3). Error bars indicate ± one standard error of the mean. A peak can be observed at volume 3. In the center region the response is higher for the Figure than for the No Figure condition, whereas the surround shows the opposite pattern.*



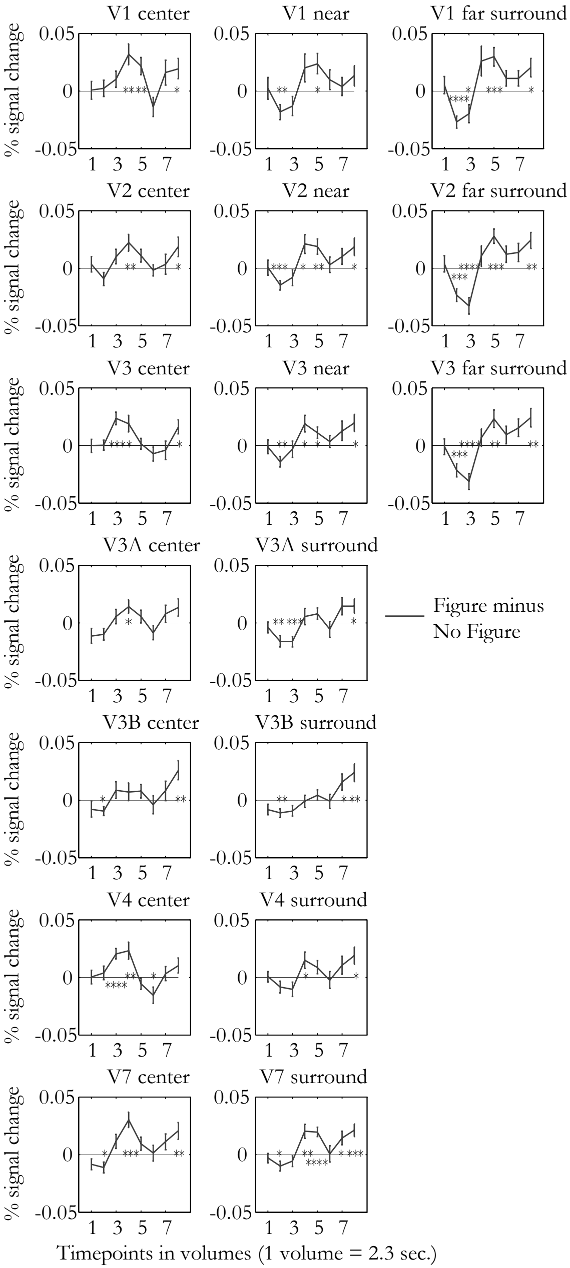
down attention is equal for both stimuli, and thus any influence of anticipatory attention is subtracted out.

Thus, the graphs in Figure 4.5 show the haemodynamic response to processing a figure, barring influence of textures, other stimuli in the stimulus sequence and pre-stimulus attentional effects. Paired t-tests were performed at each time point. Significant differences at a session population level are indicated by asterisks. Areas V1, V2 and V3 were subdivided in three regions of different eccentricities: center, near and far surround (see Figure 4.3a). Almost all areas, starting in V1, show a stronger response to the Figure than to the No Figure stimulus in the center region, as evidenced by the positivity around the time of volume 3 and 4. Similar CM effects have been found in single unit studies of figure ground segregation in macaques. These have shown that neurons in V1 with their receptive field on the figure surface start showing an enhanced response after about ~80-100 ms when compared to the response of these same neurons when they are stimulated by the background (e.g. Lamme, 1995; Zipser, et al., 1996).

The regions corresponding to the near and far surround of the Figure show a largely reversed pattern, with negativity around the time of volume 2 and 3. Thus, in the surround, there is a larger haemodynamic response to the presentation of a homogenous No Figure stimulus than to the presentation of a Figure. An analysis of variance (ANOVA) of these effects at volume 3 for V1, V2 and V3 shows a massive main effect of eccentricity $F(1,2) = 46.14$ ($p<10^{-12}$) revealing enhancement in the center and increasingly strong suppression moving outwards to the surround. The suppression is strongest in the far surround.

After initial suppression in the surround around volume 3, an enhancement can be seen around the time of volume 5. These enhancements are most likely due to a larger undershoot for responses that have an initially larger primary response. As such, we think they should be interpreted as a reversal of the initial effect seen at volume 3, and meaningless when considered in isolation. Inspect Figure 4.4 for improved understanding of how the course of the raw haemodynamic responses influences the subtracted responses in Figure 4.5. Also note that it is tempting to draw conclusions regarding

Figure 4.5 *Shows the Figure minus No Figure difference at different eccentricities (see Figure 4.3) for visual areas V1, V2, V3, V4, V3A, V3B and V7. Note that this is not a raw BOLD signal but a difference signal. For a better understanding of how the raw BOLD signal differs between Figure and No Figure trials please inspect Figure 4.4. Error bars indicate ± one standard error of the mean. Differences were also tested using a paired t-test at each volume. Significant differences are indicated by asterisks:*
**p<.05,*
***p<.01,*
****p<.001,*
*****p<.0001.*



the temporal order of neural processing on the basis of the differences in the timing of peak activation between Figure and No Figure stimuli. For example, as the suppression observed in the far surround of V1, V2 and V3 precedes the timing of the peak response in the center V7 (see Figure 4.5) one may be tempted to conclude that the cortical activity associated with it precedes it as well. However, such putative temporal differences are caused by differences in the shape of the haemodynamic response (see Figure 4.4) and these shape differences are not only influenced by temporal differences in the sequence of neural processing but also by differences in neurovascular effects between cortical sites that are not related to neural activity per se (Logothetis, Pauls, Augath, Trinath, & Oeltermann, 2001). Moreover, abundant reactivation of cortical areas (e.g. Fahrenfort, Scholte, & Lamme, 2008) potentially causes smearing of the peak of the haemodynamic response function over time, further making between-site inferences regarding the temporal sequence of neuronal activation on the basis of the BOLD signal unwarranted.

Large surround suppressive effects in early visual areas as observed in Figure 4.5 have also been reported in the cortical surround of locations receiving voluntary top down attention (Müller & Kleinschmidt, 2004; Slotnick, et al., 2003; A. T. Smith, et al., 2000). Even though anticipatory attention should be the same for both Figure and No Figure stimuli, it is quite conceivable that the presentation of the Figure stimulus captured additional endogenous attentional resources, indirectly causing suppression of the surround through attention networks. In order to preclude this interpretation, we separately modeled Figure stimuli that were seen and Figure stimuli that were not seen. If suppression of the surround is also present in unseen trials (where the Figure does not attract top down attention), this would show that the suppressive effect is in fact stimulus driven and not caused by endogenous top-down attention.

Figure 4.6 shows the Seen/Unseen analysis. Here, the same Figure minus No Figure difference from Figure 4.5 is shown, but now separately for Seen and Unseen trials. Figure and No Figure trials were carefully balanced with respect to orientation and mask duration so that the subtraction procedure detailed above retains its validity.

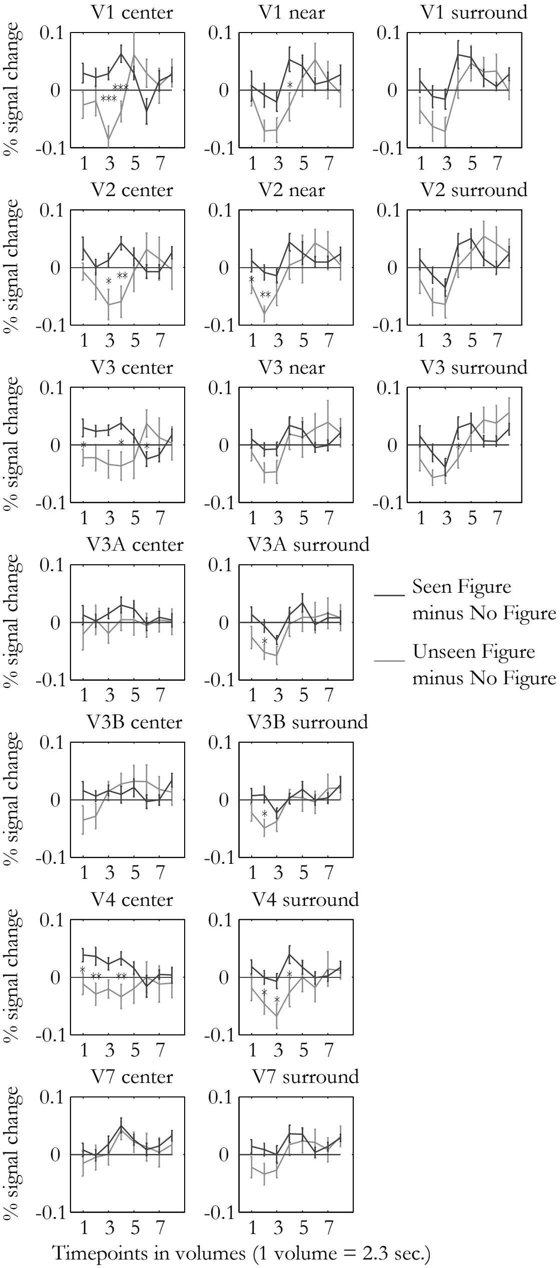


Figure 4.6 *Shows the Figure minus No Figure difference separately for Seen and Unseen Figure trials at different eccentricities (see Figure 4.3) for visual areas V1, V2, V3, V4, V3A, V3B and V7. Note that these are not raw BOLD but difference signals. For a better understanding of how the raw BOLD signal differs between Figure and No Figure trials please inspect Figure 4.4. Error bars indicate ± one standard error of the mean. The pattern for Seen trials is the same as the overall pattern observed in Figure 4.5. Unseen trials, like Seen trials, show suppression in the surround, but the enhancement in the center representation which is present in Seen trials is abolished. Differences between Seen and Unseen trials were ascertained using a paired t-test at each volume. Significant differences are indicated by asterisks:*
$*p<.05$,
$**p<.01$,
$***p<.001$.

Differences between Seen and Unseen trials were determined using paired t-tests, significant differences are indicated with asterisks. Overall, the pattern for Seen trials is the same as in Figure 4.5. Baseline levels at volume 1 seem to be slightly higher for Seen than for Unseen trials, although paired t-tests show that these differences are by and large not significant. These may be due to incidental differences in overall vigilance throughout the experiment, leading to more Figures being detected when vigilance is higher.

The graphs representing the surround in Figure 4.6 (rightmost column) show that even when Figure stimuli were Unseen (classified as No Figure stimuli) they still evoked strong stimulus driven suppression in the surround. The shape of the stimulus induced trough in the surround is roughly the same for Unseen and for Seen Figures, strongly suggesting that the suppressive effect is not caused by endogenous top down attention. The pattern is quite different in the center representations of V1, V2 and V3. Here, the enhancement that is found for Seen Figures is abolished and in fact is reversed when Figures are Unseen. The difference between Seen and Unseen trials is most striking in the center region of V1, resulting in a highly significant difference at volume 3 and 4. As a result of this reversal in the center, Unseen Figure trials seem to cause suppression across the entire visual field in V1, V2 and V3.

Discussion

We observed a highly significant cortical tunnel effect, in which the retinotopic region corresponding to the figure is enhanced, while the surround is strongly suppressed in all early visual areas, including V1. Because cortical responses to masked Figure trials are contrasted with masked No Figure trials (both of which are created from the same set of oriented textures) the effect cannot be attributed to anticipation, to differences in local stimulation or to any effects originating from the presentation of the mask itself. This cortical tunnel effect is reminiscent of effects found in the spatial attention literature, where activity levels in areas on which attention is focused are enhanced, whereas activity levels at locations where attention is not directed are suppressed (Müller & Kleinschmidt, 2004; Slotnick, et al., 2003; A. T. Smith, et al., 2000).

It is therefore conceivable that these effects are partially caused by the allocation of top down endogenous attention to visible Figure stimuli. To rule out this possibility, we separately analyzed seen and unseen trials. The result of this analysis shows that suppressive effects in the surround also occur when a Figure stimulus is completely unseen due to masking. This effectively rules out the influence of endogenous attention as a cause for the extra-classical suppressive effects in the far surround. However, it does not imply that invisible figures are free from attentional influences altogether. Some experiments have shown that invisible stimuli may capture exogenous bottom up attention (McCormick, 1997; Mulckhuyse, Talsma, & Theeuwes, 2007). Nevertheless, such bottom up effects are different from endogenous top down attention in that they are automatic and stimulus driven, as supported by recent evidence in which top down and bottom up attention were dissociated (Buschman & Miller, 2007). Therefore the claim we make that to our knowledge has not been made prior, is that the massive extra-classical suppressive effects we find in the surround are caused by stimulus driven activity originating from sensory areas (which may include exogenous bottom up attention), as they can also be observed for unseen stimuli.

Regardless of whether the term bottom up attention is applicable in this context, it is clear that these extra-classical field suppressive effects are not caused by endogenous top down attention but driven by mechanisms originating from sensory cortex, and it is therefore worthwhile to examine what neurophysiological mechanisms may underpin it. If top down attention is not driving the effects, what is? In directed attention, enhancement has been observed for retinotopic locations that are attended at the expense of non-attended locations in V1 and up, plausibly due to feedback initiated from frontoparietal areas (Müller & Kleinschmidt, 2004; Slotnick, et al., 2003; A. T. Smith, et al., 2000). In automatic texture segregation on the other hand, instead of being triggered ‘at will’, feedback processing is thought to be triggered from within visual cortex, as a consequence of stimulus classification in the feedforward sweep. How this process might lead to figure ground segregation and grouping is worked out in some detail in the incremental grouping theory (Roelfsema, 2006). In this theory, base groupings with increasing complexity are established in the feedforward sweep as a result of lateral interactions between neurons with increasingly large receptive field sizes. This feedforward cascade eventually leads to initial stimulus classification, after which feedback processing serves to incrementally group image regions corresponding to particular perceptual surfaces or objects in retinotopic cortex. In this way, retinotopic space in striate and extrastriate cortex is labeled through feedback as belonging to objects, thereby binding their constituting elements together. Recently it was shown that such feedback mechanisms may serve different roles during different feedback iterations, initially enabling figure-ground segregation, later on serving as a mechanism by which attentional selection operates (Roelfsema, et al., 2007)

Aside from providing a conceptual framework for how figure ground segregation, binding and attentional selection may be implemented in visual cortex, incremental grouping theory accounts for a number of neurophysiological findings, amongst which that of extra-classical receptive field effects, such as contextual modulation in V1 (Lamme, 1995; Zipser, et al., 1996). We suggest here that feedback processes that enable incremental grouping not only serve to bind the figure surface together, but that areas that do not correspond to an object or surface, may be actively suppressed to improve figural assignment and discriminability, in effect binding the background together. The combined role of enhancement and suppression could be akin to the process of contrast enhancement for objects in retinotopic cortex. A biologically inspired neural network model of texture segregation by Roelfsema (2002), explicitly models inhibitory feedback connections in order to reliably counter faulty assignment of background elements to figural regions in early visual areas. As a result, in this network, regions corresponding to the background of an image containing a figure are suppressed due to feedback after about 200 ms. Although speculative, our data nicely fits with such a model.

This suppression is different from endogenous attentional suppression in that it is an automatic, and not a cognitively driven process, as evidenced by the fact that it occurs for unseen stimuli. Rather, it may be driven by feedback following automatic stimulus classification in the feedforward sweep. The fact that unseen trials in this experiment seem to show suppression for *all* retinotopic locations in early visual areas ties in with this idea. There is strong evidence that figure-ground detection in visual cortex as a result of initial feedforward processing can occur completely unimpaired, even when texture stimuli are strongly masked and figure presence is unreportable by the subject (Fahrenfort, et al., 2007). We conjecture that even though unseen figures may be detected by high visual areas in the feedforward sweep (e.g. V7 center in Figure 4.6), binding this figure representation to the corresponding retinotopic surface in low visual areas is unsuccessful in the case of unseen trials due to masking. This then results in effective suppression of the entire visual field as not belonging to the figure, which would explain why the entire early visual cortex seems to be suppressed as a result of unseen figures. In contrast, seen trials are marked by strong enhancement in the center of early visual areas, which is compatible with the notion of figural regions being successfully labeled as such through feedback, eventually leading to stimulus detection by the subject.

Aside from theoretical considerations when inferring neurophysiological mechanisms that may underpin these effects, the distribution of responses across retinotopic space in V1 found in this study is also informative. We show large scale differences between V1 processing of Figure and No Figure trials, the distribution of which extends far beyond the V1 cRF, while local physical stimulation and anticipatory attentional levels are roughly the same. Estimates of V1 cRF size in macaques are less than 1° even at sizeable eccentricities (A. T. Smith, Singh, Williams, & Greenlee, 2001), and estimates of early visual cRF size in humans using subdural electrodes also point to (much) smaller than 1° sizes (Yoshor, Bosking, Ghose, & Maunsell, 2007). The effects observed in this study take place at considerably larger distances, suggesting the involvement of feedback mechanisms originating from higher visual areas with larger receptive fields. Studies explicitly investigating the role of feedback connections in the integration of global organization to local responses have also shown the importance of feedback in this respect (Angelucci, et al., 2002; Hupe, et al., 1998; Lamme, Super, et al., 1998).

Nevertheless, one should be cautious when inferring feedback mechanisms from fMRI data. For example, some have noted that suppressive effects as observed in this study might be explained in terms of “blood stealing”, haemodynamic changes that have no neural correlate (A. T. Smith, Williams, & Singh, 2004). However, in the present study, suppression of the BOLD signal for Unseen trials happens across the visual field for early retinotopic areas, so that it is unclear how blood stealing could have caused it. Recently, V1 BOLD has also been implicated in anticipatory arterial pumping (Sirotin & Das, 2009), but as the amount of anticipation was closely matched between conditions in this study, such mechanisms cannot account for our findings.

Conclusions:

1. Suppression of the surround occurs in V1 and up, and is not exclusively a top down attentional phenomenon but may result from sensory-driven processing within visual cortex due to the global organization of the stimulus, possibly related to exogenous bottom up attention and/or figure-ground segregation.
2. Enhanced activity at retinotopic locations in V1 corresponding to the surface of a figure is indicative of whether that figure is seen or not.

Chapter 5.

Object classification in the absence of visual awareness and figure-ground segregation

Abstract

It is well known that neurons in the temporal lobe classify objects, such as faces, and it is generally assumed that the activity of such neurons is necessary for conscious awareness of these objects. However, object categorization may also occur unconsciously, as has been shown by the selective activation of object selective neurons by masked objects. So what distinguishes conscious from unconscious object recognition? We constructed schematic images containing objects such as faces and houses while keeping local retinal stimulation between conditions identical. Using a dichoptic fusion paradigm, we manipulated stimulus visibility such that objects were either visible or not visible. Confirming earlier results, we found that both consciously perceived and non-perceived objects result in category specific BOLD activation, even if they are task irrelevant and non-attended. Critically however, we show that objects that are consciously seen show a distinct neural signature of figure-ground segregation in early and midlevel visual areas, which is completely absent when objects are not seen. Although counterintuitive, this implies that consciousness is more intimately related to processes of figure-ground segregation and perceptual organization than to object categorization. We propose that figure-ground segregation is a prerequisite for visual awareness, and that both phenomena share part of their neural correlate, which is recurrent processing within visual cortex.

Fahrenfort, J. J., Heinen, K., van Gaal, S., Scholte, H. S. & Lamme, V. A. F. (Submitted) Object classification in the absence of visual awareness and figure-ground segregation.

Introduction

Some regions in the visual cortex are particularly sensitive to processing specific object categories. Amongst the first experiments showing this were a study identifying an area in the fusiform gyrus engaged in processing faces (Kanwisher, et al., 1997) and a study showing a region in the parahippocampus engaged in processing houses or scenes containing a spatial layout (Epstein & Kanwisher, 1998), both of which extended the observation that a large region termed the lateral occipital complex (LOC) responds more strongly to objects than to scrambled objects or other visual patterns (Malach, et al., 1995).

It has also been shown that the brain is able to perform such object categorization even when objects cannot be consciously perceived (Kouider, et al., 2008; G. Kovacs, et al., 1995; Marois, et al., 2004; Moutoussis & Zeki, 2002). This implies that object categorization is not the same as object perception, although it is still unclear what cortical mechanisms distinguish the two. So what does it take to go from object classification to actually seeing something like a face? A number of studies suggest that object categorization is highly distributed (e.g. Haxby, et al., 2001; Schwarzlose, Swisher, Dang, & Kanwisher, 2008). However, the objects in these studies can be readily perceived and are consciously categorized. As many studies have implicated recurrent processing in conscious visual perception (e.g. Boehler, Schoenfeld, Heinze, & Hopf, 2008; Haynes, et al., 2005; Silvanto, et al., 2005; Super, et al., 2001), it may be that distributed category representations emerge only as a result of conscious visual perception. There is evidence that object categorization can also proceed in a highly automated fashion, based on feedforward processes alone (G. Kovacs, et al., 1995; Thorpe, et al., 1996; VanRullen & Koch, 2003). Although speculative, we hypothesize from these results that conscious object perception may be tied to distributed representations resulting from recurrent processing, whereas unconscious object categorization is tied to fleeting focal representations resulting from feedforward processing.

This also has bearing on how object categorization is related to processes of figure-ground segregation. Traditionally, the logical sequence of events was thought to be that objects were first segregated from their background before they could be classified (e.g. Rubin, 1958). As figure-ground segregation has repeatedly been associated with recurrent processing (e.g. Lamme, 1995) and in light of the hypothesis that object categorization is an unconscious feedforward process, we postulate the reverse, that object categorization precedes figure-ground segregation and as such may proceed independently from it. In our proposed scheme of events, both unseen and seen objects are initially categorized by category selective neurons, whereas only objects that are subsequently segregated from their background are consciously perceived. Although counterintuitive, this would imply that consciousness is more intimately related to processes of figure-ground segregation and perceptual organization than to object categorization.

To test these hypotheses we used a dichoptic fusion paradigm in which stimuli were either visible or invisible (see Figure 5.1a). Using classical contrasts specific to object categorization and figure-ground segregation we were able to directly test the relationship between object categorization, figure-ground segregation and visual awareness.

Methods

Participants

Eighteen subjects participated in the experiment for a monetary reward. All subjects provided written informed consent. All procedures were approved by the ethical committee of the University of Amsterdam. Two subjects were not included in the analyses: one subject's data was lost due to human error, the other subject showed excessive movement during scanning. Sixteen subjects (3 male, 2 left handed) with normal vision and an average age of 21.1 years (± 3.4) were included in data analysis.

Stimuli

Stimuli were presented at 800 × 600 resolution with a projected size measuring 16.9° × 12.7° visual angle. They were composed of a 28 by

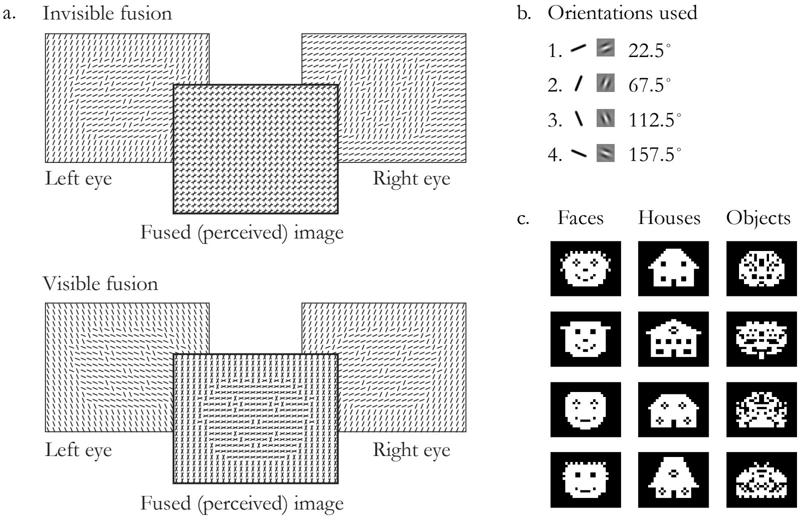


Figure 5.1 *(a) Example of invisible and visible dichoptic fusion using schematic line elements. In either eye, objects were created in the image by using a different orientation for the Gabor elements in the foreground (depicting a face in this example) and the background. There was always a 45° orientation difference between foreground and background. For invisible fusion (top panel), objects were created by using the same orientation between elements belonging to the foreground of one eye and elements in the background of the other eye. This way, images in either eye contain an object whereas the percept of the dichoptically fused image consists of a homogenous screen of crossed Gabor elements. For visible fusion (bottom panel) objects were created by using a different orientation for the elements in the foreground of one eye and elements in the background of the other eye, resulting in differently crossed Gabor elements for foreground and background. (b) Schematic lines depicting Gabor elements, as well as the actual Gabor elements used in the experiment. Four orientations were used: 22.5°, 67.5°, 112.5° and 157.5°. (c) Schematic depiction of all four instances of three of the four stimulus categories: faces, houses and nonsense-objects (homogenous screens are not shown). Black represents one orientation, white the other orientation.*

22 matrix of oriented 2-dimensional Gabor functions, oriented elements with luminance features that roughly model the receptive field structure of V1 simple cells (period ≈ 0.58°, $\sigma \approx 0.10°$, schematically represented by lines in the left and right eye of Figure 5.1a). By using a different orientation for the Gabor elements representing the figure (the foreground) and for elements representing the background, objects could be constructed in the images. Four

orientations were used: 22.5°, 67.5°, 112.5° and 157.5° (see schematic lines and actual Gabor elements in Figure 5.1b). An image containing an object would always be composed of two orientations with a 45° difference between them (for examples see left and right eye of Figure 5.1a). Four stimulus categories were created in this way: faces, houses, nonsense objects and homogenous screens using all possible orientations (see schematic representation in Figure 5.1c, white is one orientation, black the other orientation; only stimulus categories containing an object are shown, homogenous screens are omitted in this schematic). There were four instances of each of the object categories (see Figure 5.1c). All objects consisted of the same number of oriented elements.

Visible versus invisible dichoptic fusion

Stimuli containing different orientations of Gabor elements were presented to the left and the right eye. When different images are presented to the left and right eye for a short duration, they are fused to a single percept, a phenomenon known as dichoptic fusion (see Kolb & Braun, 1995; Moutoussis & Zeki, 2002; Zipser, et al., 1996). The fusion phenomenon was exploited to create conditions under which objects in the images were either visible or not visible, while keeping physical stimulus characteristics locally identical.

Dichoptically invisible objects were created by using the same orientation between the elements belonging to the object (the foreground) of one eye and the elements in the background of the other eye, while maintaining a 45° difference between foreground and background within each eye. This way, images in either eye contain an object whereas the percept of the dichoptically fused image consists of a homogenous screen of crossed Gabor elements (see Figure 5.1a top). Dichoptically visible objects were created in the same way, but using a different orientation for the elements in the foreground of one eye and the elements in the background of the other eye, resulting in differently crossed elements for foreground and background (see Figure 5.1a bottom).

This was done for all objects and homogenous screens, thus creating 8 conditions:

dichoptically fused visible faces, houses, nonsense objects and homogenous screens, and dichoptically fused invisible faces, house, nonsense objects and homogenous screens. For brevity, dichoptically fused visible and dichoptically fused invisible conditions are henceforth referred to simply as visible and invisible conditions.

Stimulus sequence

The sequence of events in a trial can be found in Figure 5.2a. A red fixation dot was present in the middle of the screen throughout the experiment. Each trial started with 167 ms in which a grey screen was shown. After this, the stimulus sequence was presented four times in succession. Each sequence began with a 50 ms mask, followed by an isoluminant grey screen for 67 ms, the stimulus screen for 83 ms and ending with another isoluminant grey screen for 50 ms. The (repeated) short stimulus presentation of 83 ms was used to aid dichoptic fusion. The masking screens (acting both as forward and backward masks) consisted of a 28 by 22 sized matrix of Gabor elements of which half were randomly rotated to an orientation of 0° and the other half were rotated to 90°. If the orientation of an element was at 0° in the left eye it would be at 90° in the right eye and vice versa, resulting in a homogenously fused percept consisting of plus-like Gabor elements.

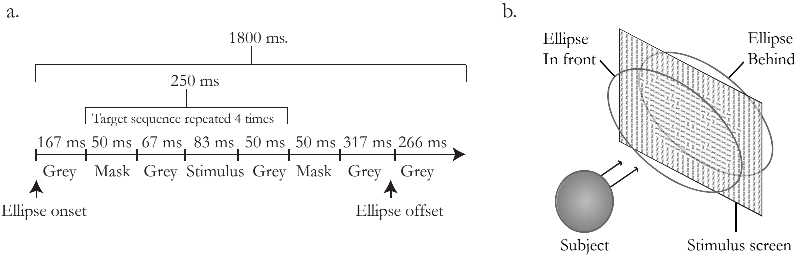


Figure 5.2 *(a) Schematic depiction of the timeline of a trial. (b) 3D representation of a subject performing the ellipse "in front" / "behind" task (both "in front" and "behind" are shown for illustrative purposes). An illusion of dimensionality was created of an ellipse hovering "in front" or "behind" the stimulus screen by offsetting ellipse location horizontally from center in different directions for the left and the right eye.*

On each trial, the sequence of four stimulus presentations always contained one stimulus category. Half of the trials contained a sequence of four of the same instances of its stimulus category (Same trials); the other half of the trials contained four different instances of its stimulus category (Different trials). These were originally intended to look at adaptation effects, but explorative analyses showed that statistical power proved insufficient for useful analysis of such adaptation effects. Since the Same-Different manipulation had no bearing on the rest of the experiment, Same-Different trials were pooled in all subsequent analyses and are mentioned no further.

In addition, during the first 1533 ms of a trial, an elliptic circle of size 15.45° × 11.76° visual angle overlaying the stimulus sequence was presented to each eye (large enough to encompass faces/houses/objects in the stimulus sequence). By displacing each ellipse 7 pixels from the fixation center in opposite direction for each eye, a retinal disparity was created. This disparity created an illusion of dimensionality, in which the perceived ellipse either seemed to hover in front or behind the stimulus screen, but only when properly focusing with both eyes. A three dimensional impression of what this looks like can be found in Figure 5.2b. On half of the trials, the disparity was such that the circle seemed to hover in front of the stimulus screen, on the other half of the trials the circle appeared to hover behind the stimulus screen, although the disparity was small enough to make this difficult to see when not paying attention, and impossible to see when viewing the stimuli monocularly. During the experiment, subjects were instructed to indicate whether the circle was hovering ‘in front’ or ‘behind’ the stimulus sequence. This was done to induce proper convergence for stereoscopic viewing, a necessary condition for dichoptic fusion to occur. As soon as target stimulus sequence began, subjects were able to give a response on the ellipse task, until the next trial would begin. Trials were presented in immediate succession, although sometimes empty dummy trials lasting 1.8 or 3.6 seconds were presented in between (see fMRI acquisition).

The orientations of the Gabor elements in the stimuli and all other elements in the stimulus sequence were balanced out in such a way that local physical stimulation was on average the same for faces, houses, nonsense-objects and homogenous screens within the visible and invisible conditions. This allowed us to make contrasts between conditions that are sensitive to higher order differences in perceptual organization, while largely shutting out the influence of local physical stimulation on cortical processing (such as the influence of the Gabor elements themselves on orientation selective neurons in early visual areas, Hubel & Wiesel, 1962, 1968; Lamme, 1995; Zipser, et al., 1996).

Tasks

On each trial during the experiment, subjects had to indicate whether an ellipse was hovering ‘in front’ or ‘behind’ the stimulus sequence using a two-button response panel in their right hand (see under Stimulus Sequence above for a description of ellipse presentation). The reason for using this task was threefold: [1] using this task forced subjects to view the stimuli stereoscopically, which aids dichoptic fusion [2] using an attentionally demanding task that does not focus on the stimuli themselves, makes it less likely that differences between the visible and invisible conditions are caused by differences in attention or task relevance and [3] it is less likely that any correlates of visibility found in this study are related to the act of report about the stimulus.

To determine whether subjects were successfully fusing stimuli in the experiment, a control task was performed after scanning (but still inside the scanner), in which subjects had to make two consecutive responses on each trial: [1] indicate whether the circle was hovering ‘behind’ or ‘in front’ of the stimulus sequence, and subsequently [2] indicate the object category in the stimulus sequence (face, house, nonsense object or homogenous screen). If subjects were successfully fusing the objects, one would predict that invisible faces, houses and nonsense objects would be classified as homogenous screens, whereas visible stimuli would be correctly categorized. Responses were made on a two-button response panel in their right hand (in front / behind) and a four-button response panel in their left hand (faces, houses, nonsense objects and homogenous screens). This control task consisted of 32 trials per condition in random order, a total of 256 trials.

Apparatus for stereo presentation

To be able to present different images to the left and the right eye, all visual stimulation was presented using two stacked beamers projecting at exactly the same location on a projector screen in the scanner room. Using polarizing filters, the upper beamer projected light of opposite polarity compared to the light of the bottom beamer. Subjects viewed the screen via a mirror through polarizing glasses. The glass of the left eye only allowed light of the upper beamer to pass through, the glass of the right eye only allowed light of the lower beamer to pass through.

fMRI acquisition, preprocessing and analysis

Data was collected on a Philips 3T Intera scanner. All subjects took part in one scanning session, which always started with the acquisition of a high resolution three dimensional anatomical image of the head using a T1 image sequence lasting about 6 minutes. Subsequently, three runs of functional data were collected, except for three subjects from whom four runs were acquired. A functional run started and ended with 16 seconds rest and lasted approximately 12 minutes. The functional recordings were acquired using a T*-weighted sequence (TR 2301 ms, TE 28 ms, 35 slices, slice thickness 3 mm; slice gap 0.3 mm, FOV 220 × 220 mm). Each run contained 256 target trials, in which each of the eight stimulus categories were presented 32 times in random order. In addition, 100 dummy trials containing only an isoluminant grey screen intended to improve deconvolution and counteract BOLD saturation in visual areas were randomly interspersed between these trials. Of these dummy trials, 20 lasted 3.6 seconds, and 80 lasted 1.8 seconds.

Preprocessing and data analysis was performed using FSL (FMRIB's Software Library, http://www.fmrib.ox.ac.uk/fsl) and Matlab (The Mathworks Inc., Natick, MA, USA). Using FEAT (fMRI Expert Analysis Tool, v5.92) fMRI images were motion corrected using MCFLIRT, slice-time aligned, brain areas were extracted using BET, the data were spatially smoothed using a Gaussian kernel of 5 mm and high-pass temporally filtered using a Gaussian envelope (σ = 35 s).

All functional data were aligned to the structural image of the subject. Using the structural image, the data of each subject was transformed to the standard space of the Montreal Neurological Institute (MNI) using FLIRT. Subsequently all conditions and relevant contrasts were specified in a General Linear Model (GLM) and all runs were analyzed using this GLM. Runs were pooled on a per subject basis using a fixed effects model. Subsequently, a mixed effects group analysis was done (FMRIB's FLAME stage 1) in which relevant lower level contrasts were combined. These were used to help determining the regions of interest. From these regions of interest, parameter estimates from the GLM were transformed into estimates representing percent signal change and exported for each subject and each condition to perform final analyses.

Regions of interest (ROI's)

We looked at four regions of interest: [1] the FFA (Fusiform Face Area), [2] a region in the subiculum of the hippocampus that responded to the contrast normally used to locate the PPA (Parahippocampal Place Area), [3] the LOC (Lateral Occipital Cortex), which was subdivided into a superior and an inferior part and [4] V1 and V2. The exact regions for which we compared visible and invisible stimuli were chosen because of their theoretical interest, and defined on the basis of a combination of anatomical and functional characteristics. All functional contrasts used in defining the ROI's (FFA, PPA, LOC and V1/V2) were created by pooling visible and invisible stimuli, so as not to bias the results in any direction towards the processing of visible or invisible stimuli.

The left and right FFA ROI's were defined using a classical contrast used to locate the FFA (Kanwisher, et al., 1997): voxels responding more strongly to Faces than to houses and other objects. We included voxels that exceeded an uncorrected threshold of $Z > 2.3$ within the area of the Temporal Occipital Fusiform Cortex as defined by the Harvard-Oxford Cortical Structural Atlas in MNI space. This resulted in two well-defined clusters corresponding to the FFA as described in the literature, with a larger cluster for the right hemisphere than for the left hemisphere (see Kanwisher, et al., 1997). Voxels in the ROI's were weighted more strongly as they responded more strongly to faces than to houses and nonsense objects, as this is the contrast that classically defines the FFA.

Originally, we had intended to look at activity in the PPA (Parahippocampal Place Area), which has been implicated in perceiving the local visual environment, such as represented by rooms, scenes and houses (Epstein & Kanwisher, 1998). However, the houses > faces & objects contrast which is typically used to identify the PPA, did not result in convincing clusters representing typical PPA activation in the parahippocampus. However, a symmetrical bilateral cluster of voxels did show up slightly more dorsal, in the subiculum of the hippocampus itself, as structure that has been implicated in, amongst other things, the encoding of novel complex pictures and other information (e.g. Clark, Broadbent, Zola, & Squire, 2002; Stern, et al., 1996). The reason for the classical PPA contrast not showing up in the parahippocampus may well have something to do with the fact that the PPA is specifically sensitive to representations containing the layout of the local environment, such as scenes and rooms and much less to objects without a three-dimensional spatial context (even if they contain a spatial representation, Epstein & Kanwisher, 1998). Our iconic reproduction of a house using Gabor elements may simply have not had enough flavor of the local spatial environment. However, we did include the slightly more dorsal bilateral cluster in the hippocampus. It was defined by voxels exceeding an uncorrected threshold of $Z > 2$ in the subiculum of the hippocampus as defined by the Juelich Histological Atlas in MNI space. Voxels were weighted more heavily as they responded more strongly to houses than to faces and nonsense objects.

The LOC regions of interest (inferior and superior LOC separately) were chosen because of the implication of the LOC in human object recognition (Grill-Spector, Kourtzi, & Kanwisher, 2001). The regions were defined as those clusters of voxels responding more to stimuli containing a figure-ground relationship (faces, houses and nonsense objects) than to homogenous screens (thresholded by a $Z > 2.3$ and a corrected cluster significance threshold of $p = .05$) lying either in the inferior LOC or superior LOC as defined by the probabilistic Harvard-Oxford Cortical Structural Atlas in MNI space. Voxels in the ROI's were weighted more strongly as they had a bigger probability of belonging to either the superior or to the inferior LOC as defined by the atlas.

Finally, V1 and V2, were chosen because of their traditional implication in low level vision and the more recent implication of V1 in figure-ground segregation and visual awareness through recurrent processing (Lamme, 1995; Lamme, et al., 2000). The ROI's were defined as those clusters of voxels responding more strongly to stimuli containing a figure-ground relationship (faces, houses and nonsense objects) than to homogenous screens (thresholded by a $Z > 2.3$ and a corrected cluster significance threshold of $p = .05$) lying either in V1 or in V2, as defined by the probabilistic Juelich Histological Atlas in MNI space. Again, voxels in these ROI's were weighted according to the probability that they belonged to either V1 or V2.

Results

Behavior

In a behavioral control task performed directly after scanning (but still inside the scanner), subjects had to make two responses: [1] the ellipse localization response which they also performed during scanning (see Methods), and [2] a stimulus categorization response. The stimulus categorization task was intended as confirmation that subjects were fusing well enough to render stimuli invisible in the invisible condition. The average response frequency for each of the stimulus categories can be found in Figure 5.3, separately for invisible (left panel) and for visible trials (right panel). As can be seen in Figure 5.3 (left panel), almost all of the invisible trials were categorized as homogenous screens, thus indicating that fusion and our visible-invisible manipulation was successful. One sample *t*-tests against zero confirmed that none of the invisible categories (except homogenous screens) were identified above chance level. Figure 5.3 (right panel) shows that all of the categories in the visible condition were identified well above chance level.

During scanning, subjects only had to perform the ellipse localization ('in front'/'behind') task. Despite task difficulty, the average proportion correct was .89 (SEM .02), indicating that subjects were viewing the stimuli stereoscopically. The percentage correct did not differ between 'in front' or 'behind' trials [$F_{1,15}$ = .35, p = .57], or between visible and invisible trials [$F_{1,15}$ = .54, p = .47], as determined by a repeated measures test. On the control task, subjects scored .91 (SEM .03) on the ellipse localization task. Subjects did not score differently on ellipse localization during scanning and ellipse localization during the control task [t_{15} = 1.28, p = .22], confirming that stereoscopic viewing conditions between the control task and the experiment itself were highly comparable.

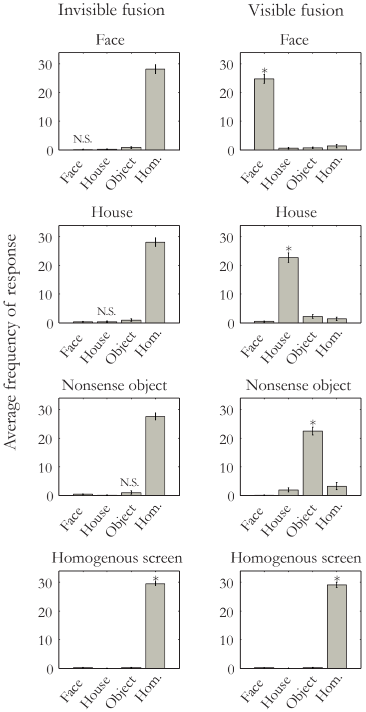
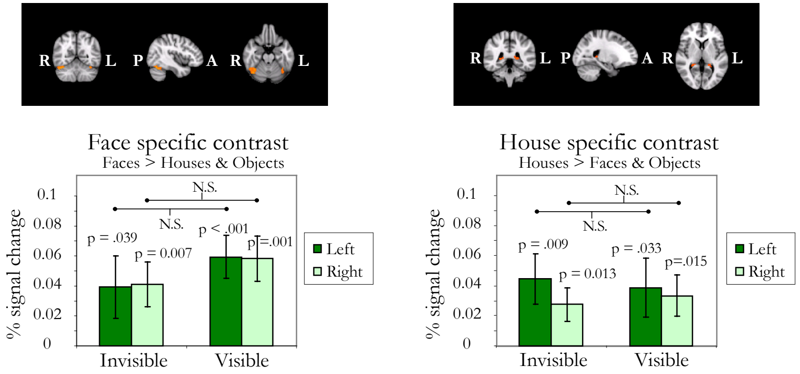


Figure 5.3 *Behavioral response frequency data for (a) invisible conditions and (b) visible conditions. Each graph represents a stimulus condition. Mean response frequency (±1 s.e.m.) is given for all possible responses. For each condition a one sample t-test against 0 was performed to determine whether that category was identified above chance level [N.S. = $p > .05$; * = $p < 10^{-15}$].*

Object specific activation for face and house stimuli

We aimed to reveal whether face and house specific activity could be found for visible and invisible stimuli. Parameter estimates for faces were tested against parameter estimates for houses & objects using one-tailed paired t-tests, separately for visible and invisible stimuli, and separately for the left and right FFA (see Figure 5.4a). Both visible and invisible faces caused significantly more activation in both the left and right FFA when compared to houses & objects, thus showing face specific activation for both visible and invisible faces. Moreover, as can bee seen in the bottom panel of Figure 5.4a, there was no significant difference between the visible and invisible FFA contrast for either the left or the right FFA (see figure and legend for *t*- and *p*-values).



a. FFA ROI's in fusiform cortex

b. ROI's in hippocampal area

Figure 5.4 *(a) face and (b) house specific activation for visible and invisible stimuli.*
(a, top panel) Left and right ROI's showing face selective voxels in orange. ROI's were determined using the faces > houses & objects contrast (see methods). Voxels with higher Z-values have lighter colors. (a, bottom panel) Mean % signal change (±1 s.e.m.) for the faces > houses & objects contrast, separately for visible and invisible stimuli in the left and right FFA. Paired t-tests confirmed that both invisible [left FFA: $t_{15} = 2.27, p = .039$, right FFA: $t_{15} = 2.78, p = .007$] and visible stimuli [left FFA: $t_{15} = 4.50, p < .001$, right FFA: $t_{15} = 4.15, p = .001$] showed above chance % signal change. Moreover, paired t-tests between visible and invisible conditions showed that visible stimuli did not show significantly more activation than invisible stimuli in these areas [left FFA: $t_{15} = 1.29, p = .215$, right FFA: $t_{15} = 1.29, p = .218$].
(b, top panel) Left and right ROI's showing house selective voxels in orange. ROI's were determined with the houses > faces & objects contrast (see methods). Voxels with higher Z-values have lighter colors. (b, bottom panel) Mean % signal change (±1 s.e.m.) for the houses > faces & objects contrast, separately for visible and invisible stimuli in the left and right ROI. Paired t-tests confirmed that both invisible [left ROI: $t_{15} = 2.99, p = .009$, right ROI: $t_{15} = 2.83, p = .013$] and visible stimuli [left ROI: $t_{15} = 2.35, p < .033$, right ROI: $t_{15} = 2.76, p = .015$] showed above chance % signal change. Moreover, paired t-tests between visible and invisible conditions showed that visible stimuli did not show significantly more activation than invisible stimuli [left ROI: $t_{15} = .83, p = .415$, right ROI: $t_{15} = .91, p = .38$].

One-tailed paired *t*-tests of houses against faces & other objects revealed house specific activation, both for visible and invisible houses, and both for the left and the right hemispheric ROI (see Figure 5.4b). There was no significant difference between the visible and invisible contrasts for the ROI in the left or for the ROI in the right hemisphere (see bottom panel of figure 4b and legend for *t*- and *p*-values). In summary, it appeared that both houses and faces evoked category specific activation of high level areas, regardless of their visibility, confirming earlier results using a similar dichoptic fusion paradigm (Moutoussis & Zeki, 2002).

Figure-ground activity in LOC and early visual areas for visible, but not for invisible stimuli

Subsequently, we aimed to reveal activity related to figure-ground segregation, by contrasting stimuli containing a surface with those that do not contain a surface (houses, faces & other objects > homogenous screens). From the voxels in this contrast we created ROI's using anatomical criteria (see Methods section), resulting in four ROI's: [1] inferior LOC, [2] superior LOC, [3] V1 and [4] V2 (see Figure 5.5a and b, top panels). Note that all figure-ground activity was covered by these four occipital ROI's, there was no additional activity in more frontal or other regions of the cortex.

Activity caused by faces, houses and objects was compared to activity caused by homogenous screens using paired t-tests, separately for visible and invisible stimuli. This resulted in significant activation for visible stimuli in inferior LOC, superior LOC (Figure 5.5a, bottom panel) and V1 and V2 (Figure 5.5b, bottom panel). Invisible stimuli however, showed no significant activation whatsoever for this contrast in any of these areas. Moreover, when comparing visible and invisible contrasts directly, visible contrasts show significantly more activity than invisible contrasts in all of these areas (see bottom panels of Figure 5.5a and 5.5b, legend for *t*- and *p*-values). The same was true for all object categories separately. When doing separate *t*-tests, invisible faces, houses and objects did not show more activity than homogenous screens in these areas, whereas visible faces, houses and objects did.

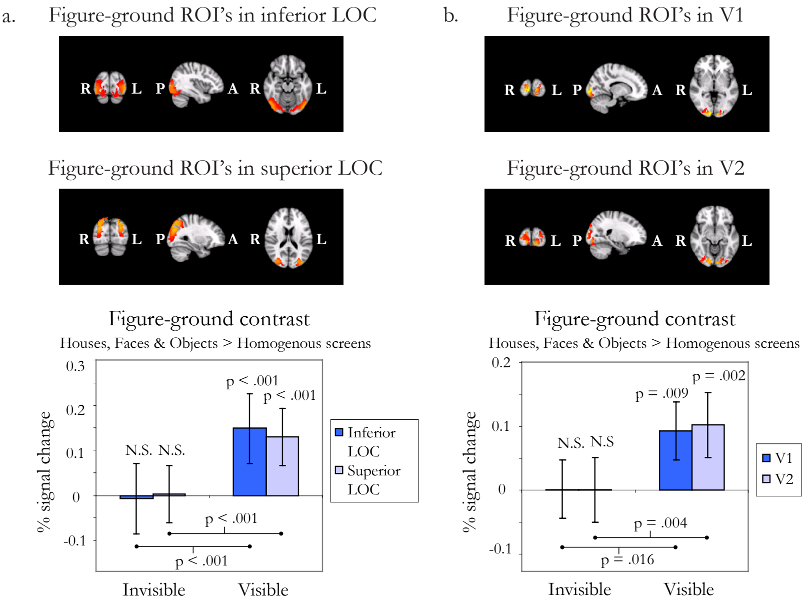


Figure 5.5 *Figure-ground activation in (a) inferior and superior LOC and in (b) V1 and V2. (a, top panel) Inferior and superior LOC ROI's in orange containing those voxels that are selective for stimuli containing a figure-ground relationship (see methods). Voxels with a higher probability of belonging to inferior or superior LOC are indicated by lighter colors. (a, bottom panel) Mean % signal change (±1 s.e.m.) for the faces, houses & objects > homogenous screens contrast, separately for visible and invisible stimuli in inferior and superior LOC. Paired t-tests confirmed that invisible did not show above chance % signal change [LOC inferior:* $t_{15} = .96, p = .352$*, LOC superior:* $t_{15} = .77, p = .454$*], whereas visible stimuli did [LOC inferior:* $t_{15} = 6.51, p < .001$*, LOC superior:* $t_{15} = 6.19, p < .001$*]. Moreover, paired t-tests between visible and invisible conditions showed that visible stimuli showed significantly more activation than invisible stimuli [LOC inferior:* $t_{15} = 5.41, p < .001$*, LOC superior:* $t_{15} = 4.42, p < .001$*].*
(b, top panel) V1 and V2 ROI's in orange containing those voxels that are selective for stimuli containing a figure-ground relationship (see methods). Voxels with a higher probability of belonging to V1 or V2 respectively are shown in lighter colors. (b, bottom panel) Mean % signal change (±1 s.e.m.) for the faces, houses & objects > homogenous screens contrast, separately for visible and invisible stimuli in V1 and V2. Paired t-tests confirmed that invisible did not show above chance % signal change [V1: $t_{15} = .73, p = .478$*, V2:* $t_{15} = .71, p = .490$*], whereas visible stimuli did [V1:* $t_{15} = 3.03, p = .009$*, V2:* $t_{15} = 3.74, p = .002$*]. Moreover, paired t-tests between visible and invisible conditions showed that visible stimuli showed significantly more activation than invisible stimuli [V1:* $t_{15} = 2.72, p = .016$*, V2:* $t_{15} = 3.44, p < .004$*].*

This was true for all areas, except for the visible faces versus homogenous t-test in V1 (p = .075), possibly due to a lack of statistical power (see Table 5.1 for *t*- and *p*-values for all *t*-tests in all areas). Inferior LOC, superior LOC, V1 and V2 all seem to be engaged in figure-ground segregation, but only when the segregated object is consciously perceived.

Table 5.1 *Figure ground contrast for individual stimulus categories. t- and p-values for the individual t-tests between the object categories and homogenous screens, separately for visible and invisible conditions in inferior and superior LOC as well as V1 and V2. All visible categories except faces in V1 show significant p-values as indicated by an asterisk (*), whereas none of the invisible categories do so.*

		LOC inferior		LOC superior		V1		V2	
		t-value	p-value	t-value	p-value	t-value	p-value	t-value	p-value
invisible	faces > homogenous	0.95	0.356	1.56	0.140	1.07	0.300	1.06	0.305
	houses > homogenous	1.25	0.230	1.03	0.319	0.76	0.457	0.85	0.410
	objects > homogenous	1.13	0.276	0.74	0.472	1.06	0.307	1.02	0.325
visible	faces > homogenous	7.74	< 0.001 *	6.06	< 0.001 *	1.91	0.075	3.14	0.007 *
	houses > homogenous	5.83	< 0.001 *	5.51	< 0.001 *	2.63	0.019 *	3.17	0.006 *
	objects > homogenous	4.23	0.001 *	4.47	< 0.001 *	3.05	0.008 *	3.26	0.005 *

Discussion

As in previous studies (Kouider, et al., 2008; Marois, et al., 2004; Moutoussis & Zeki, 2002), this study shows that invisible stimuli can cause stimulus specific activity in object selective parts of the human brain. Moreover, we are able to produce this effect even when the stimuli are task-irrelevant and non-attended, thus suggesting that such categorizing can be established automatically. Additionally, we show that under these conditions there is no significant difference in the level of stimulus specific activity between visible and invisible stimuli, making accounts relying on activity thresholds for visual awareness (e.g. Moutoussis & Zeki, 2002) somewhat unlikely. Apparently, as noted by these and other authors (Beck, Rees, Frith, & Lavie, 2001; Dehaene, et al., 2001), stimulus specific activation (or stimulus categorization by the brain) is insufficient to generate visual awareness. If stimulus specific activity is insufficient to generate visual awareness, what is?

We identified robust non-stimulus specific differences between visible and invisible stimuli in areas of the visual cortex that have previously been implicated in object recognition (LOC, also see

Grill-Spector, et al., 2001) and low level vision (V1 and V2). In these areas, visible stimuli containing a figure-ground relationship generated much more activity than homogenous screens, whereas invisible stimuli containing such a relationship showed no increased levels of activation in this contrast, despite the fact that both visible and invisible stimuli were able to generate stimulus specific activity in higher and/or more specialized areas.

It is especially surprising that this pattern of results was also found in the earliest visual area V1, an area with small receptive fields that has traditionally been found to respond only to simple stimulus characteristics such as line orientation (Hubel & Wiesel, 1962, 1968). Estimates of V1 receptive field size in macaques are ~0.25° in the fovea and less than 1° even at sizeable eccentricities (A. T. Smith, et al., 2001). Estimates of early visual receptive field size in humans using subdural electrodes also point to (much) smaller than 1° sizes (Yoshor, et al., 2007). As Gabor elements in this study subtended a visual angle comparable to the receptive field size of V1 neurons (0.58°) and conditions were balanced out at this level (both monocularly and binocularly, and both within the visible conditions and within the invisible conditions) one would not expect selective differences between conditions to arise here, as was confirmed for invisible stimuli but not for visible stimuli. What then brings about these differences between conditions for visible stimuli in V1? It seems unlikely that these are caused by long-range horizontal intracortical connections. On a monocular level, both visible and invisible stimuli contain figure-ground relationships. If long range horizontal connections would be responsible for the figure-ground signals, it is unclear why such connections would selectively come into play for binocularly driven V1 neurons but not for monocularly driven V1 neurons, each of which comprise approximately 50% of the neurons in V1 (Trotter, 1995).

Alternatively, it may be that recurrent interactions with areas higher up the cortex modulate V1 activity for visible, but not for invisible stimuli. Over the past 15 years, percept-dependent modulation of V1 activity has been repeatedly associated with the influence of feedback connections which have been suggested to play a role in figure-ground segregation (Lamme, 1995), perceptual grouping (Roelfsema, 2006) and visual awareness (Haynes, et al., 2005;

Lamme, 2006). In this interpretation, objects and scenes are categorized quickly and automatically in the feedforward sweep (e.g. Thorpe, et al., 1996; VanRullen & Koch, 2003), whereas conscious representations in which objects acquire their phenomenal figure-ground properties, require recurrent interactions with lower visual areas. Thus a face may be categorized as such in the FFA in the feedforward sweep, but to consciously see the face requires figure-ground segregation and possibly grouping through recurrent interactions.

Perceptual hypothesis testing (Enns & Di Lollo, 2000) may be an integral part of this process, allowing low and high level neurons to be locked into certain perceptual interpretations depending on how well they match. As the match fails, recurrent interactions may fall apart, resulting in a lack of visibility for (part) of the visual input or a transition from one interpretation to another, as may also be the case in pattern masking (Fahrenfort, et al., 2007; Lamme, et al., 2002), flash suppression (Wilke, Logothetis, & Leopold, 2003) and binocular rivalry (Lee, Blake, & Heeger, 2005). In the present study, this would explain why modulatory activity in early and midlevel visual areas is seen for visible, but not for invisible stimuli, as the monocular higher level interpretation does not match the binocular input for invisible stimuli, whereas it does for visible stimuli. For such an explanation to hold, one needs to assume that binocularly plausible interpretations are favored over monocular interpretations, evidence for which exists in the literature (I. Kovacs, Papathomas, Yang, & Feher, 1996).

Our results also have implications for theories of object categorization at large. Traditionally, object categorization is thought to follow figure-ground segregation (Rubin, 1958). More recently, categorization has been thought to influence figure-ground segregation (e.g. Peterson & Gibson, 1993; 1994) or to be based on the same process as figure-ground segregation (Grill-Spector & Kanwisher, 2005; but see Mack, Gauthier, Sadr, & Palmeri, 2008). Here we show that stimulus specific activity can occur even without the neural and behavioral signature of figure-ground segregation. On the basis of these data we are inclined to conclude that initial object categorization is the first (automatic) step in visual processing, and as such precedes figure-ground segregation. This would also explain why one is able to categorize an object as soon as one becomes aware of its presence (Grill-Spector & Kanwisher, 2005), as in the proposed scheme of events the brain has already categorized the object before it is segregated from the rest of the scene.

At first glance, it may seem puzzling how an object can be categorized without it being segregated from the rest of the image prior to categorization. However, recent modeling efforts have shown the plausibility of strictly feedforward object categorization in natural scenes (Serre, et al., 2007). In addition, incremental grouping theory by Roelfsema (2006; 2000) outlines in some detail how image elements can initially be grouped and classified in feedforward base-groupings causing complex tuning properties in high visual areas, whereas incremental grouping that depends on recurrent interactions enhances the responses of neurons coding features that are bound in perception later on. We hypothesize that in this study, both invisible and visible faces and houses are detected and categorized in feedforward base-groupings, whereas only visible faces and houses are incrementally grouped in recurrent interactions, leading to visual awareness and figure-ground segregation.

Conclusion

The results of this study indicate that dichoptically invisible stimulation is sufficient to generate face and house specific activation, even when stimuli are not task relevant and there is no neural signature of figure-ground segregation. A neural signature of visual awareness and figure-ground segregation can only be found when stimuli are dichoptically visible. This suggests that initial categorization of stimuli is automatic and unconscious and can take place in the absence of visual awareness and figure-ground segregation, plausibly as a result of feedforward processing. Furthermore, these results seem to show that when attention and/or behavioral report of a stimulus is not required, the neural signature of phenomenal visual awareness and figure-ground segregation can be identified in early and midlevel visual areas. It seems likely that this signature is caused by recurrent signals between these areas, reaching all the way back to visual area V1. These findings have important implications for theories of visual awareness and object categorization.

Chapter 6.

Discussion

Summary

The experiments laid out in this thesis show that object detection (chapter 2, 3 and 4) and categorization (chapter 5) can occur outside of consciousness in the FFS. Conscious vision on the other hand goes hand in hand with successful figure-ground segregation and is selectively associated with RP (all chapters), reaching back all the way to area V1 (chapter 4 and 5). Together, these data challenge the notion that figure-ground segregation is a prerequisite for object detection and object categorization (Marr, 1982; Nakayama, et al., 1995; Rubin, 1958, see Figure 6.1a), but rather that object detection and categorization are a starting point for figure-ground segregation. In this view the FFS initially causes objects to be detected and categorized in high visual areas, subsequently triggering the recurrent processes that are necessary for figure-ground segregation and conscious vision to occur (see Figure 6.1b).

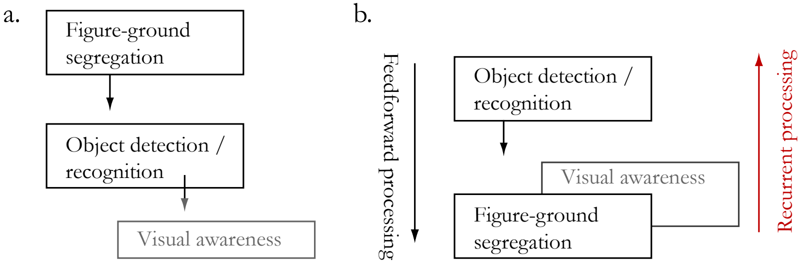


Figure 6.1 *Models of visual information processing specifying the relationship between figure-ground segregation, object detection/categorization and visual awareness.(a) Figure-ground segregation precedes object detection/recognition and visual awareness. (b) Object detection/categorization is realized in the FFS, Figure-ground segregation and visual awareness is enabled through RP. Object detection/categorization therefore precede Figure-ground segregation and visual awareness.*

The function of consciousness

If object detection and categorization take place prior to the processes effecting consciousness, it becomes questionable whether object detection and categorization can be considered primary functions of consciousness, as many people do on intuitive or definitional grounds (e.g. Dehaene, et al., 2006). This brings up the question what – if anything – consciousness is good for. As shown in this thesis as well as many other studies (e.g. Boehler, et al., 2008; Lamme, Super, et al., 1998; Pascual-Leone & Walsh, 2001; Super, et al., 2001), consciousness is consistently associated with RP. Therefore, a point of departure for answering this question maybe to inquire what RP is good for. As outlined in the introduction of this thesis, a variety of largely mutually non-exclusive proposals have been put forward in the literature, such as figure-ground segregation, binding, perception of detail and ambiguity resolution (Di Lollo, et al., 2000; Hochstein & Ahissar, 2002; Lamme, 1995; Rao & Ballard, 1999; Roelfsema, et al., 1998). Given the co-occurrence of consciousness and RP and the phenomenological nature of consciousness, it becomes appealing to view consciousness as the phenomenological expression of functions supported by RP (see Figure 6.2). From this perspective, rather than talk about the 'function' of consciousness, a more accurate description might be that RP supports certain functions, and that consciousness is the phenomenological expression of these functions.

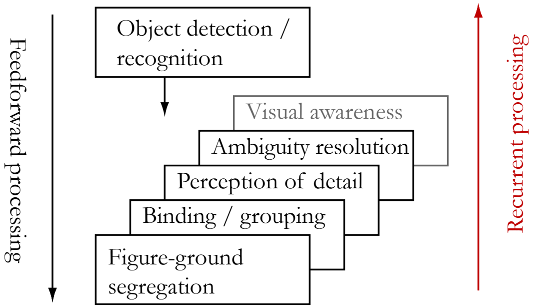


Figure 6.2 *Visual awareness is the phenomenological expression of functions carried out through RP.*

The function of RP and how it relates to attention and visual consciousness

Although the proposed functional roles of RP are diverse, a unifying principle behind them seems to be that they enable the integration of visual information across cortical areas. For example, in figure-ground segregation, the role of RP is to use information about the presence and shape of objects in high visual areas to automatically label neurons in low visual areas as either foreground or background (Lamme, 1995; Roelfsema, et al., 2002). A similar role is played out by RP in object based attention by intentionally driving a labeling operation of low level neurons representing object segments by making use of high level neurons representing the objects themselves (Roelfsema, et al., 1998). In perceptual hypothesis testing and reverse hierarchy theory, RP is used to resolve ambiguity by integrating information represented in high visual areas with large receptive fields with information represented in low visual areas with small receptive fields (Di Lollo, et al., 2000; Hochstein & Ahissar, 2002). Despite the fact that informational integration across visual areas is a central theme, aspects of these theories are incompatible, especially in the way they relate RP to attention and visual consciousness.

For example, in Roelfsema's account of object based attention as well as in reverse hierarchy theory, RP enhances spatial resolution through focused attention (Hochstein & Ahissar, 2002; Roelfsema, et al., 1998). This seems to suggest that RP and attention are closely related, or may even lead some to conclude that RP itself is the neural correlate of attention. Studies of figure-ground segregation (Marcus & Van Essen, 2002; Scholte, et al., 2006) and theories of visual consciousness (Lamme, 2003, 2006) on the other hand assert that RP is automatic (pre-attentive), and occurs independently of focused attention. In fact, a number of arguments favor the hypothesis that RP is *not* the correlate of attention. For example, it has been shown that RP continues to be in effect even when subjects do not attend to a stimulus (Scholte, et al., 2006, but also in chapter 5 of this thesis). Conversely, it has also been shown that fully attended stimuli that are masked do not evoke RP (Fahrenfort, et al., 2007, chapter 2 of this thesis; Lamme, et al., 2002) and even fully attended stimuli in plain view sometimes fail to evoke RP (Super, et al., 2001).

Such arguments are incompatible with the idea that RP is the correlate of attention. However, this does not mean that attentional selection could not operate through RP, for instance by biasing cortical pathways for processing certain objects, locations or features through feedback (e.g. Gandhi, et al., 1999; Martinez, et al., 1999; Roelfsema, et al., 1998; Zhang & Luck, 2009). Alternatively (or additionally), it may be that RP has different functional roles during different feedback iterations, first carrying signals responsible for figure-ground segregation and later on in time carrying signals related to attentional selection (Roelfsema, et al., 2007).

Another divergence of views between some of these theories concerns how they relate to consciousness, in particular the assumed minimal level of RP that is required for a subject to become conscious of a stimulus. In Hochstein and Ahissar's reverse hierarchy theory, "vision at a glance" (in which the gist of a scene is determined), comes about as soon as high cortical levels are activated. It is therefore implicitly assumed in their theory that this form of visual consciousness can be established on the basis of the FFS alone, however sparse its representation may be. If true, this would invalidate the idea that RP is necessary to establish visual consciousness of a stimulus. It contrasts quite sharply with the view laid out by Lamme (2006) and DiLollo & Enns (2000), in which RP is required for visual consciousness to emerge in any way. Although there is evidence that visual selective behavior can be triggered on the basis of the FFS (VanRullen & Koch, 2003) and that the information required to extract the gist of a scene can be processed in the FFS (Thorpe, et al., 1996), there is little evidence that subjects can *explicitly* detect the gist of a scene on the basis of the FFS alone. In fact, much evidence to the contrary exists, showing that mere activation of high level areas is insufficient to reach a conscious state (chapters 2, 4 and 5 of this thesis, but also Kouider, et al., 2008; G. Kovacs, et al., 1995; Marois, et al., 2004; Moutoussis & Zeki, 2002). Nevertheless, it remains an open question what level of RP is necessary for consciousness to emerge or how to determine this. An operational way out may be to *define* consciousness as the phenomenological expression of RP's functions, in which case the degree to which a system experiences consciousness is directly on par with the degree to which RP's functions (e.g. figure-ground segregation, binding, perception of detail and ambiguity resolution) are successfully carried out. A similar view has been put forward by Lamme (2003, 2006) in which it is proposed to define consciousness as RP.

Some speculation: consciousness as an epiphenomenon?

When defining consciousness as the phenomenological expression of functions supported by RP, the question remains why (or how) this phenomenological expression occurs. Some may argue that consciousness should be branded as an epiphenomenon: a secondary phenomenon caused by physical phenomena, but without any causal or functional role in these phenomena. This would be true if consciousness were a superfluous property altogether, one that could be taken out without consequence. In thought experiments, philosophers have invoked the notion of a 'zombie', an apparatus that acts like a human but has no phenomenal experience (e.g. in Shear, 1999). Images like these are brought up to investigate the idea that it is logically possible to imagine a world identical to our own, but without consciousness. This then is used as proof that consciousness has a special status separate from the physical world (Chalmers, 1995).

However, leaving aside the philosophical and epistemological problems associated with disproving the existence of zombies, consciousness and RP have been consistently observed together in many neuroscientific experiments, making the zombie scenario far-fetched. In practice it seems, whenever RP and its functions are in effect, their phenomenological counterpart is also expressed, and when RP is taken out, consciousness goes out the door as well. If this experimental evidence is accepted, saying that consciousness is an epiphenomenon of RP becomes a bit like saying that a bellied out sail is an epiphenomenon of a sailboat gaining speed in the wind: although being bellied out itself may not be a driving functional force (the wind is), it is an inextricable expression of the sailing event that cannot be separated from it. And although many people have the ability to imagine a sailing sailboat without a bellied out sail, few will argue that such a thought experiment now requires being bellied out to be separated from the list of events that can be explained in physical terms, and even fewer will argue that being bellied out has no functional role in sailing.

Although it is beyond the scope of this thesis to prove that consciousness is functionally relevant or how the mental world can be explained in physical terms (Chalmers' 'hard' problem, 1995), on the basis of current empirical data it seems plausible that there is a functional reason for phenomenal experience to be associated with RP. Some may find it hard to imagine how neuroscience with its strong physicalist basis can bring anything to bear on what this role may be, but we should not forget the important advances that have already been made. Importantly, being able to tie consciousness closely to functions requiring the integration of information across (visual) areas, such as figure-ground segregation, binding and ambiguity resolution (rather than to object detection and categorization), gives us a first glimpse. It suggests that phenomenal experience has something to with activating or locking into distributed representations. Although this does not answer the question why these are phenomenal in nature, it hints at the idea that there is something special about RP-induced distributed (non-local) representations that they inevitably express themselves in the brain in ways that we call phenomenal experience. The distributed nature of consciousness may therefore well be the starting point for inquiring how physical events can lead to (seemingly) non-physical phenomena. Although neuroscience has not solved the mystery of consciousness, it has more than any other discipline, started to give us pointers as to where to look.

References

Angelucci, A., Levitt, J. B., Walton, E. J. S., Hupe, J. M., Bullier, J., & Lund, J. S. (2002). Circuits for local and global signal integration in primary visual cortex. *Journal of Neuroscience, 22*(19), 8633-8646.

Atkinson, J., & Braddick, O. J. (1989). Where and what in visual-search. *Perception, 18*(2), 181-189.

Bachmann, T., Luiga, I., & Poder, E. (2005). Variations in backward masking with different masking stimuli: II. The effects of spatially quantised masks in the light of local contour interaction, interchannel inhibition, perceptual retouch, and substitution theories. *Perception, 34*(2), 139-153.

Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmid, A. M., Dale, A. M., et al. (2006). Top-down facilitation of visual recognition. *Proceedings of the National Academy of Sciences of the United States of America, 103*(2), 449-454.

Bar, M., Tootell, R. B. H., Schacter, D. L., Greve, D. N., Fischl, B., Mendola, J. D., et al. (2001). Cortical mechanisms specific to explicit visual object recognition. *Neuron, 29*(2), 529-535.

Beck, D. M., Rees, G., Frith, C. D., & Lavie, N. (2001). Neural correlates of change detection and change blindness. *Nature Neuroscience, 4*(6), 645-650.

Becker, M. W., & Anstis, S. (2004). Metacontrast masking is specific to luminance polarity. *Vision Research, 44*(21), 2537-2543.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate - a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B-Methodological, 57*(1), 289-300.

Boehler, C. N., Schoenfeld, M. A., Heinze, H. J., & Hopf, J. M. (2008). Rapid recurrent processing. gates awareness in primary visual cortex. *Proceedings of the National Academy of Sciences of the United States of America, 105*(25), 8742-8747.

Boyer, J. L., Harrison, S., & Ro, T. (2005). Unconscious processing of orientation and color without primary visual cortex. *Proceedings of the National Academy of Sciences of the United States of America, 102*(46), 16875-16879.

Brefczynski, J. A., & DeYoe, E. A. (1999). A physiological correlate of the 'spotlight' of visual attention. *Nature Neuroscience, 2*(4), 370-374.

Breitmeyer, B. G. (1984). *Visual masking: An integrative approach.* Oxfordshire/New York: Oxford/Clarendon Press/Oxford University Press.

Breitmeyer, B. G., & Ogmen, H. (2000). Recent models and findings in visual backward masking: A comparison, review, and update. *Perception & Psychophysics, 62*(8), 1572-1595.

Breitmeyer, B. G., Ro, T., & Ogmen, H. (2004). A comparison of masking by visual and transcranial magnetic stimulation: implications for the study of conscious and unconscious visual processing. *Consciousness and Cognition, 13*(4), 829-843.

Brewer, A. A., Liu, J. J., Wade, A. R., & Wandell, B. A. (2005). Visual field maps and stimulus selectivity in human ventral occipital cortex. *Nature Neuroscience, 8*(8), 1102-1109.

Bridgeman, B. (1980). Temporal response characteristics of cells in monkey striate cortex measured with metacontrast masking and brightness discrimination. *Brain Research, 196*(2), 347-364.

Buschman, T. J., & Miller, E. K. (2007). Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices. *Science, 315*(5820), 1860-1862.

Caputo, G., & Casco, C. (1999). A visual evoked potential correlate of global figure-ground segmentation. *Vision Research, 39*(9), 1597-1610.

Cavanaugh, J. R., Bair, W., & Movshon, J. A. (2002). Nature and interaction of signals from the receptive field center and surround in macaque V1 neurons. *Journal of Neurophysiology, 88*(5), 2530-2546.

Chalmers, D. J. (1995). Facing up to the problem of consciousness. *Journal of Consciousness Studies, 2*, 200-219.

Clark, R. E., Broadbent, N. J., Zola, S. M., & Squire, L. R. (2002). Anterograde amnesia and temporally graded retrograde amnesia for a nonspatial memory task after lesions of hippocampus and subiculum. *Journal of Neuroscience, 22*(11), 4663-4669.

Dehaene, S., Changeux, J. P., Naccache, L., Sackur, J., & Sergent, C. (2006). Conscious, preconscious, and subliminal processing: a testable taxonomy. *Trends in Cognitive Sciences, 10*(5), 204-211.

Dehaene, S., Naccache, L., Cohen, L., Le Bihan, D., Mangin, J. F., Poline, J. B., et al. (2001). Cerebral mechanisms of word masking and unconscious repetition priming. *Nature Neuroscience, 4*(7), 752 - 758.

Dehaene, S., Naccache, L., Le Clec'H, G., Koechlin, E., Mueller, M., Dehaene Lambertz, G., et al. (1998). Imaging unconscious semantic priming. *Nature, 395*(6702), 597-600.

Dehaene, S., Sergent, C., & Changeux, J. P. (2003). A neuronal network model linking subjective reports and objective physiological data during conscious perception. *Proceedings of the National Academy of Sciences of the United States of America, 100*(14), 8520-8525.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual-attention. *Annual Review of Neuroscience, 18*, 193-222.

Di Lollo, V., Enns, J. T., & Rensink, R. A. (2000). Competition for consciousness among visual events: The psychophysics of reentrant visual processes. *Journal of Experimental Psychology-General, 129*(4), 481-507.

Donchin, E., & Coles, M. G. H. (1988). Is the P300 component a manifestation of context updating? *Behavioral and Brain Sciences, 11*(3), 357-374.

Eimer, M., & Schlaghecken, F. (1998). Effects of masked stimuli on motor activation: Behavioral and electrophysiological evidence. *Journal of Experimental Psychology-Human Perception and Performance, 24*(6), 1737-1747.

Enns, J. T. (2004). Object substitution and its relation to other forms of visual masking. *Vision Research, 44*(12), 1321-1331.

Enns, J. T., & Di Lollo, V. (2000). What's new in visual masking? *Trends in Cognitive Sciences, 4*(9), 345-352.

Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature, 392*(6676), 598-601.

Fahrenfort, J. J., Scholte, H. S., & Lamme, V. A. F. (2007). Masking disrupts reentrant processing in human visual cortex. *Journal of Cognitive Neuroscience, 19*(9), 1488-1497.

Fahrenfort, J. J., Scholte, H. S., & Lamme, V. A. F. (2008). The spatiotemporal profile of cortical processing leading up to visual perception. *Journal of Vision, 8*(1), -.

Fang, F., & He, S. (2005). Cortical responses to invisible objects in the human dorsal and ventral pathways. *Nature Neuroscience, 8*(10), 1380-1385.

Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex, 1*(1), 1-47.

Foxe, J. J., & Simpson, G. V. (2002). Flow of activation from V1 to frontal cortex in humans - A framework for defining "early" visual processing. [Article]. *Experimental Brain Research, 142*(1), 139-150.

Francis, G. (1997). Cortical dynamics of lateral inhibition: Metacontrast masking. *Psychological Review, 104*(3), 572-594.

Gandhi, S. P., Heeger, D. J., & Boynton, G. M. (1999). Spatial attention affects brain activity in human primary visual cortex. *Proceedings of the National Academy of Sciences of the United States of America, 96*(6), 3314-3319.

Genovese, C. R., Lazar, N. A., & Nichols, T. (2002). Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *Neuroimage, 15*(4), 870-878.

Glover, G. H. (1999). Deconvolution of impulse response in event-related BOLD fMRI. *Neuroimage, 9*(4), 416-429.

Gratton, G., Coles, M. G., & Donchin, E. (1983). A new method for off-line removal of ocular artifact. *Electroencephalography and Clinical Neurophysiology, 55*(4), 468-484.

Greenwald, A. G., Klinger, M. R., & Liu, T. J. (1989). Unconscious processing of dichoptically masked words. *Memory & Cognition, 17*(1), 35-47.

Grill-Spector, K. (2003). The neural basis of object perception. *Current Opinion in Neurobiology, 13*(2), 159-166.

Grill-Spector, K., & Kanwisher, N. (2005). Visual recognition - As soon as you know it is there, you know what it is. *Psychological Science, 16*(2), 152-160.

Grill-Spector, K., Kourtzi, Z., & Kanwisher, N. (2001). The lateral occipital complex and its role in object recognition. *Vision Research, 41*(10-11), 1409-1422.

Gross, C. G., Rocha-Miranda, C. E., & Bender, D. B. (1972). Visual properties of neurons in inferotemporal cortex of the Macaque. *Journal of Neurophysiology, 35*(1), 96-111.

Hartline, H. K. (1949). Inhibition of activity of visual receptors by illuminating nearby retinal areas in the Limulus eye. *Federation Proceedings, 8*(7), 69.

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science, 293*(5539), 2425-2430.

Haynes, J. D., Driver, J., & Rees, G. (2005). Visibility reflects dynamic changes of effective connectivity between V1 and fusiform cortex. *Neuron, 46*(5), 811-821.

Heilman, K. M., & Abell, T. V. D. (1980). Right hemisphere dominance for attention: The mechanism underlying hemispheric asymmetries of inattention (neglect). *Neurology, 30*(3), 327-.

Heinrich, S. P., Andrés, M., & Bach, M. (2007). Attention and visual texture segregation. *Journal of Vision, 7*(6), 1-10.

Hillyard, S. A., Hink, R. F., Schwent, V. L., & Picton, T. W. (1973). Electrical signs of selective attention in the human brain. *Science, 182*(108), 177-180.

Hochstein, S., & Ahissar, M. (2002). View from the top: Hierarchies and reverse hierarchies in the visual system. *Neuron, 36*(5), 791-804.

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology, 160*, 106-154.

Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology, 195*(1), 215-243.

Hupe, J. M., James, A. C., Payne, B. R., Lomber, S. G., Girard, P., & Bullier, J. (1998). Cortical feedback improves discrimination between figure and background by V1, V2 and V3 neurons. [10.1038/29537]. *Nature, 394*(6695), 784-787.

Jolij, J., Scholte, H. S., Van Gaal, S., & Lamme, V. A. F. (in press). The brain decides late: long latency visual evoked potentials reflect perceptual decisions. *Journal of Cognitive Neuroscience.*

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience, 17*(11), 4302-4311.

Kastner, S., Pinsk, M. A., De Weerd, P., Desimone, R., & Ungerleider, L. G. (1999). Increased activity in human visual cortex during directed attention in the absence of visual stimulation. *Neuron, 22*(4), 751-761.

Kastner, S., & Ungerleider, L. G. (2000). Mechanisms of visual attention in the human cortex. *Annual Review of Neuroscience, 23*, 315-341.

Knierim, J. J., & Van Essen, D. C. (1992). Neuronal responses to static texture patterns in area-V1 of the alert macaque monkey. *Journal of Neurophysiology, 67*(4), 961-980.

Koivisto, M., Revonsuo, A., & Lehtonen, M. (2006). Independence of visual awareness from the scope of attention: An electrophysiological study. *Cerebral Cortex, 16*(3), 415-424.

Koivisto, M., Revonsuo, A., & Salminen, N. (2005). Independence of visual awareness from attention at early processing stages. *Neuroreport, 16*(8), 817-821.

Kok, A. (2001). On the utility of P3 amplitude as a measure of processing capacity. *Psychophysiology, 38*(3), 557-577.

Kolb, F. C., & Braun, J. (1995). Blindsight in normal observers. *Nature, 377*(6547), 336-338.

Konen, C. S., & Kastner, S. (2008). Two hierarchically organized neural systems for object information in human visual cortex. *Nature Neuroscience, 11*(2), 224-231.

Kouider, S., Eger, E., Dolan, R., & Henson, R. N. (2008). Activity in face-responsive brain regions is modulated by invisible, attended faces: evidence from masked priming. *Cerebral Cortex, 19*(1), 13-23.

Kovacs, G., Vogels, R., & Orban, G. A. (1995). Cortical correlate of pattern backward-masking. *Proceedings of the National Academy of Sciences of the United States of America, 92*(12), 5587-5591.

Kovacs, I., Papathomas, T. V., Yang, M., & Feher, A. (1996). When the brain changes its mind: Interocular grouping during binocular rivalry. *Proceedings of the National Academy of Sciences of the United States of America, 93*(26), 15508-15511.

Kranczioch, C., Debener, S., & Engel, A. K. (2003). Event-related potential correlates of the attentional blink phenomenon. *Cognitive Brain Research, 17*(1), 177-187.

Kranczioch, C., Debener, S., Schwarzbach, J., Goebel, R., & Engel, A. K. (2005). Neural correlates of conscious perception in the attentional blink. *Neuroimage, 24*(3), 704-714.

Lamme, V. A. F. (1995). The neurophysiology of figure ground segregation in primary visual-cortex. *Journal of Neuroscience, 15*(2), 1605-1615.

Lamme, V. A. F. (2001). Blindsight: the role of feedforward and feedback corticocortical connections. *Acta Psychologica, 107*(1-3), 209-228.

Lamme, V. A. F. (2003). Why visual attention and awareness are different. *Trends in Cognitive Sciences, 7*(1), 12-18.

Lamme, V. A. F. (2004). Separate neural definitions of visual consciousness and visual attention; a case for phenomenal awareness. *Neural Networks, 17*(5-6), 861-872.

Lamme, V. A. F. (2006). Towards a true neural stance on consciousness. *Trends in Cognitive Sciences, 10*(11), 494-501.

Lamme, V. A. F., Rodriguez-Rodriguez, V., & Spekreijse, H. (1999). Separate processing dynamics for texture elements, boundaries and surfaces in primary visual cortex of the macaque monkey. *Cerebral Cortex, 9*(4), 406-413.

Lamme, V. A. F., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neurosciences, 23*(11), 571-579.

Lamme, V. A. F., Super, H., Landman, R., Roelfsema, P. R., & Spekreijse, H. (2000). The role of primary visual cortex (V1) in visual awareness. *Vision Research, 40*(10-12), 1507-1521.

Lamme, V. A. F., Super, H., & Spekreijse, H. (1998). Feedforward, horizontal, and feedback processing in the visual cortex. *Current Opinion in Neurobiology, 8*(4), 529-535.

Lamme, V. A. F., Van Dijk, B. W., & Spekreijse, H. (1992). Texture segregation is processed by primary visual cortex in man and monkey. Evidence from VEP experiments. *Vision Research, 32*(5), 797-807.

Lamme, V. A. F., Van Dijk, B. W., & Spekreijse, H. (1993). Contour from motion processing occurs in primary visual cortex. *Nature, 363*(6429), 541-543.

Lamme, V. A. F., Zipser, K., & Spekreijse, H. (1998). Figure-ground activity in primary visual cortex is suppressed by anesthesia. *Proceedings of the National Academy of Sciences of the United States of America, 95*(6), 3263-3268.

Lamme, V. A. F., Zipser, K., & Spekreijse, H. (2002). Masking interrupts figure-ground signals in V1. *Journal of Cognitive Neuroscience, 14*(7), 1044-1053.

LaRock, E. (2007). Disambiguation, binding, and the unity of visual consciousness. *Theory & Psychology, 17*(6), 747-777.

Lee, S. H., Blake, R., & Heeger, D. J. (2005). Traveling waves of activity in primary visual cortex during binocular rivalry. *Nature Neuroscience, 8*(1), 22-23.

Liu, J., Harris, A., & Kanwisher, N. (2002). Stages of processing in face perception: an MEG study. [10.1038/nn909]. *Nature Neuroscience, 5*(9), 910-916.

Logothetis, N. K., Pauls, J., Augath, M., Trinath, T., & Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fMRI signal. *Nature, 412*(6843), 150-157.

Luck, S. J., Chelazzi, L., Hillyard, S. A., & Desimone, R. (1997). Neural mechanisms of spatial selective attention in areas V1, V2, and V4 of macaque visual cortex. *Journal of Neurophysiology, 77*(1), 24-42.

Luck, S. J., Girelli, M., McDermott, M. T., & Ford, M. A. (1997). Bridging the gap between monkey neurophysiology and human perception: An ambiguity resolution theory of visual selective attention. *Cognitive Psychology, 33*(1), 64-87.

Luck, S. J., Vogel, E. K., & Shapiro, K. L. (1996). Word meanings can be accessed but not reported during the attentional blink. *Nature, 383*(6601), 616-618.

Lumer, E. D., & Rees, G. (1999). Covariation of activity in visual and prefrontal cortex associated with subjective visual perception. *Proceedings of the National Academy of Sciences of the United States of America, 96*(4), 1669-1673.

Mack, M. L., Gauthier, I., Sadr, J., & Palmeri, T. J. (2008). Object,detection and basic-level categorization: Sometimes you know it is there before you know what it is. *Psychonomic Bulletin & Review, 15*(1), 28-35.

Macknik, S. L., & Livingstone, M. S. (1998). Neuronal correlates of visibility and invisibility in the primate visual system. *Nature Neuroscience, 1*(2), 144-149.

Macknik, S. L., & Martinez-Conde, S. (2004). Dichoptic visual masking reveals that early binocular neurons exhibit weak interocular suppression: Implications for binocular vision and visual awareness. *Journal of Cognitive Neuroscience, 16*(6), 1049-1059.

Malach, R., Reppas, J. B., Benson, R. R., Kwong, K. K., Jiang, H., Kennedy, W. A., et al. (1995). Object-related activity revealed by functional magnetic-resonance-imaging in human occipital cortex. *Proceedings of the National Academy of Sciences of the United States of America, 92*(18), 8135-8139.

Marcus, D. S., & Van Essen, D. C. (2002). Scene segmentation and attention in primate cortical areas V1 and V2. *Journal of Neurophysiology, 88*(5), 2648-2658.

Marois, R., Yi, D. J., & Chun, M. M. (2004). The neural fate of consciously perceived and missed events in the aftentional blink. *Neuron, 41*(3), 465-472.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco: W.H. Freeman.

Martinez, A., Anllo-Vento, L., Sereno, M. I., Frank, L. R., Buxton, R. B., Dubowitz, D. J., et al. (1999). Involvement of striate and extrastriate visual cortical areas in spatial attention. *Nature Neuroscience, 2*(4), 364-369.

McCormick, P. A. (1997). Orienting attention without awareness. *Journal of Experimental Psychology-Human Perception and Performance, 23*(1), 168-180.

Mesulam, M. M. (1999). Spatial attention and neglect: parietal, frontal and cingulate contributions to the mental representation and attentional targeting of salient extrapersonal events. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences, 354*(1387), 1325-1346.

Miller, B. T., & D'Esposito, M. (2005). Searching for "the top" in top-down control. *Neuron, 48*(4), 535-538.

Milner, A. D., & Goodale, M. A. (1995). *The visual brain in action*: Oxford University Press, Oxford.

Moutoussis, K., & Zeki, S. (2002). The relationship between cortical activation and perception investigated with invisible stimuli. *Proceedings of the National Academy of Sciences of the United States of America, 99*(14), 9527-9532.

Mulckhuyse, M., Talsma, D., & Theeuwes, J. (2007). Grabbing attention without knowing: Automatic capture of attention by subliminal spatial cues. *Visual Cognition, 15*(7), 779-788.

Müller, N. G., & Kleinschmidt, A. (2004). The attentional 'spotlight's' penumbra: center-surround modulation in striate cortex. *Neuroreport, 15*(6), 977-980.

Murray, S. O., Boyaci, H., & Kersten, D. (2006). The representation of perceived angular size in human primary visual cortex. [10.1038/nn1641]. *Nature Neuroscience, 9*(3), 429-434.

Naccache, L., Gaillard, R., Adam, C., Hasboun, D., Clemenceau, S., Baulac, M., et al. (2005). A direct intracranial record of emotions evoked by subliminal words. *Proceedings of the National Academy of Sciences of the United States of America, 102*(21), 7713-7717.

Nakayama, K., Zijiang, J. H., & Shinsuke, S. (1995). Visual surface representation: A critical link between lower-level and higher-level vision. In D. N. Osherson, L. R. Gleitman & S. M. Kosslyn (Eds.), *An Invitation to Cognitive Science* (pp. 1-70). Cambridge: MIT Press.

Nieuwenhuis, S., Aston-Jones, G., & Cohen, J. D. (2005). Decision making, the p3, and the locus coeruleus-norepinephrine system. *Psychological Bulletin, 131*(4), 510-532.

Nunez, P., & Srinivasan, R. (2006). *Electric fields of the brain*. Oxford: Oxford University Press.

Ogmen, H., Breitmeyer, B. G., & Melvin, R. (2003). The what and where in visual masking. *Vision Research, 43*(12), 1337-1350.

Oram, M. W., & Perrett, D. I. (1992). Time course of neural responses discriminating different views of the face and head. *Journal of Neurophysiology, 68*(1), 70-84.

Pascual-Leone, A., & Walsh, V. (2001). Fast backprojections from the motion to the primary visual area necessary for visual awareness. *Science, 292*(5516), 510-512.

Perrin, F., Pernier, J., Bertrand, O., & Echallier, J. F. (1989). Spherical splines for scalp potential and current-density mapping. *Electroencephalography and Clinical Neurophysiology, 72*(2), 184-187.

Peterson, M. A., & Gibson, B. S. (1993). Shape-recognition inputs to figure ground organization in 3-dimensional displays. *Cognitive Psychology, 25*(3), 383-429.

Peterson, M. A., & Gibson, B. S. (1994). Must figure-ground organization precede object recognition - an assumption in peril. *Psychological Science, 5*(5), 253-259.

Philiastides, M. G., Ratcliff, R., & Sajda, P. (2006). Neural representation of task difficulty and decision making during perceptual categorization: A timing diagram. *Journal of Neuroscience, 26*(35), 8965-8975.

Philiastides, M. G., & Sajda, P. (2006). Temporal characterization of the neural correlates of perceptual decision making in the human brain. *Cerebral Cortex, 16*(4), 509-518.

Pins, D., & ffytche, D. (2003). The neural correlates of conscious vision. *Cerebral Cortex, 13*(5), 461-474.

Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. [10.1038/4580]. *Nature Neuroscience, 2*(1), 79-87.

Ro, T., Breitmeyer, B., Burton, P., Singhal, N. S., & Lane, D. (2003). Feedback contributions to visual awareness in human occipital cortex. *Current Biology, 13*(12), 1038-1041.

Rodriguez, E., George, N., Lachaux, J. P., Martinerie, J., Renault, B., & Varela, F. J. (1999). Perception's shadow: long-distance synchronization of human brain activity. *Nature, 397*(6718), 430-433.

Roelfsema, P. R. (2006). Cortical algorithms for perceptual grouping. *Annual Review of Neuroscience, 29*, 203-227.

Roelfsema, P. R., Lamme, V. A. F., & Spekreijse, H. (1998). Object-based attention in the primary visual cortex of the macaque monkey. *Nature, 395*(6700), 376-381.

Roelfsema, P. R., Lamme, V. A. F., & Spekreijse, H. (2000). The implementation of visual routines. *Vision Research, 40*(10-12), 1385-1411.

Roelfsema, P. R., Lamme, V. A. F., Spekreijse, H., & Bosch, H. (2002). Figure-ground segregation in a recurrent network architecture. *Journal of Cognitive Neuroscience, 14*(4), 525-537.

Roelfsema, P. R., Tolboom, M., & Khayat, P. S. (2007). Different processing phases for features, figures, and selective attention in the primary visual cortex. *Neuron, 56*(5), 785-792.

Rolls, E. T., & Tovee, M. J. (1994). Processing speed in the cerebral-cortex and the neurophysiology of visual masking. *Proceedings of the Royal Society of London Series B-Biological Sciences, 257*(1348), 9-15.

Rubin, E. (1958). *Figure and ground.* New York: Van Nostrand.

Scholte, H. S., Witteveen, S. C., Spekreijse, H., & Lamme, V. A. F. (2006). The influence of inattention on the neural correlates of scene segmentation. *Brain Research, 1076*(1), 106-115.

Schultz, W. (2001). Reward signaling by dopamine neurons. *Neuroscientist, 7*(4), 293-302.

Schwarzlose, R. F., Swisher, J. D., Dang, S. B., & Kanwisher, N. (2008). The distribution of category and location information across object-selective regions in human visual cortex. *Proceedings of the National Academy of Sciences of the United States of America, 105*(11), 4447-4452.

Sereno, M. I., Dale, A. M., Reppas, J. B., Kwong, K. K., Belliveau, J. W., Brady, T. J., et al. (1995). Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science, 268*(5212), 889-893.

Sergent, C., Baillet, S., & Dehaene, S. (2005). Timing of the brain events underlying access to consciousness during the attentional blink. *Nature Neuroscience, 8*(10), 1391-1400.

Serre, T., Oliva, A., & Poggio, T. (2007). A feedforward architecture accounts for rapid categorization. *Proceedings of the National Academy of Sciences of the United States of America, 104*(15), 6424-6429.

Shear, J. (1999). *Explaining consciousness: The hard problem*: MIT Press, 1999.

Silvanto, J., Cowey, A., Lavie, N., & Walsh, V. (2005). Striate cortex (V1) activity gates awareness of motion. *Nature Neuroscience, 8*(2), 143-144.

Sirotin, Y. B., & Das, A. (2009). Anticipatory haemodynamic signals in sensory cortex not predicted by local neuronal activity. *Nature, 457*(7228), 475-U476.

Skiera, G., Petersen, D., Skalej, M., & Fahle, M. (2000). Correlates of figure-ground segregation in fMRI. *Vision Research, 40*(15), 2047-2056.

Slotnick, S. D., Schwarzbach, J., & Yantis, S. (2003). Attentional inhibition of visual processing in human striate and extrastriate cortex. *Neuroimage, 19*(4), 1602-1611.

Smith, A. T., Singh, K. D., & Greenlee, M. W. (2000). Attentional suppression of activity in the human visual cortex. *Neuroreport, 11*(2), 271-277.

Smith, A. T., Singh, K. D., Williams, A. L., & Greenlee, M. W. (2001). Estimating receptive field size from fMRI data in human striate and extrastriate visual cortex. *Cerebral Cortex, 11*(12), 1182-1190.

Smith, A. T., Williams, A. L., & Singh, K. D. (2004). Negative BOLD in the visual cortex: Evidence against blood stealing. *Human Brain Mapping, 21*(4), 213-220.

Smith, M. L., Gosselin, F., & Schyns, P. G. (2004). Receptive fields for flexible face categorizations. *Psychological Science, 15*(11), 753-761.

Spratling, M. W., & Johnson, M. H. (2004). A feedback model of visual attention. *Journal of Cognitive Neuroscience, 16*(2), 219-237.

Stern, C. E., Corkin, S., Gonzalez, R. G., Guimaraes, A. R., Baker, J. R., Jennings, P. J., et al. (1996). The hippocampal formation participates in novel picture encoding: Evidence from functional magnetic resonance imaging. *Proceedings of the National Academy of Sciences of the United States of America, 93*(16), 8660-8665.

Super, H., Spekreijse, H., & Lamme, V. A. F. (2001). Two distinct modes of sensory processing observed in monkey primary visual cortex (V1). *Nature Neuroscience, 4*(3), 304-310.

Super, H., van der Togt, C., Spekreijse, H., & Lamme, V. A. F. (2003). Internal state of monkey primary visual cortex (V1) predicts figure-ground perception. *Journal of Neuroscience, 23*(8), 3407-3414.

Thompson, K. G., & Schall, J. D. (1999). The detection of visual signals by macaque frontal eye field during masking. *Nature Neuroscience, 2*(3), 283-288.

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature, 381*(6582), 520-522.

Tong, F., Nakayama, K., Vaughan, J. T., & Kanwisher, N. (1998). Binocular rivalry and visual awareness in human extrastriate cortex. *Neuron, 21*(4), 753-759.

Tootell, R. B. H., Hadjikhani, N., Hall, E. K., Marrett, S., Vanduffel, W., Vaughan, J. T., et al. (1998). The retinotopy of visual spatial attention. *Neuron, 21*(6), 1409-1422.

Treisman, A. M., & Gelade, G. (1980). Feature-integration theory of attention. *Cognitive Psychology, 12*(1), 97-136.

Trotter, Y. (1995). Cortical representation of visual three-dimensional space. *Perception, 24*(3), 287-298.

Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. In D. J. Ingle, M. A. Goodale & R. J. W. Mansfield (Eds.), *Analysis of visual behavior* (pp. 549–586): Cambridge, MA: The MIT Press.

VanRullen, R., & Koch, C. (2003). Visual selective behavior can be triggered by a feed-forward process. *Journal of Cognitive Neuroscience, 15*(2), 209-217.

Vigneau, M., Beaucousin, V., Herve, P. Y., Duffau, H., Crivello, F., Houde, O., et al. (2006). Meta-analyzing left hemisphere language areas: Phonology, semantics, and sentence processing. *NeuroImage, 30*(4), 1414-1432.

Vorberg, D., Mattler, U., Heinecke, A., Schmidt, T., & Schwarzbach, J. (2003). Different time courses for visual perception and action priming. *Proceedings of the National Academy of Sciences of the United States of America, 100*(10), 6275-6280.

Wandell, B. A. (1999). Computational neuroimaging of human visual cortex. *Annual Review of Neuroscience, 22*, 145-+.

Wandell, B. A., Brewer, A. A., & Dougherty, R. F. (2005). Visual field map clusters in human cortex. *Philosophical Transactions of the Royal Society B-Biological Sciences, 360*(1456), 693-707.

Weidner, R., Shah, N. J., & Fink, G. R. (2006). The neural basis of perceptual hypothesis generation and testing. *Journal of Cognitive Neuroscience, 18*(2), 258-266.

Weiskrantz, L., Barbur, J. L., & Sahraie, A. (1995). Parameters affecting conscious versus unconscious visual-discrimination with damage to the visual-cortex (V1). *Proceedings of the National Academy of Sciences of the United States of America, 92*(13), 6122-6126.

Weiskrantz, L., Warrington, E. K., Sanders, M. D., & Marshall, J. (1974). Visual capacity in the hemianopic field following a restricted occipital ablation. *Brain, 97*(1), 709-728.

Wickens, T. D. (2002). *Elementary signal detection theory*. Oxford ; New York: Oxford University Press.

Wilenius-Emet, M., Revonsuo, A., & Ojanen, V. (2004). An electrophysiological correlate of human visual awareness. *Neuroscience Letters, 354*(1), 38-41.

Wilke, M., Logothetis, N. K., & Leopold, D. A. (2003). Generalized flash suppression of salient visual targets. *Neuron, 39*(6), 1043-1052.

Woodman, G. F., & Luck, S. J. (2003). Dissociations among attention, perception, and awareness during object-substitution masking. *Psychological Science, 14*(6), 605-611.

Yoshor, D., Bosking, W. H., Ghose, G. M., & Maunsell, J. H. (2007). Receptive fields in human visual cortex mapped with surface electrodes. *Cerebral Cortex, 17*(10), 2293-2302.

Zhang, W. W., & Luck, S. J. (2009). Feature-based attention modulates feedforward visual processing. *Nature Neuroscience, 12*(1), 24-25.

Zipser, K., Lamme, V. A. F., & Schiller, P. H. (1996). Contextual modulation in primary visual cortex. *Journal of Neuroscience, 16*(22), 7376-7389.

Summary in English

This thesis is about conscious and unconscious vision. When somebody enters 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 an elevator has the same sensation as our own when you see somebody come in. What makes us conscious of the things we see? What brain processes are responsible for conscious perception? And what brain processes have nothing to do with consciousness and are more similar to the way electronic eyes process visual information? The aim of this thesis is to begin to answer questions like these.

Descartes' dualism is often mentioned in jest, and although Descartes is probably most known for the notion of dualism, he is actually one of the founding fathers of the way we look at the relationship between mind and body today: the human body as a machine in 'Traité de l'homme'. In this book, Descartes introduces the concept of a 'reflex': an involuntary movement that is executed directly, without intervention by the conscious mind. Reflexes are obviously mechanistic and it is clear how they come about. How the conscious mind comes about is less clear.

As the brain processes visual information, it initially responds as in a reflex: fast, automatically and unconsciously. This initial reflex-like process is known as the Fast Feedforward Sweep (FFS). In this thesis, I defend the position that the FFS supports functions that have previously been associated with consciousness, such as the detection and categorization of objects. Experiments in this thesis show that the brain is able to detect and categorize objects without a person becoming conscious of these objects. Furthermore, I attempt to prove that the processes responsible for conscious perception only emerge after the FFS, in what is referred to as feedback or Recurrent Processing (RP). RP is the stream of processing through which early visual areas are reactivated, making use of nerve pathways that flow back from higher to lower visual areas. I show that this reactivation as a result of RP is what causes visual consciousness.

Nederlandse samenvatting

Dit proefschrift gaat over bewuste en onbewuste visuele waarneming. Wanneer iemand een lift instapt, merkt het electronisch oogje bij de deur van de lift dit op, en zorgt dat de deur open blijft. Niemand zal echter serieus beweren dat de sensatie die de lift heeft dezelfde is als onze eigen sensatie wanneer iemand binnen komt stappen. Wat zorgt ervoor dat we ons bewust zijn van de dingen die we zien? Welke hersenprocessen zijn verantwoordelijk voor onze bewuste waarneming? En welke hersenprocessen hebben daar juist niets mee te maken en lijken meer op de processen van het lift-oogje? In dit proefschrift probeer ik dit soort vragen te beantwoorden.

Hoewel misschien nog wel het meest bekend om zijn vaak schertsend aangehaalde dualisme, is Descartes een van de eerste aanzetters tot de manier waarop we vandaag de dag naar de relatie tussen geest en lichaam kijken: het menselijk lichaam als machine in 'Traité de l'homme'. In dit boek introduceert Descartes het begrip 'reflex': bewegingen die direct en zonder tussenkomst van het bewustzijn worden uitgevoerd. Van de reflex is het duidelijk dat deze een mechanistische oorsprong heeft, en hoe deze tot stand komt. Van de geest is dit minder duidelijk.

Wanneer het brein de visuele buitenwereld verwerkt, reageert het in eerste instantie ook als in een reflex: snel, automatisch en onbewust. Deze initiële reflexmatige verwerkingsstroom wordt wel de Fast Feedforward Sweep (FFS) genoemd. In dit proefschrift verdedig ik de stelling dat in de FFS functies tot stand komen die voorheen werden beschouwd als behorend tot het domein van het bewustzijn, zoals het detecteren en categoriseren van objecten. Zo blijkt dat het brein objecten kan detecteren en categoriseren zonder dat iemand zich bewust wordt van die objecten. Verder probeer ik aan te tonen dat de processen die zorgen voor bewuste gewaarwording pas na de FFS op gang komen, met wat ook wel feedback, of Recurrent Processing (RP) genoemd wordt. RP is hersenactiviteit die ervoor zorgt dat vroegvisuele hersengebieden ge-heractiveerd worden via zenuwbanen die teruglopen van hoger gelegen visuele hersengebieden naar lager gelegen visuele hersengebieden. Ik laat zien dat het juist de her-activatie ten gevolge van RP is die voor bewustzijn zorgt.

Dankwoord

Tot slot het zo belangrijke dankwoord, waarschijnlijk het meestgelezen onderdeel van de meeste proefschriften. Een promotietraject is eufemistisch uitgedrukt niet altijd even makkelijk. Steeds ben ik echter weer tot de conclusie gekomen dat ik niet iets leukers kon doen dan dit onderzoek, ondanks (en misschien ook wel een beetje dankzij) alle persoonlijke en professionele hobbels die daarbij genomen moesten worden. Het is moeilijk je voor te stellen hoe je was vóórdat je de kennis en ervaring had opgedaan die je hebt opgedaan, en zo lijkt het eigenlijk altijd een beetje alsof het altijd al zo was, alsof je niets geleerd hebt. Maar als ik mijn best doe kan ik me nog herinneren dat ik het onderzoeksvoorstel voor deze onderzoeksplek las, en dat er een romantisch beeld van Bewustzijns Onderzoek boven kwam drijven. Dat ik niet eens goed wist wat een techniek als fMRI nu eigenlijk precies inhield. Als ik me dat weer voor de geest haal dan wordt duidelijk hoeveel ik geleerd heb in de afgelopen jaren. In dit dankwoord wil ik allen bedanken die hier direct of indirect aan hebben bijgedragen.

Allereerst natuurlijk de mensen die deze promotie mogelijk hebben gemaakt. Mijn promotor Victor, die me heeft aangenomen en zonder wie deze promotie niet had plaatsgevonden. Als ik weer eens met stapels grafieken en lastig interpreteerbare data binnen kwam stormen nam je altijd een stap terug en stelde de vraag wat we nou eigenlijk ook alweer wilden weten en hoe de data daar dan een antwoord op geven. Dat is op zo'n moment vaak een rottige rotvraag die je na twee maanden analyseren vreemd genoeg meestal geheel uit het oog verloren bent, maar natuurlijk wel de enige juiste vraag. Bedankt voor je nuchtere kijk en kernachtige commentaar. Ik heb er veel van geleerd, zonder dat had ik de uitgang denk ik niet gevonden. Bedankt ook voor het vertrouwen dat je in me gesteld hebt door mij opnieuw voor drie jaar aan te stellen.

Steven, ons 'aparte geval'. Soms hebben we wat wrijving gehad, maar op kritieke momenten heb ik me altijd erg door je gesteund gevoeld. Je optimisme en onwrikbare vertrouwen dat alles altijd goed komt zijn soms bevreemdend maar ook ontzettend fijn wanneer je het even niet meer ziet. Natuurlijk ben ik je veel dank verschuldigd op operationeel en inhoudelijk vlak. Zonder je ruime kennis van de nieuwste fMRI en EEG technieken was ik niet gekomen waar ik nu ben. Je bereidheid altijd een helpende hand uit te steken zijn een grote impuls voor zowel mijn onderzoek als voor ons lab.

De leden van de beoordelingscommissie prof. dr. C.M.A. Pennartz, prof. dr. K.R. Ridderinkhof, prof. dr. P. Roelfsema, prof. dr. F. Verstraten en prof. dr. B.L.M.F. De Gelder wil ik bedanken zich bereid te hebben gesteld mijn proefschrift te lezen en van commentaar en vragen te voorzien.

Tevens wil ik graag alle Lammetjes bedanken die in de jaren voorbij zijn gekomen. Grijs-verleden-Lammetje Jacob, gevolgd door Lammetje Myriam die sneller schaap werd dan ik door mij met een lammergangetje in te halen :-) Jammer dat jullie er niet meer zijn. Gelukkig heeft Myriam een andere Vandenbroucke ter vervanging gestuurd. Later kwamen oud-student Lammetjes Ilja en Simon. Ilja en Simon: jullie zijn fijne kamergenoten, altijd behulpzaam, sociaal en gezellig. Ik zal jullie missen, mochten jullie of ik ooit vertrekken. Ook niet te vergeten de andere hoekkamer met Martijn en het buitenechtelijke semi-Lammetje Andries. Recent zijn daar de Europees-subsidiegeld-Lammetjes Anouk, Julia en Annelinde en het Beta-Gamma-Lammetje Iris er bij gekomen. En dan hadden we ook nog de korte-snack-tussendoor-Lammetjes Klaartje en Titia. Allemaal ontzettend fijne collega's die werken hier gezellig gemaakt hebben en nog steeds maken.

Dank ook aan de andere (oud) psychonomie collega's, zoals het tegenwoordig uitgestorven Van Berkum-groepje Mante en Marte. Ik mis jullie nog steeds een beetje ondanks de waardige opvolgers in die hoekkamer. Nog plannen om terug te komen? Dank aan ons secretariaat Ellen en Hubert voor jullie hulpvaardigheid, jammer dat je weg bent Ellen. En dank natuurlijk aan alle andere gezellige collega's van de 6e!

Dan de mensen van andere etages in (of nabij) het psychologiegebouw. Bedankt Richard en Frans, dat jullie mij in de eindtijd van mijn promotie een jaar hebben aangesteld op het SECOND project. Een interessant project dat mij kennis heeft laten maken met een totaal andere tak van onderzoek. Een verdieping hoger ziet de onderzoekswereld er al anders uit, laat staan wanneer je via de overloop helemaal naar een ander gebouw loopt! Interdisciplinair onderzoek heet dat heel modern. Door de aanstelling van vier dagen heb net die ene dag armslag gehad om de laatste loodjes van mijn promotie af te kunnen ronden. Nu nog de laatste loodjes van het SECOND project!

Niet te vergeten: alle mensen die mij op enig moment ondersteund hebben gedurende mijn promotietijd. Dank aan Marcus en de andere mensen van TOP. Altijd klaar om een knoppenkast in elkaar te draaien, een dichgeslibte fan schoon te blazen, een opgeblazen voeding te vervangen of een idiote beamerkastconstructie in elkaar te lassen. Zonder jullie was het niet gelukt. Dank aan de studenten die stages en werkstukken bij mij hebben gedaan: IJsbrand, Anne, Thomas en Niels, jullie inzet bij dataverzameling was een enome hulp.

Op het persoonlijke vlak wil ik mijn vrienden en famlie bedanken, met wie ik de ups en downs van het AIO-schap heb gedeeld. Het is fijn te weten dat jullie er altijd zijn, in mooie tijden en in moeilijke tijden. Een apart woordje van dank voor Janet en Richard, die ik bereid heb gevonden mijn paranimfen te zijn en de rompslomp die dat met zich mee brengt op zich te nemen. Dank aan mijn lieve familie, die het gemekker en gezeur over de onfortuinlijkheden van de promovendus aan heeft moeten horen. Ik hoop dat ik wat meer tijd kan vinden om jullie op te zoeken nu de grootste stofwolken van de afgelopen anderhalf jaar zijn neergedaald.

En dan wil ik natuurlijk mijn liefste Janneke bedanken, die ik tijdens (en dankzij) mijn promotie heb ontmoet en die het met me uit heeft weten te houden. Wat hebben we veel gedaan en beleefd samen de afgelopen jaren! De toren die we gebouwd hebben is stevig, en bovenop zit ons lieve kleine meisje te glunderen en te kletsen (Axi! Appel! Aardbei!). Dank je voor alles.

Johannes
Amsterdam, augustus 2009