Hierarchical information processing in the visual system

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor aan
de Vrije Universiteit Amsterdam,
op gezag van de rector magnificus
prof.dr. V. Subramaniam,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de Faculteit der Aard- en Levenswetenschappen
op vrijdag 9 februari 2018 om 11.45 uur
in de aula van de universiteit,
De Boelelaan 1105

door

Bram van Vught

geboren te Amsterdam
promotor: prof.dr. P.R. Roelfsema

copromotor: dr. C v/d Togt
Overige commissieleden:

prof.dr. S. Dehaene
prof.dr. S. Dumoulin
prof.dr. T. Knapen
prof.dr. C. N. Levelt
prof.dr. S. Spijker
# Content

<table>
<thead>
<tr>
<th>CHAPTER 1</th>
<th>Introduction</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHAPTER 2</td>
<td>At the threshold of conscious perception: information loss and response bias in visual and frontal cortex</td>
<td>27</td>
</tr>
<tr>
<td>CHAPTER 3</td>
<td>The contribution of AMPA and NMDA receptors to persistent firing in the dLPFC during working memory delays</td>
<td>73</td>
</tr>
<tr>
<td>CHAPTER 4</td>
<td>Texture segregation causes early figure enhancement and later ground suppression in areas V1 and V4 of visual cortex</td>
<td>109</td>
</tr>
<tr>
<td>CHAPTER 5</td>
<td>The segmentation of proto-objects in monkey primary visual cortex</td>
<td>145</td>
</tr>
<tr>
<td>CHAPTER 6</td>
<td>Discussion</td>
<td>187</td>
</tr>
<tr>
<td></td>
<td>Summary</td>
<td>211</td>
</tr>
<tr>
<td></td>
<td>Acknowledgement</td>
<td>215</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction
Sensation and Perception

Evolution equipped us with a stunning capacity to sense signals from the environment. Through random genetic mutations, and natural selection as the driving force, our sensory systems and our brain adjusted to physical properties of the environment that we lived in throughout our evolutionary past. This way most of us are able to hear changes in air vibrations as sound, feel mechanical pressure on our body as touch, detect molecules in our mouth and nose as taste and smell respectively, and are able to see the outside world through the detection of electromagnetic waves by our eyes.

Although this capacity to perceive our environment is very remarkable, our sensation of the outside world is at the same time also very limited. The receptors in our ears through which we detect air vibrations are sensitive to frequencies between 20 to 20 KHz. Many other animal species have much broader hearing ranges or are able to hear frequencies well beyond the human hearing range. Elephants for instance can hear air vibrations at 14-16 Hz, while mice can hear air vibrations up to 80 KHz. Similarly, a dog has more than 220 million olfactory receptors while humans have only 5 million. Although the olfactory sensitivity of dogs clearly outperforms that of humans, we have been clever enough to use their superior capacity for our own benefit by letting dogs find hidden objects like drugs, mines, buried people and recently even cancer for us. The photoreceptors in our eyes are also sensitive to only a small portion of electromagnetic radiation that exists in the outside world, and only this part of the electromagnetic spectrum is therefore visible to us through the experience of different colors (Fig. 1).

Besides the limited capacity of our senses, our brain is also limited in the amount of incoming sensory information that it is able to
process. As our brain cannot process all the information that reaches our senses at once and we are therefore not capable of processing the entire environment around us in a single take, our perception at any given moment depends on a selection of features from the outside world that are behaviorally relevant.

**Figure 1 | Electromagnetic spectrum.** Illustration of the range of frequencies or wavelengths over which electromagnetic radiation extends, from very short wavelength and high energy gamma rays to very long wavelength and low energy radio waves. Inset: The visible part of the spectrum is in our experience subdivided in different colors; from violet and blue colors that have short wavelengths in the beginning of the spectrum to reddish colors that have longer wavelengths at the end of the visible spectrum.

In addition to the limited capacity of the senses in detecting the physical properties of the outside world and the limited capacity of our brain to process incoming sensory information, our perception of the outside world is also limited to the way we represent these physical properties of the outside world. As our senses and our brain translate the physical properties of the outside world into changes in electrical charge across cell membranes, our
representation of the outside world does not resemble but corresponds to the underlying physical properties in the outside world; we don’t directly experience vibrations of air when we hear something, nor do we directly experience electromagnetic radiation at certain wavelengths when we see colors.

**Visual Perception**

For humans the dominant sense through which we experience the world is vision. Probably as a direct consequence of this, it were mostly observations in the visual modality that slowly made us realize that our brain not only translates the incoming information from our senses, but also actively interprets this incoming information. In the 20th century the Gestalt psychologists like Wertheimer and Koffka proposed that our perceptual experience is shaped by organizational principles that for instance allow us to perceive elements as a group when they are close together or when they are similar and perceive wholes by filling in missing information. These and many other principles are implemented in the anatomical organization of our brains, and structure our perception without us being aware of their influence. A nice illustration of the strength of these organizational principles in the visual system comes from the observation that the brain automatically assigns a more or less consistent local color to an object under different illumination conditions. In the checkerboard illusion by Edward Adelson (Fig. 2A) this color constancy effect is exploited by the presence of a shadow, in which our visual system infers that in an area where there is a shadow a grey surface must represent a lighter surface compared to an identical gray surface in an area where there is no shadow. We therefore perceive areas A and B as having different colors, but if you remove the context of the shadow you see that the
color of areas A and B are actually identical (Fig. 2B). Another example in which it is evident that our brain automatically interprets incoming information is the Ponzo illusion. In this visual illusion the two green lines of identical size appear to be of different sizes when placed over parallel lines that converge as they recede into the distance. Because the convergence of the parallel black lines gives us the illusion of depth, our brain uses these depth cues in interpreting the scene and therefore infers that the vertical line that is more near must be smaller than the vertical line that is farther away, while in fact these lines are exactly the same size (Fig. 2C).

Figure 2 | Organizational principles of our visual system. a, The checkerboard illusion of Edward Adelson demonstrating color constancy. b, The checkerboard illusion with the context of the shadow removed. c, The Ponzo illusion that produces misjudgment of relative line length.

These examples illustrate that although perception begins in the senses by providing sensory information from the outside world, perception truly takes place in the brain. In order to represent and understand the environment our brain therefore actively shapes incoming sensory
information and thereby actively constructs our perception of the world. This interpretation of sensory information is implemented in the organization of our sensory systems and occurs, in part, outside our conscious awareness. It is important to realize however that these assumptions and inferences of the brain are based on and shaped by regularities and natural laws in the environment in which the sensory system evolved; shadows make surfaces look darker than they actually are and objects that are farther away look smaller than they actually are. From these natural laws the brain infers that a surface is lighter when it is covered by a shadow (Fig. 2A-B), and that an object is bigger when it is far away (Fig. 2C). In interacting directly with the environment our brains’ assumptions and inferences would be right, thereby allowing us to faithfully represent the outside world. In this sense these illusions simply illustrate how sophisticated and well adapted the neural machinery underlying our perception really is.

The Visual System

From retina to cortex

The translation of electromagnetic waves from the outside world into changes in electrical charge across cell membranes takes place in the retina, which is a layer of light sensitive cells at the back of the eye. These so called photoreceptors only absorb photons from a small region of the visual scene, and this region of the visual scene to which a cell is responsive is commonly referred to as the classical receptive field of the cell. The visual information leaves the retina through the axons of the ganglion cells, which bundle together to form the optic nerve. At the optic chiasm the axonal fibers from the optic nerve partially cross over to the other side of the brain (Fig. 3),
thereby allowing the lateral geniculate nucleus (LGN) in the thalamus to receive visual information from the same, although contralateral, visual field.

Figure 3 | Anatomy of the human visual system. The changes in electrical charge across cell membranes in the retina as a consequence of light entering the eye first travel through the optic nerves. At the optic chiasm some fiber bundles partially cross over into the optic chiasm, and then travel via the optic tracts to the lateral geniculate nucleus (LGN). From the LGN, the signals continue to the primary visual cortex, from which point further visual processing takes place along the visual cortical hierarchy. Figure adapted from Gazzaniga et al. Cognitive Neuroscience: The biology of the mind, 2nd edition. W.W. Norton and Company (2002).

The receptive fields of neurons in the LGN\textsuperscript{2} are similar to the receptive fields of the ganglion cells in the retina\textsuperscript{3}, and are divided in two categories.
depending on the properties of their receptive fields. So called ON cells are activated when small spots of light on the retina are directed to the center of their receptive field and their surrounding regions are dark, OFF cells on the other hand have the opposite pattern and are activated when small spots of light on the retina are directed to the surroundings of their receptive field and their center region is dark (Fig. 4A). The visual information from the LGN is send to the primary visual cortex (V1) in the back of the brain, and in this area the outputs from the LGN neurons that are selective for small points of light at specific locations in the visual field are combined and transformed into orientation selective representations⁴ (Fig. 4B). Although V1 cells have been reported to respond to color⁵,⁶ or spatial frequency⁷,⁸, most neurons in V1 are therefore most optimally activated by a stimulus inside their receptive field that has a specific orientation (Fig. 4C).
Figure 4 | Receptive fields in the earliest stages of the visual system. **a,** An illustration of receptive fields in the retina and the LGN. Receptive fields in the retina and the LGN are concentric and have a positive center and a negative surround, or vice versa. **b,** An illustration of receptive fields in V1. Receptive fields in V1 are elongated and have a positive center and a negative surround, or vice versa. **c,** Example of a simple connection circuit by which an elongated receptive field in V1 that is sensitive to stimuli of a specific orientation can be made up by aligning three concentric receptive fields in the LGN. Figure adapted from Kandel et al. Principles of Neuroscience, 4 edition. McGraw-Hill (2000).
Chapter 1

The functional organization of neurons in V1 is highly structured, as neurons that share similar properties are grouped together in so called columns. These columns are perpendicular to the surface of the cortex and consist of 6 separate layers that contain neurons with similar orientation preferences. Together with the retinotopic organization of these columns, in which neighboring columns in the cortex represent neighboring locations in the outside world, this columnar organization in V1 preserves the spatial relations between stimuli in the visual scene. At this level the structured functional organization of the visual system therefore constructed a highly detailed map of contrast differences, thereby allowing us to represent differences in light patterns in the world around us.

The hierarchically organized visual cortex

Representation of contrast differences in a highly detailed map of the visual scene is not sufficient to account for our visual experience of the world around us. We segregate contrast differences belonging to individual objects from the background, we recognize these objects and we figure out how they relate to each other to enable visually guided action and navigation. During perception behaviorally relevant information has to be flexibly selected in order to enhance its representation in the brain. In some cases, behaviorally relevant information has to be remembered over a short period of time, thereby making it necessary to store and manipulate these representations in the absence of sensory information.

In the primate brain, these higher level functions of visual processing are implemented in a hierarchically organized system that extends from V1 to the prefrontal cortex (PFC). This hierarchically organized visual system
Introduction

consists of two anatomically and functionally separated, but also heavily interconnected, parallel processing pathways (Fig. 5). The so called ‘what’ pathway along the ventral areas of the brain supports object identification, while the pathway along the dorsal areas of the brain gives rise to the so called ‘where’ and ‘how’ pathways\(^\text{10}\) that support spatial perception and visually guided actions\(^\text{11,12}\), respectively. In both these systems it is thought that the computational principle that Hubel and Wiesel uncovered in response transformations from retina to V1 is repeated at each level of the cortical hierarchy, in which combinations of simpler inputs from earlier stages results into progressively more complex representations along the cortical hierarchy.

![Image of hierarchically organized visual system](image)

**Figure 5** | The hierarchically organized visual system: Visual information is processed in the cortex via two anatomically and functionally separated parallel processing pathways. Both these pathways begin in V1 and extend through a ventral pathway into the temporal lobe and through a dorsal pathway into the parietal cortex and prefrontal cortex. Initially, visual information flows from lower to higher areas through feedforward connections (represented by the blue arrows). After the initial
Chapter 1

feedforward sweep, feedback connections (represented by the red arrows) provide descending top-down influences that mediate recurrent processing. AIP, anterior intraparietal area; IT, inferior temporal area; LIP, lateral intraparietal area; MIP, medial intraparietal area; MST, medial superior temporal area; MT, medial temporal area; PMd, dorsal premotor area; PMv, ventral premotor area; TeO, tectum opticum.

Figure from Charles D. Gilbert & Wu Li. Top-down influences on visual control. Nature Reviews Neuroscience 14, 350-363. (2013).

In the ventral pathway that includes the temporal areas (Fig. 5) and supports object identification, outputs from orientation and spatial frequency tuned cells in V1 and V2 are combined to form selectivity for complex curvature in V4\(^{13,14}\). Although some cells in V4 are known to be selective for color\(^{15}\) or spatial frequency\(^{16}\), the sensitivity of V4 cells to shapes of intermediate complexity is well explained by linear pooling of local orientation responses from V1 and V2 cells. V4 is therefore thought to be an intermediate stage in creating the observed selectivity for complex three dimensional forms in inferior temporal (IT) cortex. In IT, nearby cells are functionally clustered and selectively activated by particular object categories like houses, bodies, fruits or faces\(^{17,18}\). The capacity to recognize faces for instance is supported by millions of face-selective cells in IT. These face cells are organized into a network of highly interconnected face areas in the temporal lobes of the brain\(^{19,20}\) (Fig. 6a) that are specifically involved in processing particular facial dimensions. By combining these facial features into progressively more complex representations along the cortical hierarchy, cells in AM that exhibit selectivity for face identity in a view-invariant manner are generated (Fig. 6b).
Introduction

A)

Figure 6: Complex form representation along ventral pathway. a, Visual areas V1, V2 and V4 (blue) and the face patches in the temporal lobe (orange) in the macaque b, The receptive field size of cells as well as the complexity of the cell’s receptive field properties progressively increase along the ventral stream. Cells in V1 represent the visual scene by line segments of specific orientations. The output of V1 is thought to be combined in V4 to form cells that are selective to complex curvatures, thereby in some cases already showing some rudimentary geometrical outlines of object features like eyes, noses. All the way up in the visual hierarchy cells only respond to specific objects like faces or houses. Within the face patch system cells in PL are thought to selectively respond to specific facial features. Along the face patch system these features are combined to form representations that become increasingly invariant and identity tuned. Cells in ML for instance are tuned to one head orientation, while cells in AM are almost invariant to head orientation.
Chapter 1

Much less is known about the response transformations along the dorsal pathway that includes the parietal areas (Fig. 5). Several studies however showed that this system is involved in detecting and analyzing movements\textsuperscript{21}, the perception and interpretation of spatial relationships and multisensory integration\textsuperscript{22}. Although several studies suggest that there are also representations of objects in the dorsal stream\textsuperscript{23}, these object representations probably do not give rise to object recognition as is the case in the ventral pathway, but instead support spatial perception and visually guided actions\textsuperscript{24}.

*Feedforward processing*

When light hits the photoreceptors in our retina the visual information first propagates along the visual hierarchy through feedforward connections, the so called feedforward sweep\textsuperscript{25}. This initial phase of activity along the visual cortical hierarchy is very fast, with response latencies of around only 50 ms for neurons in the early visual areas and around 100 ms for neurons in the highest areas of the visual hierarchy\textsuperscript{26}. During the feedforward sweep the axons of each single neuron project to thousands of other neurons, thereby allowing an enormous amount of patterns to be generated by combining these outputs at neurons in higher levels of the visual cortical hierarchy. This computational principle is repeated throughout all levels of the cortical hierarchy, and it is during the feedforward sweep that visual representations in lower areas are combined in higher areas to shape the receptive field properties of neurons. By combining and pooling outputs this way, transformations along the visual cortical hierarchy take place in the way visual information is represented. Not only does combining outputs underlie the progressive increase in receptive field complexity that was explained earlier, but the receptive field size of neurons also progressively enlarges.
Introduction

along the visual cortical hierarchy by pooling outputs from lower visual areas\(^4\). Higher areas therefore integrate stimuli over a greater spatial extent, which for instance is essential for size-invariant object recognition in the ventral pathway\(^{17,27}\).

Recurrent processing

In addition to the feedforward sweep, there is also a phase of information processing that is conveyed in the opposite direction\(^{28}\). Feedforward and feedback connections have a different origin and termination profile in the different layers of the cortex; feedforward projections originate mainly in layers 2 and 3 and terminate in layer 4, whereas feedback projections originate primarily in superficial and deep layers and terminate in layer 1 and 5\(^{29}\). During recurrent processing horizontal connections within areas and feedback connection between areas modulate neural responses, which take effect during and after the feedforward sweep when information travels from lower to higher visual areas. The functional role of these recurrent connections is distinguishable from feedforward connections in a very important way. Feedforward processing is based on fixed anatomical connections within and between cortical areas that shape the receptive field properties of neurons. On the other hand, horizontal and feedback connections modulate the activity of neurons dependent on the context in which its receptive field is embedded\(^{30,31,32}\) or the behavioral relevance of the sensory information that is presented at the receptive field of the neuron\(^{33,34}\). Furthermore, these top-down influences do not only enhance the representations of behaviorally relevant information, but also alter the tuning properties of neurons to carry more information about stimulus components that are relevant to the task at hand\(^{35,36}\). This way recurrent processing thus
flexibly influences sensory representations in lower areas, thereby serving many important functions during visual perception. Contextual effects that are implemented by feedback connections for instance enhance specific features of sensory input that match patterns that are embedded in the weights of the excitatory feedback connections. This allows our brain to establish a stable neural assembly that is consistent with the input pattern, thereby enabling us to impose an interpretation based on previous experience on incomplete, ambiguous or noisy sensory information. Furthermore, by enhancing sensory representations in lower areas, top-down influences are thought to be essential for visual awareness, spatial, feature- and object-based attention, working memory, perceptual learning, expectation and figure-ground segregation. This way the brain would be able to dynamically select information from the feedforward flow that is behaviorally relevant for us.

In this thesis we recorded brain activity from neurons at different stages of the macaque visual system while the monkey performed complex visual tasks. Several studies have shown that the anatomical and functional organization (Self et al., Plos Biol, in press) of the visual system of the macaque monkey is very similar to that of the visual system of humans. Furthermore, as macaque monkeys are able to learn complex visual tasks, research in these non-human primates is essential in addressing and answering questions related to the neural implementation of cognitive functions like visual awareness, attention and working memory. By recording neural activity in the macaque visual system we addressed fundamental questions related to the organizational and computational principles that underlie our perception of the visual world. How does the brain processes sensory information that is picked up by our eyes to create visual awareness? How do cognitive processes like working memory influence neural representations of the outside world?
Introduction

How is the segregation of an object from its background implemented in the visual system, and what are the neural mechanisms that underlie this fundamental visual operation?

Chapter 2 addresses the question of how the brain gives rise to conscious experience. Surprisingly, a visual stimulus that is not consciously perceived and therefore remains invisible can nevertheless still affect behavior and elicits brain activity at multiple levels of the visual cortical hierarchy. How awareness emerges in the brain and what the differences are in neural representation at different levels of the visual cortical hierarchy between a sensory stimulus that enters awareness and a stimulus that does not is still heavily debated.

To investigate how activity propagates along the visual cortical hierarchy during the emergence of awareness we compared information transfer during the feedforward sweep in V1, V4 and the dorsolateral PFC using both natural stimuli and electrically induced V1 phosphenes, which is an artificial percept of light at the location of the receptive field when a neuron is electrically stimulated. In almost all studies that investigate the neural correlates of awareness, sensory input is held constant while the perception of the presented stimulus varies. This way the neuronal activity that is elicited by the sensory input can be compared between perceived and identical unperceived stimuli, and neural responses that correlate with basic sensory processing can be distinguished from those that correlate with perception. In our paradigm we presented a brief sensory stimulus of low intensity. Because the stimulus is very hard to perceive, this leads to a situation in which the stimulus sometimes enters awareness and sometimes does not, and thereby allows us to directly compare brain activity between perceived and identical unperceived stimuli that are close to the threshold of perception.
Chapter 1

Using this paradigm, we will address the following questions: Where in the visual hierarchy does the information about unperceived stimuli get lost? Do stronger stimuli penetrate deeper into the visual cortical hierarchy before they are lost than weaker stimuli? Which processes cause the variation in signal strength elicited by a fixed stimulus?

Chapter 3 investigates how sensory information is maintained in the brain when we have to remember something for a short period of time. This ability to store and manipulate information over short periods of time in the absence of sensory stimulation is called working memory. Several studies found that at the higher levels in the visual cortical hierarchy some neurons that represent the spatial location of the presented stimulus remain active throughout the period of time that the location of the stimulus has to be remembered. As the absence or presence of this memory trace predicts whether the monkey remembers the location of the visual stimulus, this shows that this internally sustained neural firing in the absence of sensory stimulation is the neuronal substrate for working memory.

An outstanding question however is what neural mechanism underlies this internally sustained activity. Neurons contain two major classes of channels by which they receive excitatory input, AMPA and NMDA channels. A pharmacological study that investigated the contribution of different glutamate channels during a figure-ground segregation task showed that AMPA channels contribute to the feedforward response but do not contribute to figure-ground modulation in V1 during the sustained response. NMDA channels on the other hand contribute to the enhancement of activity during the sustained response, and do not contribute to the feedforward response. Based on this observation some people that assume that feedback processes are involved during working memory, argue that NMDA receptors are crucial
for internally sustained activity. Another popular idea, that is based mostly on the kinetics and characteristics of the NMDA channel and on related modeling work, is that intrinsic dynamics of single neurons in higher cortical regions can account for the internally sustained activity that underlies working memory\textsuperscript{50}. In contrast to other glutamate channels like AMPA, NMDA channels have slow kinetics and contain a voltage-dependent magnesium block that is relieved only once a cell is depolarized. This way, it is argued that the slow decay time of NMDA channels produces a stable and robust working memory network, while the voltage dependent property of these receptors would serve as a gate for persistent activity\textsuperscript{51}.

By studying the roles of AMPA and NMDA channels during the internally sustained activity that is observed in dorsolateral PFC neurons, we directly test this idea of selective involvement of specific glutamate receptors in maintaining task relevant information. To this end we pharmacologically manipulated the contribution of AMPA and NMDA channels during a working memory task in which the monkey has to remember the spatial location of a visual stimulus, while simultaneously recording activity from single neurons in the dorsolateral PFC.

\textbf{Chapters 4 and 5} investigate how our brain segregates objects from the background in a visual scene. The visual scene consists of many complex objects (figures), which must be segregated from their background and recognized. To correctly segregate these figures from the background the visual system uses information from across the visual scene. As this requires large receptive fields it was generally thought that this process was performed by cells in areas that are high up in the visual cortical hierarchy, but previous studies have shown that even cells in V1 modulate their firing-rate depending on whether their receptive field falls on the figure or the background, so called
Chapter 1

Figure-ground modulation\textsuperscript{52}. The stimuli that are presented during these studies are controlled in such a way that the visual input in the neuron’s receptive field is identical for figure and background stimuli, and therefore these contextual effects must occur through recurrent processing in which information from larger parts of the visual field can be incorporated. Several experimental observations support this. First, the time course of figure-ground modulation is consistent with recurrent processing; while the initial feedforward response is the same for figure and background stimuli, it is only during the later sustained response after 100 ms that the neuron enhances its response when its receptive field is on a figure compared to when its receptive field is on the background. Second, figure-ground modulation is absent during anesthesia when recurrent processing is heavily reduced\textsuperscript{53}. Third, the pharmacological study\textsuperscript{49} that dissociated the contribution of different glutamate channels to the feedforward and the sustained response, and established its subsequent differential effects on figure-ground modulation furthermore supports the importance of recurrent processing in figure-ground segregation.

Although these observations therefore show that recurrent processes is involved in figure-ground modulation, it is still not known whether this modulation is due to feedback from higher visual areas or arises from purely local horizontal interactions between neurons within V1\textsuperscript{54}. In chapters 4 and 5 we will shed light on this open question. In chapter 4 we furthermore investigate whether figure-ground modulation in V1 and V4 is caused by an enhanced response to the figure, a suppressed response to the background, or a combination of both. To assess this, we compared texture-defined figure-ground displays to homogeneous textures that lack a figure-ground organization. By using complex figures during a figure segregation task we will in chapter 5 go deeper into the question of whether figure-ground
Introduction

modulation is due to feedback from higher visual areas or arises from purely local horizontal interactions between neurons within V1. Complex figures enable us to control for local differences in visual input by surrounding the input to the receptive field of the neuron with edges that are matched for figure and background stimuli up to several visual degrees outside the receptive field of the neuron. If figure-ground modulation is still present in V1 with these configurations, local computations within V1 cannot account for this modulation because the horizontal connectivity in V1 is not sufficient to implement these distant contextual effects\textsuperscript{55,56}. In that case figure-ground segregation must be an evolving process, which requires interactions between neurons at multiple levels of the visual hierarchy.
References


12. Nauhaus I, Nielsen KJ, Disney AA, Callaway EM. Orthogonal micro-


52. Lamme AF. The Neurophysiology of Figure-Ground Segregation in Primary Visual Cortex. J. Neuroscience (1995).

53. Lamme VA, Zipser K, Spekreijse H. Figure-ground activity in primary visual cortex is suppressed by anesthesia. Proc Natl Acad Sci U S A
Chapter 1


Chapter 2

At the threshold of conscious perception: information loss and response bias in visual and frontal cortex

Bram van Vugt, Bruno Dagnino, Devavrat Vartak, Houman Safaai, Stefano Panzeri, Stanislas Dehaene & Pieter R. Roelfsema

This chapter is under review at Science
Abstract

How a visual stimulus gives rise to a conscious experience is an important, yet unresolved question in neuroscience. It was proposed that sensory stimuli first propagate to higher cortical areas, where recurrent interactions give rise to “ignition”, a hypothetical process causing awareness by making information about the stimulus broadly available in the brain\textsuperscript{1,2}. However, the mechanisms that determine whether a stimulus reaches awareness are not yet well understood. Here we investigate the respective roles of visual areas V1 and V4 and of dorsolateral prefrontal cortex (dLPFC), in monkeys trained to report their perception of weak stimuli. We find that the efficiency of propagation of stimuli from visual to frontal cortex varies across trials. Stimuli only enter awareness on trials in which they elicit a minimum level of activity in dLPFC, but remain subliminal if dLPFC activation is weaker. We find that the level of these failures of propagation to frontal cortex depends on stimulus strength; weaker stimuli are lost in lower areas than stronger ones. We examine corticocortical propagation by inducing phosphenes, illusory percepts of light, with electrical microstimulation in V1 and find that perceived V1-stimuli elicit stronger V4 activity than missed V1-stimuli. To investigate the cause of the variability of conscious perception, we examine neuronal and behavioral markers of pre-stimulus brain-state and reveal separable influences on the animal’s response bias and sensitivity. Our results can be explained by a relatively simple model in which stimuli become conscious when they can cause ignition in higher cortical areas. They provide a starting point for the study of interactions between brain areas that are responsible for conscious access.
Understanding how awareness emerges in the brain is one of the major challenges that remain to be addressed in neuroscience. One approach to investigate how stimuli enter awareness uses very brief or weak stimuli, which are sometimes perceived and sometimes not so that neuronal activity can be compared between these perceptual states. The classical model that describes how weak stimuli are perceived or missed is the Signal Detection Theory (SDT)\textsuperscript{3}. It posits that stimuli elicit a stochastic signal, which has to reach a threshold for perception (Fig. 1a). Stimuli that fail to reach the threshold will be missed. In the absence of a stimulus, the signal usually stays below the threshold (correct rejection), but it may cross the threshold on occasion, giving rise to a false alarm. SDT also explains variations in the degree conservativeness in responding, because a higher threshold decreases the number of false alarms but also increases the number of misses.

The SDT does not specify the processes in the brain that determine the stochasticity of the stimulus-induced signal nor the mechanism that determines the threshold. In contrast, the Global Neuronal Workspace theory (GNWT)\textsuperscript{4} proposed that stimuli reach awareness by propagating to the higher levels of the cerebral cortex, where they can lead to ‘ignition’, a state that causes information about a brief stimulus to become sustained and broadcasted back through recurrent interactions between many brain areas\textsuperscript{5,6} (Fig. 1b). According to the GNWT, there are two reasons why a stimulus may fail to enter awareness. First, the propagation of activity to higher levels may be too weak. Second, global ignition may fail, for example if the system is refractory because another stimulus caused ignition or if attention is diverted\textsuperscript{4,7}. Combining insights from SDT and GNWT\textsuperscript{8}, we hypothesized that the threshold for perception might equal the signal strength required for the hypothetical ignition process. Furthermore, it is conceivable that the stochasticity of signal strength might be related to variations in the
propagation of activity from lower to higher cortical levels, but these hypotheses remain to be tested.

We therefore trained monkeys to detect low contrast stimuli (Fig. 1c) and examined activity in areas V1 and V4 of visual cortex and in the dorsolateral prefrontal cortex (dIPFC) in response to these stimuli to address the following questions: (1) Where in the visual hierarchy does the information about subliminal stimuli get lost? (2) What are the neuronal mechanisms underlying the perceptual threshold, and how are they related to the putative ignition process? (3) Which processes cause the variation in signal strength elicited by a fixed stimulus?

The monkeys directed gaze to a fixation point, and on half of the trials we presented a 2° low contrast stimulus in the neurons’ receptive field (RF) for 50ms. After a delay of 450ms the monkey reported the stimulus by making a saccade to its previous location. On the other half of trials, we did not present the stimulus and the monkey made a saccade to a grey circle (the reject dot). Accuracy on such stimulus-absent trials was high (~5-10% of false alarms). We adjusted the contrast on stimulus-present trials with a staircase procedure that kept it close to the threshold of perception, at an accuracy of ~80% (see Methods). The contrast threshold ($\theta_{\text{High}}$; accuracy of 80%) varied with stimulus eccentricity between 2.5 and 7% (Supplementary Fig. 1a-c). To examine perception of very weak stimuli we also defined a second threshold, $\theta_{\text{Low}}$, associated with an accuracy of 40% and categorized stimulus strength into three categories; easy (contrast>$\theta_{\text{High}}$), intermediate ($\theta_{\text{Low}}<$ contrast<$\theta_{\text{High}}$) and difficult (contrast<$\theta_{\text{Low}}$) (Fig. 1d). We recorded multi-unit activity (MUA) from V1, V4 and dIPFC and normalized the activity to the response elicited by a high contrast stimulus (Methods). As expected, stimuli with higher contrasts
elicited more activity than stimuli with lower contrasts in these areas (time-window 0-300ms after stimulus onset, t-tests, all ps<10^{-3}) (Fig. 1e).

**Figure 1 | Visual perception at low contrast.** a, Signal Detection Theory (SDT) holds that stimuli need to cause an internal signal strength larger than a threshold to be perceived. The internal signal strength is stochastic and false alarms result if the signal crosses threshold in the absence of a stimulus. In SDT the subject’s bias can be changed
Chapter 2

by shifting the threshold. A high (low) threshold, makes the subject more (less) conservative in his ‘target present’ judgements. D-prime is a measure of the subject’s sensitivity. It corresponds to the distance between the distributions of signal strength for stimulus-present and stimulus-absent trials, measured in units of the standard deviation. 

b, According to Global Neuronal Workspace Theory (GNWT), sensory activity first needs to be propagated to the higher stages of the cortical hierarchy. If it is strong enough, it can access awareness by causing ‘global ignition’, a process which enables maintenance and sharing of information about the stimulus between cortical processors. 

c, Contrast detection task. On half of the trials, a low contrast stimulus was presented for 50 ms after 300-500ms of fixation. On the other half of the trials there was no stimulus. After a delay of 450ms, the monkey reported the stimulus by making a saccade to its previous location and the absence of the stimulus by making a saccade to a grey circle (reject dot). 

d, Psychometric detection curve for an example session in monkey B. We determined two thresholds, \( \theta_{\text{low}} \) (accuracy of 40%) and \( \theta_{\text{high}} \) (accuracy of 80%), based on the psychometric function. 

e, Responses elicited in V1, V4 and dlPFC by easy (contrast>\( \theta_{\text{high}} \)), intermediate (\( \theta_{\text{low}} < \text{contrast}<\theta_{\text{high}} \)) and difficult stimuli (contrast<\( \theta_{\text{low}} \)). In V1 the number of sites \( \{N_{\text{Easy}}, N_{\text{Intermed}}, N_{\text{Diff}}\} \) was \([33, 25, 23]\), in V4 \([36, 34, 29]\) and in dlPFC \([17, 20, 14]\). The vertical scale bar is in units of normalized activity, because the MUA-signal at all sites was normalized to the response elicited by a high contrast stimulus (Methods and Supplementary Fig. 8). 

Within each category, we compared neuronal activity between Seen and Miss-trials with identical stimulus contrast (see Methods). Seen stimuli elicited stronger activity in V1, V4 and dlPFC than missed stimuli, at every difficulty level (Fig. 2a, see Supplementary Fig. 2 for example recording sites) (window 0-300ms, paired t-tests, all areas and categories \( p < 0.01 \)). Hence during Miss-trials, information is lost during the propagation of visual information to higher cortical areas. In dlPFC and, to a lesser extent, in V4 and V1, the extra neuronal activity on Seen-trials was maintained until the saccade (Supplementary Figure 3).
To determine the locus of the information loss, we computed the miss-fraction; the percentage of activity remaining for non-perceived stimuli \( \text{Activity}_{\text{miss}} / \text{Activity}_{\text{seen}} \times 100\% \), time-window 0-300ms) (Fig. 2b). For the difficult stimuli, the miss-fraction was 46% in V1 and 14% in V4, implying a substantial loss of activity before V1 and a further loss between V1 and V4 (significant difference between miss-fractions in V1 and V4, \( t_{46}=2.5, p < 0.01 \), \( N_{V1}=27, N_{V4}=26 \)). For the intermediate and easy stimuli, the miss-fractions were much higher (around 60% and 80% respectively) and did not differ significantly between V1 and V4 (both \( p > 0.05 \)), indicating substantial propagation of neural activity on subliminal trials. Now, however, extra activity was lost between V4 and dlPFC, both for the intermediate (62% vs. 33%, \( t_{50}=3.4, p < 0.05 \), \( N_{V4}=33, N_{dlPFC}=19 \)) and easy stimuli (83% vs. 22%, \( t_{49}=4.0, p < 10^{-3} \), \( N_{V4}=34, N_{dlPFC}=17 \)). Thus, activity elicited by weak stimuli tends to get lost at lower levels than that elicited by stronger stimuli.

Importantly, the activity levels in V1 and V4 on easy Miss-trials (red curves in right panels of Fig. 2a) were at least as strong as those on difficult Seen-trials (green curves in left panels). Hence the neuronal activity level in these areas does not fully predict perception. In dlPFC, however, at all contrast levels the activity elicited by seen stimuli was stronger than that elicited by missed stimuli, implying these neurons signaled awareness in a more categorical manner, which is presumably related to the planning of an eye movement towards the neurons’ RF. These results were consistent between monkeys (Supplementary Fig. 4).
Figure 2 | Activity in V1, V4 and dIPFC in the contrast detection task. 

a, Activity elicited in V1 (upper panels), V4 (middle panels) and dIPFC (lower panels) by contrasts lower than $\theta_{\text{Low}}$ (difficult, left), between $\theta_{\text{Low}}$ and $\theta_{\text{High}}$ (intermediate, middle) and higher than $\theta_{\text{High}}$ (easy, right) for contrast-matched Seen-trials (green curves) and Miss-trials (red). The black curves illustrate activity on trials in which the monkeys correctly reported the absence of a stimulus and the blue curves activity on trials with false alarms.

b, Miss fraction ($\text{Activity}_{\text{Miss}} / \text{Activity}_{\text{Hit}} \times 100\%$) in V1 (blue bars), V4 (yellow) and dIPFC (red) in the different stimulus categories (time-window, 0-300ms after stimulus onset).
Neuronal activity also differed between false alarms and correct rejections. In dLPPC and V4, neuronal activity was higher on false alarm trials (time-window 200 ms before saccade; dLPPC, \( t_{27}=4.5, p < 10^{-3} \); V4, \( t_{36}=4.8, p < 10^{-3} \)), with a trend in the same direction in V1 (\( t_{34}=1.8, p = 0.07 \)). The extra activity on false alarm trials was already present in a 300ms time window before stimulus onset in all areas (all ps < 0.05), suggesting that increased cortical excitation is a pre-stimulus brain-state marker that predicts false alarms. To compute the reliability of this marker, we computed the area under the receiver-operating curve (AUROC). An AUROC of 0.5 indicates no predictive power and a value of 1 perfect prediction. We obtained AUROCs of 0.53, 0.53, 0.59 in V1, V4 and dLPPC, respectively (MUA; striped green bars in Fig. 3a), and the values increased around the time that a stimulus could have been presented (300ms time-window; striped black bars). On stimulus-present trials, we focused on trials of intermediate difficulty level, for which we had enough trials, and obtained relatively low pre-stimulus AUROCs (V1: 0.52, V4: 0.50, dLPPC: 0.51). The values after stimulus onset, also known as “choice probabilities”, were higher (V1: 0.71, V4: 0.71, dLPPC: 0.74, solid black bars in Fig. 3a).

The finding that some pre-stimulus brain states are more permissive for stimulus detection than others is of great interest. We therefore also evaluated other markers that might predict perceptual outcome, including the diameter of the pupil (Pu), its time-derivative (\( \Delta Pu \))\(^9\),\(^10\), the power in the alpha, beta, and low/high gamma bands of the local field potential\(^11\)–\(^14\) and the time that the monkeys took to initiate a new trial, which is informative about their motivation. When considered individually, all markers gave weak predictions (Fig. 3a). We sought to increase predictive power in distinguishing between hits and misses by linearly combining pre-stimulus brain-state measures into a joint measure \( J \), using a cross-validation method to avoid
overfitting (including sessions with more than 15 hits and misses at the intermediate difficulty level, see Methods). \( J \) predicted perceptual outcome with an accuracy close to 60\% (V1: 0.59, V4: 0.58, dlPFC: 0.59, all ps<0.001).

To examine the influence of \( J \) on neuronal activity, we selected all trials from the highest and lowest quintile of the \( J \)-distribution across trials. Higher \( J \)s were associated with higher pre-stimulus activity and a stronger visual response in all three areas (Fig. 3c). High \( J \) values also caused a slight increase in the false alarm rate, indicating that the monkeys were more prone to report “target present” (Fig. 3b). We computed another joint measure, bias (\( B \)), which was designed to specifically predict false alarms. The AUROC-values for \( B \) were around 0.6 (V1: 0.60, V4: 0.62, dlPFC: 0.61, Fig. 3a). Higher \( B \)-values were associated with extra pre-stimulus firing in all three areas in stimulus-present (not used to define \( B \), Fig. 3d) and stimulus absent-trials (Supplementary Fig. 5). These results, taken together, suggest that higher baseline firing rates increase the probability of false alarms, presumably by making neuronal activity prone to cross the threshold for reporting a stimulus.

According to SDT, accuracy also depends on d-prime, the distance between the stimulus-present and absent distributions of signal strength (Fig. 1a). Our comparison between seen and miss trials (Fig. 2) suggested that there was variability in the propagation of neuronal activity to higher cortical levels. However, \( J \) did not have an isolated effect on d-prime because it also influenced the false-alarm rate. We therefore devised a third linear combination of pre-stimulus brain-state measures to index sensitivity (\( S \)), designed to discriminate hits from misses without influencing the false alarm rate (see Methods, Fig. 3a,b). A high \( S \) increased visually driven activity, especially in the higher areas, in accordance with an influence on the efficiency of activity propagation to higher levels (Fig. 3e). Hence, there are
separable influences of pre-stimulus brain-state on the subject’s response bias and sensitivity. Bias $B$ relates to an increase in spontaneous activity whereas sensitivity $S$ relates to an increase in the efficiency of signal propagation (Fig. 3f).

**Figure 3 | Influence of pre-stimulus brain state on neuronal activity and choice.**

*a*, Behavioral and neurophysiological markers of pre-stimulus brain state that predict
the animal’s choice. Predictive value was quantified as the area under the receiver operator curve (AUROC). Results of experiments in which we recorded activity in V1, V4 and dlPFC are shown in different panels. Positive AUROCs indicate that a higher value of the marker predicts a higher probability of hits or false-alarms. T½, time between the appearance of the fixation point and the moment that the monkey directed gaze to the fixation point; Pu, pupil diameter; ΔPu, change in pupil diameter; MUA, pre-stimulus MUA; α, power from 5-15 Hz; β, 15-25 Hz; γ<sub>λ</sub>, 25-40 Hz; γ<sub>h</sub>, 40-80 Hz. Joint, combination of markers best distinguishing between hits and misses (J), correct rejections and false alarms (B) and a measure that discriminates between hits and misses with a minimal influence on the false alarm rate (S) (see schematic in panel b). C.P., choice probability based on stimulus-driven MUA. Filled bars, seen trials vs. misses. Striped bars, false alarms vs. correct rejections. *, P<0.05, **, P<0.01, ***, P<0.001, one-tailed t-test with Bonferroni correction. b, J, B and S were based on linear combinations of pre-stimulus brain state markers. Lower panels, influence of J, B and S on the probability of reporting stimulus-present on no-stimulus trials (N.S. – false alarms) and stimulus-present trials (hits) at the three difficult levels (lowest and highest quintiles). c, Neuronal activity (smoothed with a 40ms window) on stimulus-present trials within the highest quintile (lighter colors) and lowest quintile (darker colors) of the distribution of J. The shaded regions indicate +/- s.e.m. as determined with bootstrapping. d,e, Activity on stimulus present trials within higher and lower quintiles of B (d) and S (e) during the pre-stimulus epoch. Note that B was defined based on stimulus-absent trials but is applied here to sort stimulus-present-trials. f, High values of B, which increase the false alarm rate, are associated with higher firing rates throughout the trial, causing neurons to be closer to the threshold of ignition. High values of S are associated with more efficient propagation of neuronal activity to higher processing levels, increasing the separation between the distributions of signal strength on target-present and target-absent trials.

Visual information reaches higher areas through multiple routes, some of which bypass V1. To specifically investigate the role of V1-to-V4 propagation and its failures around the threshold of perception, we activated V1 with electrical microstimulation, while recording from V4. The monkeys
now reported a phosphene, an illusory light percept at the RF of the stimulated neurons\textsuperscript{16}, elicited in V1 with 5 pulses (200Hz) while we varied stimulation strength with a staircase procedure (Fig. 4a). As before, we determined two thresholds (Fig. 4b). $\theta_{\text{high}}$ was $17\pm4\mu$A and $40\pm19\mu$A (mean ± s.d.) for monkeys B and C, respectively (Supplementary Fig. 1d). We recorded MUA in V4 from neurons with RFs that overlapped with those of the stimulated V1 neurons (Fig. 4c, right) for a total of 84 V1-V4 pairs (58 in monkey B and 26 in C). V1 microstimulation elicited V4 activity with a temporal profile that resembled the response elicited by a visual stimulus and higher V1 currents increased the V4 response (all $p$s < 0.05) (Fig. 4c; Supplementary Fig. 6). When we compared the activity between Seen-trials and Miss-trials with the same microstimulation current, we found that V4 activity was larger on Seen-trials than on Miss-trials at all current strengths (time window from 0-150ms after stimulus onset; paired t-test, all $p$s < 10\textsuperscript{-6}) (Fig. 4d; Supplementary Fig. 6c). Hence, the efficiency of activity propagation from V1 to V4 is correlated to perception. The miss-fraction increased from 16\% in difficult trials to 42\% in intermediate trials and to 83\% in easy trials (Fig. 4e) (t-tests, all $p$s < 10\textsuperscript{-3}, $N_{\text{Low}}$=47, $N_{\text{Intermediate}}$=41, $N_{\text{High}}$=67), implying that information about the stronger currents was lost at processing levels higher than V4. Indeed, V4 activity on easy Miss-trials was stronger than on difficult Seen-trials, confirming that although correlated to perception, the amplitude of the early V4 response does not fully predict if a stimulus will reach awareness.
Figure 4 | Phosphene perception and GNWT model. a, Phosphene detection task. On half of the trials a train of five microstimulation pulses was delivered to V1 to evoke a phosphene at the RF of the stimulated cells (green rectangle). The other half of the trials
were without microstimulation. After 480ms the monkey reported a phosphene by making an eye movement to its previous location and its absence with an eye movement to the grey circle. 

b, Accuracy as function of current amplitude in an example recording session in monkey B. c, V4 population activity on correct trials during V1 microstimulation with different current intensities. Right, RFs of stimulated V1 neurons (white rectangle) and a V4 recording site (false color). d, Average V4 response elicited by currents below $\theta_{\text{low}}$ (difficult, left), between $\theta_{\text{low}}$ and $\theta_{\text{high}}$ (intermediate, middle) and higher than $\theta_{\text{high}}$ (easy, right). Green curves, Seen-trials. Red curves, Miss-trials. The square in the left panel illustrates the microstimulation epoch (5 pulses with an interval of 5ms). e, Miss fraction in V4 ($\text{Activity}_{\text{Miss}}/\text{Activity}_{\text{Seen}} \times 100\%$) for the different performance categories (time window 0-150ms). f, Structure of the GNWT model. The visual stimulus activated the LGN and feedforward connections (FF) propagated activity from the visual cortex (areas V1 and V4) to parietal and frontal cortex. Self-connections were within the areas and feedback connections (FB) propagated activity from higher back to lower areas. g, Probability of target-present response as function of stimulus strength (a.u., arbitrary units). h, Miss fractions are lower at higher levels, indicating that more activity is lost on the miss-trials. i, Activity elicited in model V1, V4 and frontal cortex (compare to Fig. 2a). The black curves illustrate activity for correct rejections and the blue curves activity for false alarms.

We investigated how well a simple mathematical model of hierarchically arranged areas and top-down feedback could reproduce these findings. The model’s architecture was based on previous modeling studies$^{17,18}$ and contained the lateral geniculate nucleus and four hierarchically arranged cortical areas (Fig. 4f). We represented the population of neurons in each area with a single, stochastic Ornstein-Uhlenbeck process so that five variables described the evolution of the network state during simulated trials, one for each brain region. The model contained feedforward connections, self-connections within the areas and feedback connections. The self-connections and feedback connections in V1 and V4 of the model were relatively weak so that the role of these areas was to propagate activity to the
higher levels. The self-connections and feedback connections were stronger in frontal to parietal cortex so that the dynamics in these regions were slower\textsuperscript{19} and a brief strong input (50ms duration) could cause self-sustained activity after the stimulus disappeared\textsuperscript{20}. Simulation of the dynamics revealed a threshold for self-sustained activity (“ignition”) and if it was reached we assumed that the model had signaled the stimulus. If the input arriving from visual cortex was weaker than this threshold, activity decreased back to baseline levels and we assumed that no stimulus was detected. Although remarkably simple, the model accounted for our main findings. It produced a realistic psychometric function with increased accuracy for higher contrasts (Fig. 4g). Furthermore, the activity of the model units was remarkably similar to that recorded in the monkeys (Fig. 4i and Supplementary Fig. 7; compare to Fig. 2) and, interestingly, weak stimuli tended to get lost at lower levels than stronger stimuli (Supplementary Fig. 7d). Once ignition occurred at the higher levels, the feedback connections from frontal and parietal regions to the visual cortex caused a small increase in activity in V1 and V4 after the stimulus had disappeared, just as in the data (Fig. 2). Finally, the model also accounted for the profile of neuronal activity on trials without a stimulus. In a fraction of these trials, stochastic fluctuations in activity caused spontaneous ignitions at variable time points, giving rise to false alarms. When averaged across trials, time-locked to trial onset, these spontaneous ignitions are visible as a ramping of activity, just as in the dLPFC of monkeys (Fig. 2a). On these trials, feedback to lower levels also caused slightly higher activity levels in V1 and V4 than on trials with correct rejections, just as observed in the monkey visual cortex (Fig. 2a). Importantly, the model confirmed our conjecture of multiple bottlenecks for conscious access: weak stimuli tend to get lost at early processing levels, while stronger stimuli transiently activate frontal cortex but fail to reach the threshold for ignition.
Information processing at the threshold of conscious perception

The present results provide new insights into how stimuli reach awareness and inspire unification of SDT and GNWT. Conscious visual perception requires stimuli to reach a threshold level in higher cortical areas that is sufficient for “ignition”, which corresponds to a categorical decision about the presence of a stimulus. Stimuli that fail to reach this level remain subliminal, whereas stimuli crossing the threshold induce a self-sustained pattern of neuronal activity at the higher processing levels (the global workspace), which corresponds to a working memory of the stimulus. False alarms occur when the stochastic neuronal activity reaches the threshold for ignition in the absence of a stimulus. We found that V1 activity was weaker on miss trials, which is in accordance with previous work, but differs from results in the primary somatosensory cortex of monkeys where neuronal responses do not predict the perception of weak tactile stimuli. Yet, V1 and V4 also did not fully predict perception because neuronal activity evoked by higher contrast stimuli that were missed was at least as strong as that elicited by low contrast stimuli that were seen, implying information loss in downstream areas. In contrast, the activity level in dlPFC categorically predicted perception, implying that dlPFC lies at or beyond the stage that determines the threshold for perception. Our experiments were not aimed at revealing all brain regions that participate in the maintenance of the working memory trace and may be part of the global neuronal workspace. We focused on neurons in dlPFC that contribute to eye movement planning and our design did not dissociate brain regions necessary for awareness from those involved in reporting about a visual stimulus, which is difficult to achieve in experimental animals. Other cortical regions, upstream from dlPFC and including, for example, the temporal and parietal cortex, also exhibit extra activity if a stimulus reaches awareness, suggesting that they take part in conscious perception. The representations in these upstream areas are likely
to be effector independent so that they can also support other behavioral responses like, for example, a key press in response to the stimulus. We found that the higher brain regions provided feedback to V1 and V4, where neuronal activity was slightly stronger on trials where the monkey planned an eye movement towards the RF. A previous fMRI study observed a comparable increase of neuronal activity in early visual cortex when subjects reported a stimulus but the fMRI signal did not differentiate between hits and false-alarms, as if it was blind to stimulus presentation. The difference with the present results may be caused by the nature of the fMRI signal, which is sensitive to processes other than spiking activity, such as synaptic activity of feedback connections.

We found that it is possible to combine markers of pre-stimulus brain-state to improve the prediction whether a stimulus will reach awareness or not, with accuracies up to 65%. Additional measures of pre-stimulus brain state may further increase predictive power, although part of the unpredictability may be caused by the intrinsic stochasticity of neuronal activity, which was an essential ingredient of the model (Fig. 4f). It proved possible to define linear combinations of pre-stimulus brain-state markers with independent information about the subject’s response bias and sensitivity. A bias to report target present was associated with a higher baseline firing rate across different brain regions, causing neurons to be closer to the threshold for ignition (Fig. 3). In contrast, a higher sensitivity was associated with an improved propagation of neuronal activity to higher processing levels, thereby increasing the difference in activity levels between target-present and absent trials at the processing stage that determines the threshold for ignition. Although our study did not address the brain mechanisms that influence the pre-stimulus firing rate and the quality of signal propagation, previous studies have established relations between pupil
size and the frequency bands of the EEG and the tone of neuromodulators such as noradrenaline\textsuperscript{9} and acetylcholine\textsuperscript{30}. Future studies could examine whether the activity of these neuromodulatory systems indeed exert separable influences on the animals’ bias and sensitivity when stimuli are near the threshold of conscious perception.
Chapter 2

References


Information processing at the threshold of conscious perception


35. Logothetis, N. K., Pauls, J., Augath, M., Trinath, T. & Oeltermann, A. Neurophysiological investigation of the basis of the fMRI signal. (2001).


37. Xing, D., Yeh, C.-I. & Shapley, R. M. Spatial spread of the local field


Methods

Surgeries and RF mapping

Five male rhesus monkeys participated in this study. We carried out V1 and V4 recordings in monkeys B and D, microstimulation experiments in monkeys B and C and dlPFC recordings in monkeys E and J. In a first operation, a head holder was implanted. In a separate surgery, arrays of 4x5 or 5x5 electrodes (Blackrock Inc.) with a thickness of 80μm and a length of 1 or 1.5mm were chronically implanted in areas V1 and V4 in monkeys B, C and D. These electrodes are most likely positioned in layers 4 and 5. For the dlPFC recordings we performed a craniotomy (stereotaxic coordinates: 21mm anterior, and 17mm lateral) and implanted a titanium chamber (Crist Instruments) for electrophysiological recordings. All surgical procedures were performed under aseptic conditions and general anesthesia and complied with the US National Institutes of Health Guidelines for the Care and Use of Laboratory Animals and were approved by the Institutional Animal Care and Use Committee of the Royal Netherlands Academy of Arts and Sciences. Further details of the surgical procedures and the postoperative care have been described elsewhere31.
We measured the RF dimensions of every V1 recording site by determining the onset and offset of the response to a slowly moving light bar for each of four movement directions. V4 RFs were mapped by presenting white squares (1°x1°) on a grey background at different positions of a grid (1° spacing). We measured RFs of the dlPFC neurons with a delayed saccade task, presenting a visual stimulus (white circle of 2° diameter) at one of 8 locations (at 8° eccentricity) to determine the preferred direction. After 150ms the visual cue was extinguished but the monkey maintained fixation for another 350ms before the fixation point was extinguished, cueing the monkey to make a memory guided eye-movement into a target-window (4° diameter) centered on the previous visual stimulus. Subsequently, we mapped the neurons' preferred eccentricity in 4° steps. The RFs of all recorded neurons in dlPFC were in the contralateral hemifield. They were classified as having a visual response (26 out of 39 neurons) when the signal to noise ratio (SNR) of this response, defined as the ratio between the visually driven activity (time-window 0-300ms after stimulus onset) and the standard deviation of spontaneous activity was higher than 2. Similarly, they were classified as having a saccadic response (13 out of 39 neurons) when the SNR of activity at the time of saccade onset was higher than 2.

**Behavioral setup**

The monkeys performed the tasks while seated in front of a 21-inch CRT monitor with a refresh rate of 70 Hz and a resolution of 1024x768 pixels. The eye position was monitored with a video-based eye tracker (Thomas Recording) and sampled at 250Hz. A trial was initiated when the monkey had maintained his gaze for 300ms within a fixation window, 1.5° in diameter,
centered on the fixation point. The monkey obtained a juice reward at the end of each correct trial.

**Contrast detection task**

We used a forced-choice, delayed-saccade task while we varied the contrast of the stimulus according to a staircase procedure (Fig. 1c). A trial started with a fixation point in the center of the screen (size of 0.3° of visual angle) and a reject dot in the periphery (0.3° size). We randomly selected 47.5% of the trials as visual stimulus trials. The duration of the fixation period was randomized between 300-500ms, according to a uniform distribution (except in four sessions with monkey B with a fixed 300ms fixation epoch). We then presented a 2° visual stimulus for 50ms at the RF location. The other 47.5% of trials were no-stimulus trials, and 5% were used to gauge the visual responsiveness of the neurons for normalization of activity (see below). After an additional delay of 500ms, the fixation dot became blue, cuing the monkeys to make a saccade. The monkeys reported the presence of the stimulus by making a saccade to its previous location and the absence of a stimulus by making a saccade to the reject dot. They obtained a juice reward after correct trials. We adjusted the contrast of the stimulus with a 3 up/1 down staircase procedure (we used contrast steps between 0.4 and 1%) to ensure that the contrast was near the threshold of visibility. The luminance of the background of the monitor was 3.9 cd/m² during recordings in monkey E, 9.4 cd/m² during recordings in monkeys B and D and 4.9 cd/m² during recordings in monkey J. We report Weber contrast: (luminance stimulus–luminance background)/luminance background. We recorded MUA at a total of 35 recording sites in V1 during (20 sessions; 9 sites in 8 sessions in monkey B; 26 sites in 12 sessions in monkey D), a total of 37 recording sites in V4 during 23
recording sessions (19 sites in 7 sessions in monkey B, 18 sites in 16 sessions in monkey D) and a total of 28 recording sites in dlPFC (15 in monkey E and 13 in monkey J). In the analysis of pre-stimulus brain state markers we included sessions with a minimal number of 15 miss and 15 hit trials at the intermediate difficulty level.

The mean $\theta_{\text{High}}$ (80% accuracy) across recording sessions during the V1 recordings was $3.5 \pm 0.3\%$ (mean$\pm$s.d.) for monkey B and $5.1 \pm 1.3\%$ for monkey D (Supplementary Fig. 1a). The mean $\theta_{\text{High}}$ during the V4 recordings for monkey B was $4.3 \pm 0.8\%$ and $3.4 \pm 0.3\%$ for monkey D (Supplementary Fig. 1b). The mean $\theta_{\text{High}}$ during the dlPFC recordings was $6.9 \pm 1.2\%$ for monkey E and $2.7 \pm 0.5\%$ for monkey J (Supplementary Fig. 1c). The mean false alarm rate during the V1 recordings was $5.0 \pm 3.3\%$ for monkey B and $8.9 \pm 2.6\%$ for monkey D. The mean false alarm rate during the V4 recordings was $5.3 \pm 2.4\%$ for monkey B and $3.6 \pm 1.8\%$ for monkey D. Finally, the mean false alarm rate during the dlPFC recordings was $13.9 \pm 10.2\%$ for monkey E and $3.6 \pm 3.5\%$ for monkey J.

To estimate the visual responsiveness of the V1 and V4 neurons we presented a high contrast homogeneous texture in 5% of the trials. On these trials, the monkeys maintained fixation and we used the activity elicited by the texture for normalization. Neurons in dlPFC did not respond well to homogeneous textures. To estimate visual responsiveness in dlPFC we therefore presented a $2^\circ$ yellow square in 5% of trials, which was the target of an eye movement. We normalized the MUA (e.g. in Figs. 1-4) so that the mean level of activity before stimulus onset was 0 and the peak response elicited by the texture (or yellow square for the dlPFC recordings) was 1 (Supplementary Fig. 8). This normalization procedure maintains all activity differences between stimulus conditions and between trials with hits, misses, false alarms and correct rejections.
Phosphene detection task

The phosphene detection task had a similar structure as the visual detection task, but now the monkeys had to report a phosphene elicited by a train of microstimulation pulses applied to area V1 (Fig. 4a). The trial started with a fixation epoch of 300ms. On 47.5% of the trials we applied five negative-first biphasic pulses of 400\(\mu\text{s}\) duration (200\(\mu\text{s}\) per phase) at a frequency of 200Hz, through one of the V1 electrodes using a custom-made two-channel constant current stimulator. An adjacent electrode on the same array was used for current return. The close proximity of the current source and sink in V1 decreases the magnitude of the stimulation artifact in V4. The other 47.5% of trials were stimulus absent trials without microstimulation. In both conditions the monkeys maintained fixation during a delay of 500ms before the fixation dot became blue, cuing the monkeys to make a saccade. This additional delay is advantageous, because it excludes reflexive saccades that might be elicited by the activation of motor structures like the superior colliculus\(^{33}\). The monkeys reported the phosphene by making a saccade to its location and the absence of a phosphene by making a saccade to the reject dot (Fig. 4a). We used a 3 up/1 down staircase procedure with a spacing between current amplitudes of 5\(\mu\text{A}\). In 5% of the trials we presented a homogeneous texture stimulus to measure the visual responsiveness of neurons at the V4 recording sites and we used the response amplitude for normalization. In previous work\(^{34}\), we found that the V4-response to V1 microstimulation was reduced or even abolished if V4 was simultaneously driven by visual stimulus, implying substantial overlap between the pathways by which a visual stimulus reaches V4 and the V1 microstimulation effect (Figure S4 in ref. \(^{34}\)). Furthermore, V1 stimulation elicited stronger activity in V4 recording sites that overlapped more with the receptive field of the stimulated V1 neurons.
We tested different combinations of V1 stimulation electrodes with multiple V4 recording electrodes across sessions. The mean threshold for phosphene detection across sessions was 17±4 μA for monkey B and 40±19 μA for monkey C (Supplementary Fig. 1d). The mean false alarm rate across recording sessions was 10±6% (mean±s.d.) for monkey B and 10±4% for monkey C.

**Data acquisition and artifact removal**

We recorded MUA in areas V1, V4 and dlPFC as the envelope of the signal filtered between 500 and 5,000Hz, as in previous studies\(^{35-37}\). Specifically, we made recordings with TDT (Tucker Davis Technology) recording equipment using a high-impedance headstage (RA16AC) and a preamplifier (either RA16SD or PZ2) and we sampled the data at a rate of 24.4 kHz. We band-pass filtered the signal (500Hz-5kHz), full-wave rectified it and used a low-pass filter (200Hz) to produce an envelope of the multi-unit activity (MUA). This MUA signal provides an average of spiking activity of a number of neurons in the vicinity of the tip of the electrode and the population response obtained with this method is therefore expected to be identical to the population response obtained by pooling across many single units\(^{32,35-37}\).

To analyze activity in V4 elicited by microstimulation of V1, we developed an offline procedure to remove the microstimulation artifact\(^{34}\). We synchronized the timing of the microstimulation pulses to the clock of the data acquisition system, ensuring that data was always sampled at identical time points relative to the microstimulation pulses. We computed the average shape of the stimulation artifact at each recording site and subtracted the artifacts from the raw signal (Supplementary Fig. 9a,b). We then band-pass filtered and
rectified the signal (Supplementary Fig. 9c, as described above), and removed a period of 1ms (24 samples) centered on each pulse from the signal to remove possible remnants of the artifacts (Supplementary Fig. 9d). We used linear interpolation to fill in the missing samples and low-pass filtered the signal to compute the MUA (Supplementary Fig. 9e). As control, we applied the same procedure to trials without microstimulation, we found that it did not influence the shape or amplitude of the MUA signal.

**Data analysis**

We computed the SNR of neurons at every V1 and V4 recording site as the ratio between the peak response elicited by the full screen texture and the standard deviation of the spontaneous activity level across trials. We only included recording sites with a SNR larger than 1 in the analysis and excluded recording sites with a drift in the signal (22 out of 220 recording sites in V1 and 25 of 340 sites in V4). In the contrast detection experiments, we ensured that the neurons’ receptive fields fell on the visual stimulus, and for the MS experiments we only included V4 recording sites with a RF that overlapped with the RF of the stimulated V1 neurons. We averaged the responses across sessions if a combination of V1 and V4 electrodes was tested in multiple sessions, so that every combination of V1 and V4 recording sites contributed at most a single data point to the statistics.

For each recording site, we calculated the average response per condition after subtracting the mean baseline activity (300ms before stimulus onset). We used the peak visual response (0-150ms after stimulus onset) that was evoked by a high contrast texture stimulus (or yellow square) for normalization (for example recording sites see Supplementary Fig. 8).
In the comparison of seen and miss trials, we only included contrasts for which we collected at least three Miss- and Seen-trials. We defined three categories for the stimulus intensities (easy, intermediate and difficult) by defining two thresholds, $\theta_{\text{High}}$ (80% correct) and $\theta_{\text{Low}}$ (40% correct) based on the psychometric curve in every recording session. When multiple stimulus intensities fell in the same category, we first calculated the average response on Seen and Miss trials per intensity before computing a weighted average across intensities, where the number of missed trials per intensity determined the weighting, thereby ensuring that differences in the distribution of intensities between Seen and Miss trials did not invalidate the comparison. We computed the miss-fraction as $\text{Activity}_{\text{miss}}/\text{Activity}_{\text{seen}} \times 100\%$ in a time-window from 0-300ms after stimulus onset.

**Pre-stimulus state variables**

We quantified the pre-stimulus brain state in each trial using three behavioral parameters and five neuronal measures. The first behavioral parameter was the time that the monkey took to attain fixation after the appearance of the fixation point, which provides a measure for the animal’s motivation. We also measured the mean pupil size and its mean derivative during a 300 ms pre-stimulus period, because both parameters have been suggested to provide useful measures of brain-state\(^9,10\). Pupil size was computed as the average of horizontal and vertical pupil radii measured by the eye tracker and we also used this average to compute the derivative. As neuronal measures, we included the average MUA across all recording sites in the 300ms pre-stimulus time-window and the pre-stimulus LFP power spectrum in four different bands: alpha (10-15 Hz), beta (15-25 Hz), low gamma (25-40 Hz) and high gamma (40-80 Hz). We computed the power spectrum using wavelet
transform of the LFP signal, using complex Gaussian wavelets of second order, denoted as the ‘cgau2’ wavelet class in Matlab\textsuperscript{38}. We computed the single trial power spectrum as the average of the power during the 300 ms pre-stimulus window across all recording sites. To ensure that the pre-stimulus power spectrum is not influenced by post-stimulus activity, we cut the LFP signal used in the analysis at the stimulus onset and used only the pre-stimulus period. To minimize edge effect problems in the analysis, we used a symmetric extension method\textsuperscript{39} by mirroring the LFP signal from -300-0 ms symmetrically around stimulus onset and using the extended signal for the computation of the wavelet coefficients and power spectra.

**Computations of the area under the ROC-curve (AUROC)**

We used the area under the ROC curve (AUROC) to determine how well two distributions of pre-stimulus brain state measurements allowed us to predict hits vs. misses and false alarms vs. correct rejections, in single trials. If there is no predictive power, the AUROC is 0.5 and if the brain-state measure perfectly predicts the behavioral outcome it has a value of 1. We computed AUROCs for every measure of brain state individually, and defined it so that the AUROC is larger than 0.5 if a higher value of the parameter predicts more hits or false-alarms. In the following definitions, we will use $x$ to denote a single variable or a linear combination of them.

To compute the AUROC between two sets of data $x_1 \in X_1$ and $x_2 \in X_2$ with probability densities $f_1$ (e.g. hit trials) and $f_2$ (e.g. miss trials), we first computed the empirical true positive TP and false positive rates FP, defined as
If the two distributions are Gaussian, the area under ROC (AUROC) is positive if and only if the mean of $X_1$ is larger than the mean of $X_2$. The choice of the ‘positive’ class only affects whether the AUROC is larger or smaller than 0.5. For each variable, we determined the sign after considering the AUROC values of the entire dataset, thereby ensuring that we only report effects that are stable and consistent across sessions and animals. We computed the AUROC using the \textit{polyarea} function of Matlab.

### Linear combinations of pre-stimulus brain-state measures

We computed three different linear combinations of the state variables, $J$, $B$ and $S$, as explained in the main text. In this analysis, we only included sessions with a minimum of 15 hit and 15 misses at the intermediate difficulty level (V1: N=15; V4, N=8; dlPFC, N=14). The first joint measure, $J$, optimized the AUROC for predicting whether the monkey would have a hit or miss on a stimulus-present trial. For each trial, we considered the joint linear combination

$$J = \sum_{i=1}^{8} w_i M_i$$

(2)

where $M_i$ is pre-stimulus brain state measure $i$, and $w_i$ are the weights, which were constrained to have a sign consistent with the AUROC sign of each individual brain-state measure. For example, because the AUROC of the alpha band in the seen vs. miss comparison was larger than 0.5, we constrained the
weight to be positive in the joint measure. Because we had to combine measures with different units and ranges, we first z-scored them and we also z-scored the resulting joint measures.

To compute the optimal weights while avoiding overfitting, we implemented a leave-one-out cross-validation procedure. For the stimulus present trials, we sampled an equal number of hit and miss trials from each luminance level to assure that the contrast values of the hit and miss trials were balanced. For each trial, we considered all other trials as the training set and we computed the optimal weights with a genetic algorithm, which is suitable for finding the global maximum of the AUROC\(^40\). We then used the weights to compute the joint measure \((J, B \text{ or } S)\) of the trial that had been left out (in 100 repeats, sampling the remaining trials with replacement). We averaged across repeats to determine the average joint measure per trial and considered its distribution across trials for the analysis. As a control, we also used a 2-fold cross-validation method, training the model with half of the trials and testing with the other half. This method gave similar results as the leave-one-out procedure.

In addition to \(J\), which maximized the difference between hits and misses, we computed \(B\) (bias), which maximally discriminated between false alarm and correct rejection trials. The third measure, \(S\), simultaneously maximizes the difference between hits and misses while ensuring that the measure does not influence the false alarm rate on stimulus-absent trials. There is no closed form solution to determine \(S\), and we therefore used Pareto optimal solutions, for which one of the objective functions (AUROC values of the miss-seen comparison and false alarm-correct rejection comparison) cannot improve without changing the other objective function. We selected all the Pareto solutions for which the AUROC is below the null AUROC value obtained after
shuffling the false alarm and correct rejection trial labels. We chose the Pareto solution with the highest AUROC.

**Statistical tests**

We used two-tailed unpaired and paired t-tests. For the tests involving multiple-comparisons, we used the Holm-Bonferroni correction of the $p$-value.

**Model description**

The model consisted of the lateral geniculate nucleus (LGN) and four cortical areas. For every contrast level, we ran 1000 trials in which the stimulus was presented to the LGN for 50ms. The membrane potential $m_A$ in area $A$ was updated according to

$$\Delta m_A = \frac{dt}{\tau}(inp_A(t) - m_A(t - 1)),$$

where $inp_A$ is the input into the area, $\tau$ is the time constant (here we used $\tau=0.04s$) and $dt$ the time step of the simulation ($dt=10^{-3}s$). The input into an area was defined as

$$inp_A = \sum_B r_B w_{B,A} + n_A,$$

where $r_B$ is the firing rate in area $B$, $w_{B,A}$ is the connection from area $B$ to $A$ (Table 1) and $n_A$ is a noise term. The external input to the model was injected into the LGN for 50ms with a strength of 0, 4, 6, 8, 10, 13 or 16. The noise $n_A$ was distributed according to an Ornstein-Uhlenbeck process:

$$\Delta n_A = \frac{dt}{\tau}(-n_A(t - 1) + g\xi).$$
where $g$ determines the noise level ($g=10$) and $\xi$ is distributed according to a standard normal distribution. The firing rate $r_A$ in area $A$ depended on the membrane potential $m_A$ as follows:

$$r_A = 1/(1 + \exp[-(m_A - \theta_A)]),$$  \hspace{1cm} (6)

where $\theta_A$ determines the spontaneous activity level of area $A$ ($\theta_{LGN} = 2.5$, $\theta_{V1} = 3.5$; $\theta_{V4} = 4.0$, $\theta_{Parietal} = 4.2$ and $\theta_{Frontal} = 4.2$).

<table>
<thead>
<tr>
<th></th>
<th>LGN</th>
<th>V1</th>
<th>V4</th>
<th>Parietal</th>
<th>Frontal</th>
</tr>
</thead>
<tbody>
<tr>
<td>LGN</td>
<td>0</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V1</td>
<td>0.25</td>
<td>0.3</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>V4</td>
<td>0.25</td>
<td>3</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parietal</td>
<td>0.25</td>
<td>0.25</td>
<td>2</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>Frontal</td>
<td>0.25</td>
<td>0.25</td>
<td>8</td>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>

**Table 1.** Strength of connections. Black, feedforward connections; green, self-connections; red, feedback connections.
Supplementary Figures

**Figure S1 | Contrast and phosphene thresholds.** a-c, Distribution of contrast thresholds $\theta_{\text{th}}$ (accuracy of 80%) across recording sessions in V1, V4 and dLIFC. d, Distribution of phosphene thresholds (accuracy of 80%) across recording sessions in monkeys B and C.
Figure S2 | Neuronal activity at example recording sites in V1, V4 and dlPFC. 

Activity elicited by seen (green curves) and missed stimuli (red curves) of equal contrast at example recording sites in V1 and (a) V4 (b) in monkey D and dlPFC in monkey J (c) with contrasts $<\theta_{\text{Low}}$ (difficult, left panels), between $\theta_{\text{Low}}$ and $\theta_{\text{High}}$ (intermediate, middle) and higher than $\theta_{\text{High}}$ (easy, right). The black curves represent activity on trials in which the monkey correctly reported the absence of a stimulus and the blue curves activity on false alarm trials.
Figure S3 | Activity aligned on saccade onset in the visual detection task. a, Activity in V1 (left), V4 (middle) and dIPFC (right) aligned on saccade onset was stronger for seen-trials (green curves) than that during contrast-matched miss-trials (red) (time window 200-0ms before saccade; V1, t_{34}=2.9, p < 10^{-3}; V4, t_{36}=5.2, p < 10^{-3} and dIPFC, t_{27}=4.5, p < 10^{-3}). Responses elicited during trials with correct rejections and false alarms are shown in black and blue, respectively. b, Difference in activity between false alarms and correct rejections in V1 (blue curve), V4 (orange) and dIPFC (red), aligned on stimulus onset. The increase in activity on false alarm trials was significant in V4 and dIPFC (time window 200-0ms before saccade; dIPFC, t_{27}=4.5, p < 10^{-3}; V4, t_{36}=4.8, p < 10^{-3}). There was a trend in the same direction in V1 (t_{34}=1.8, p = 0.07).
Figure S4 | Data of individual monkeys in the visual detection task. 

- **a**, Activity in V1 in monkey D (upper panels) and monkey B (lower panels).
- **b**, Activity in V4 in monkey D (upper panels) and monkey B (lower panels).
- **c**, Activity in dIPFC in monkey E (upper panels) and monkey J (lower panels).
Figure S5 | Activity on trials without a stimulus sorted by bias B. Activity in V1, V4 and dLPFC on trials with the highest (red) and lowest quintile of B (blue). Shaded regions indicate s.e.m. as determined by bootstrapping. The vertical lines indicate the average time point of stimulus appearance on stimulus-present trials.
**Figure S6 | V1 microstimulation induced activity in V4.**

**a.** V4 activity induced by electrical microstimulation of V1 neurons. The V1 and V4 neurons had overlapping receptive fields.

**b.** MUA elicited at an example V4 recording site by V1 microstimulation on single trials. Note that V4 activity is elicited on every trial. The open square below the plot illustrates the microstimulation epoch (5 pulses with an interval of 5ms).

**c.** Mean evoked response (time window 0-150ms after stimulus onset) for seen-trials (x-axis) and miss-trials (y-axis) across V4 recording sites in the different performance categories. Current strengths were identical on these seen and miss trials. Note that a larger fraction of the activity is lost in V4 on miss trials, in particular on trials with lower currents.
Figure S7 | Model activity as function of stimulus strength. a, Model structure. The visual stimulus activates the LGN and then propagates through visual cortex (areas V1
and V4) to parietal and frontal cortex. Feedforward connections (FF) propagated activity from lower to higher levels, self-connections within the areas and feedback connections (FB) from higher to lower areas. b, Probability of Seen-trials as function of stimulus strength (a.u., arbitrary units). c, Activity elicited in model V1 (upper panels), V4 (middle panels) and frontal cortex (lower panels) by stimuli with an amplitude of 4 (left), 8 (middle) and 16 a.u. (right). The black curves illustrate activity for correct rejections and the blue curves activity for false alarms. d, Miss fraction (ActivityMiss/ActivitySeen x 100%) in model V1 (blue bars), V4 (yellow) and frontal cortex (red) for the same stimulus strengths as in c.
Figure S8 | Responses to high contrast stimuli used for normalization of the activity. Activity in V1 (a) and V4 (b) elicited at example recording sites in response to of a full screen, high contrast homogeneous texture (gray curves). In dIPFC (c) we presented salient $2^\circ$ yellow square. Green curves, seen trials. Red curves, missed trials from the contrast detection task. We used the activity elicited by the high contrast stimuli for normalization of the MUA.
Figure S9 | Artifact removal. **a**, Example experiment where we applied microstimulation to V1 and recorded in V4. The stimulation artifacts are clearly visible in the broadband signal at five example V4 recording sites (different colors). Arrows illustrate the timing of the five pulses separated by 5 ms. Inset: average shape of the artifact at one of the sites. **b**, We subtracted the average shape of the artifact at every recording site and in every trial. **c**, A high pass filtering step (500 Hz cut off) was applied to remove the low-frequency LFP signals. This was followed by rectification (negative samples become positive). Note the remnants of the artifacts. **d**, We removed samples for a duration of 1 ms centered on the pulses, and used linear interpolation to replace these samples. **e**, A low pass filter (200 Hz cut off) was applied to compute the MUA signal. Note that the microstimulation artifacts have been completely removed.
Chapter 3

The contribution of AMPA and NMDA receptors to persistent firing in the dlPFC during working memory delays

Bram van Vugt, Timo van Kerkoerle, Devevrat Vartak, Pieter R. Roelfsema

This chapter has been submitted to Journal of Neuroscience
Abstract

Many tasks demand that information is kept online for a few seconds before it is used to guide behavior. The information is kept in working memory as the persistent firing of neurons encoding the memorized information. The neural mechanisms responsible for persistent activity are not yet well understood. Theories attribute an important role to ionotropic glutamate receptors and it has been suggested that NMDA receptors (NMDA-Rs) are particularly important for persistent firing, because they exhibit long time constants. Ionotropic AMPA receptors (AMPA-R's) have shorter time-constants and have been suggested to play a smaller role in working memory.

Here we compared the contribution of AMPA-Rs and NMDA-Rs to persistent firing in the dorsolateral prefrontal cortex (dlPFC) of macaque monkeys performing a delayed saccade to a memorized spatial location. We used iontophoresis to eject small amounts of glutamate receptor antagonists, aiming to perturb but not abolish neuronal activity. We found that both AMPA-Rs and NMDA-Rs contributed to persistent activity. Blockers of the NMDA-Rs decreased persistent firing associated with the memory of the neuron’s preferred spatial location but had comparatively little effect on the representation of the anti-preferred location. They therefore decreased the information conveyed by persistent firing about the memorized location. In contrast, AMPA-R blockers decreased activity elicited by the memory of both the preferred and anti-preferred location, with a smaller effect on the information conveyed by persistent activity. Our results provide new insights into the contribution of AMPA-Rs and NMDA-Rs to persistent activity during working memory tasks.
Introduction

Working memory refers to the ability to store and manipulate information over short periods of time, on the order of seconds. In many situations, we have to briefly remember what we perceived and we then store this information in working memory while in other situations working memories are retrieved from long-term memory. The ability to store and manipulate information is crucial for cognition in daily life and a deeper understanding of its neural basis would be of great medical and social significance because disorders such as schizophrenia and Alzheimer’s disease degrade the quality of working memory.

We can maintain memories of stimuli in any sensory modality including visual, tactile and auditory stimuli. Many previous studies focused on the maintenance of visual information. They revealed neuronal correlates for the memorization of multiple visual features, including motion, color, shape and stimulus location. A common task used to probe spatial working memory is the oculomotor delayed-response (ODR) task (Figure 1A), in which subjects keep a location in working memory in order to make a saccade to it at the end of the trial. Several studies found that the firing of so called ‘delay cells’ in the dorsolateral prefrontal cortex (dLFC) of the macaque monkey represents a spatially specific memory trace. They are activated by a visual cue in their receptive field (RF) and remain active during memory delays when the visual cue is extinguished. Delay cells are intermingled with visual cells, which are activated by a visual stimulus but return to baseline when the stimulus is no longer visible.

The mechanisms underlying persistent activity are only partially understood. On the one hand, it may involve reverberatory excitation between neurons within or between cortical areas. On the other hand, it may rely on specific membrane conductances that cause sustained excitation of
individual neurons induced, for example, by the activation of acetylcholine receptors\textsuperscript{23-25}, dopamine receptors\textsuperscript{26} and noradrenaline receptors\textsuperscript{27,28}. Several studies have also implicated NMDA-Rs in working memory\textsuperscript{29,30}. These receptors have a long time constant, which is important for stable persistent activity\textsuperscript{29,31}.

Two events need to occur before NMDA-Rs pass current. Glutamate needs to bind, but the neuron also has to be depolarized to release magnesium, which blocks the channel at resting membrane potentials\textsuperscript{32,33}. In the visual cortex, this gating of NMDA-Rs by membrane depolarization causes them to influence neuronal firing rates multiplicatively, with strong effects on neurons that are well driven by a stimulus and smaller effects for weakly activated cells\textsuperscript{34,35}. In contrast, AMPA-Rs always depolarize the postsynaptic neurons, in an additive manner. It is likely that the gating of NMDA-channels also has consequences for persistent activity in higher brain regions, such as frontal cortex. Sensory input to the neurons might release the magnesium block by activating AMPA-Rs so that currents can flow through NMDA-channels, keeping the neurons sufficiently depolarized and thereby causing persistent activity when the stimulus has disappeared.

In an elegant study, Wang et al.\textsuperscript{36} tested the role of glutamate receptors in persistent firing in the macaque dlPFC. They found that NMDA-R antagonists almost abolished persistent activity, whereas the effect of AMPA-R antagonists was weaker, in support of the specific role of NMDA-Rs in working memory. However, NMDA-R blockers had stronger effects in all epochs, making it difficult to rule out that the effects were caused by differences in efficacy of AMPA-R and NMDA-R blockers.

In the present study, we directly compared the contribution of AMPA-Rs and NMDA-Rs to neuronal activity in dlPFC in a working memory task. We used microiontophoresis with low ejection currents to perturb neuronal

76
activity without abolishing it, so that we could directly compare the contributions of these receptors in different epochs of the task. We report that AMPA-Rs and NMDA-Rs make comparable contributions to persistent activity. However, the contribution of NMDA-Rs is strongest for the preferred stimulus of a cell, in accordance with their multiplicative effect on neuronal firing rates.

**Materials and Methods**

**Surgical procedures**

All procedures complied with the NIH Guide for Care and Use of Laboratory Animals (National Institutes of Health, Bethesda, Maryland) and were approved by the institutional animal care and use committee of the Royal Netherlands Academy of Arts and Sciences.

We recorded neural activity from the dorsolateral prefrontal cortex (frontal eye fields and surrounding cortex on the convexity) of three adult macaque monkeys (Macaca Mulatta: monkeys B, J and E). During surgeries, general anesthesia was induced with ketamine (15 mg/kg injected intramuscularly) and maintained after intubation by ventilation with a mixture of 70% N₂O and 30% O₂, supplemented with 0.8% isoflurane, fentanyl (0.005 mg/kg intravenously) and midazolam (0.5 mg/kg/h intravenously). In a first surgery, the monkeys were implanted with a head post for head stabilization. The monkeys were then trained on the ODR task until they could reliably perform the task. In a second surgery, we performed a craniotomy (centered on stereotaxic coordinates: 21mm anterior, and 17mm lateral) and implanted a titanium chamber (Crist Instruments) for electrophysiological recordings and the iontophoretical administration of APV and CNQX. After implantation, the location of the arcuate and principal sulci relative to the recording chamber were determined using ultrasound imaging (Figure 1C), and the frontal eye fields (FEF) were located with electrical microstimulation (Figure 1C).
Chapter 3

Behavioral task

Monkeys B, J and E were first trained on the ODR task (Figure 1A). A fixation point (a red circle of 0.3° diameter) was presented on a grey background and the monkey started the trial by directing gaze to a 1.5° diameter fixation window centered on the fixation point. After 300 ms of fixation a visual cue (white circle of 2° diameter) was presented at either the neurons’ receptive field or the anti-preferred location (opposite sign for the x- and y-coordinates). After 150 ms the visual cue was extinguished but the monkey had to maintain fixation for another 1000 ms before the fixation point was extinguished, which indicated to the monkey that he was required to make a memory guided eye-movement into a target-window (4 degrees diameter) that was centered on the location of the previous visual cue. Correct responses were rewarded with apple juice. Trials in which the animal broke fixation before the fixation point was extinguished were aborted, and stimulus conditions were presented in a pseudorandom order. All stimuli were generated using in-house software (Tracker) and presented on a CRT monitor with a resolution of 1024x768 pixels and refresh rate of 85Hz, which was viewed from a distance of 40 cm. Eye movements were recorded with a video eye-tracker (Thomas recordings) with a sampling rate of 350 Hz.
Figure 1 | Microiontophoresis in dLPFC during an oculomotor delayed response task. a, After a 300 ms fixation epoch, a white spatial cue was presented for 150 ms and the monkey maintained fixation for another 1000 ms. After the fixation point disappeared, the monkey made a saccade to the memorized location. b, Three barrel glass electrode used for microiontophoresis. Scale bar, 100 μm. c, Coronal (left panel) and sagittal (middle panel) ultrasound images of the recording chamber of monkey B. White arrow points to the arcuate sulcus, grey arrows to the principal sulcus. Location of the different slices relative to the arcuate sulcus (white line) and principal sulcus (grey line) are indicated in the schematic drawing in the right panel. White scale bars, 10mm. d, Example wave shape of a well-isolated single unit during the block of trials before drug delivery (black), during drug delivery (red) and in the recovery period (gray).
Electrophysiology and iontophoresis

We recorded single units with tungsten-in-glass electrodes fused with two side barrels \(^{37}\) (Figure 1D) that were used for iontophoretic drug administration by applying a small electric current to a tungsten wire that was inserted into these side barrels. The impedances of the measuring electrodes ranged from 400 kOhm to 2 MOhm (median ~1 MOhm) and the impedance of the ejection barrels from 15 to 150 MOhm (median ~20 MOhm).

The signal from the recording electrode was recorded with Tucker Davis Technology (TDT) equipment using a high-impedance headstage (RA16AC) and a preamplifier (RA16SD) with a hardware high-pass filter of 2.2Hz, a low-pass filter of 7.5 kHz (-3dB point) and sampled with a rate of 24.4kHz. Spikes were initially determined by setting a voltage threshold. If necessary, spike sorting was done offline using waveclus software \(^{38}\).

For iontophoresis, we dissolved the NMDA-R antagonist 2-amino-5-phosphonovalerate (APV) (Sigma-Aldrich) or the AMPAR antagonist 6-cyano7-nitroquinoxaline-2,3-dione (CNQX) (Sigma-Aldrich) at 0.02 M in triple-distilled water (pH ~8.0). APV and CNQX are negatively charged, and we retained them in the glass-pipettes by delivering a positive potential (+15nA for APV and +20nA for CNQX) and ejected them by delivering a negative potential. The ejection currents were set to the amount needed for a noticeable difference in the spiking activity recorded while the monkey performed the ODR task. We adjusted the current in order to perturb but not abolish the activity. For APV, ejection currents ranged from -2nA to -7nA for monkey B and from -5nA to -15nA for monkey J. For CNQX, ejection currents ranged from -10nA to -20nA in both monkeys (J and E).
**Contribution of AMPA-Rs and NMDA-Rs during working memory**

**RF mapping**

RF’s were measured using the same ODR task that was used during the recordings. First the preferred and anti-preferred direction was determined using 8 locations at 8° eccentricity. The eccentricity was subsequently mapped in 4° steps for the preferred and anti-preferred direction only. Most of the RF’s of the recorded single units were at 18° eccentricity for monkey B, at 13° eccentricity for monkey J and at 18° eccentricity for monkey E.

**Data acquisition**

We determined the location of the arcuate sulcus with ultrasound imaging and recorded single unit activity anterior to this sulcus (Figure 1C). A blunt guide tube, made to tightly fit around the probe, was rigidly attached to a microdrive for mechanical stability (Narishige group). We pre-dimpled the dura with the guide tube and electrode (~1mm), penetrated the dura with the electrode, and pulled back the guide tube and electrode to un-dimple the dura. The electrode was left to settle for about 20 minutes. The probe was then carefully advanced until a single unit was encountered. After stabilizing the recording of the spiking activity of the single unit, we determined its RF properties with the ODR task. We only selected isolated single units with spatial tuning for further recording, and most of these neurons (45 out of 57 for the APV dataset, 32 out of 51 for the CNQX dataset) showed sustained firing during the memory period. For most single units (47 out of 57 for the two APV datasets, 29 out of 48 for the two CNQX datasets), three blocks of ~80 trials were recorded; a recording block of ~80 trials without drug delivery by maintaining the holding current (from now on called “pre-drug recordings”), a recording block of ~80 trials where the drugs was administered by applying the ejection current (“during-drug recordings”) and finally a recording block of ~80 trials without drug delivery, again by maintaining the holding current (“recovery
recordings"). During-drug recordings were started once an effect was noticeable in the spiking activity, usually 3 to 4 minutes after the ejection current was applied and the drugs were applied continuously throughout the recording period. Recovery recordings were started once the effect of drug delivery faded, usually 5-10 minutes after the holding current was applied after drug delivery. The waveforms of the recorded spiking activity during one example recording are shown in Figure 1D. For a small fraction of the recordings (10 out of 57 for the two APV datasets, 13 out of 48 for the two CNQX datasets) we lost the single unit during the waiting period after drug delivery so that we could perform the recovery recording.

Data analyses
All spike data was binned in 10 ms windows. The ODR task was divided into two epochs; spontaneous activity and task-related activity. The spontaneous epoch lasted from 300 ms before stimulus onset up to stimulus onset and the task-related epoch lasted from stimulus onset up to saccade onset. We also evaluated the cue-driven activity in a time-window from 50-250 ms, persistent activity in a time-window from 300-1150 ms after cue onset (starting 150 ms after cue offset) and saccade related activity in a window from 200 ms before saccade onset to the onset of the saccade. To quantify the spatial tuning for each cell individually, we calculated $d'$ for task-related activity as $(\text{mean preferred location} - \text{mean anti-preferred location})/\sqrt{(0.5\times(\text{s.d. preferred location})^2 + (\text{s.d. anti-preferred location})^2)}$, where $\sqrt{}$ is square root and s.d. the standard deviation of the firing rate across trials.

For statistical analysis, we used two-sided t-tests to compare spontaneous and task-related spiking activity between pre-drug, during drug and recovery recordings. A three-way repeated-measures ANOVAs with the factors drug (2 levels), epoch (4 levels) and monkey (2 levels) was used to compare drug
Contribution of AMPA-Rs and NMDA-Rs during working memory effects across time-windows. Results were considered significant if p-values were smaller than 0.05 for both monkeys individually as well as when averaged across monkeys.

We investigated if the influence of APV and CNQX on delay activity in the preferred direction predicted the influence on delay activity in the anti-preferred direction by computing the Pearson correlation coefficient $r$. We determined the significance of the difference between correlation coefficients for APV and CNQX by first carrying out Fischer’s $r$ to $z$ transform and then computing the $z$-value of the difference according to $z_{\text{Difference}} = (z_{\text{CNQX}} - z_{\text{APV}}) / \sqrt{1/(N_{\text{CNQX}} - 3) + 1/(N_{\text{APV}} - 3)}$.

Results

Behavioral effects of blocking glutamate receptors

At the time we started collecting the data, performance for both monkeys in the ODR task was high (99.9% for monkey B, 98% during APV recordings and 99.2% during CNQX recordings in monkey J and 96% for monkey E) (Figure 2A). To elucidate the contribution of the glutamate receptors to persistent activity during memory delays, we iontophoretically administered the NMDA-R antagonist APV or the AMPA-R antagonist CNQX. However, we only applied small dosages to perturb activity without abolishing it. At these dosages, the glutamate receptor antagonists did not have consistent effects on accuracy. Although APV decreased the accuracy of monkey B to 99.4% ($p < 0.005$; t test, $n=33$), accuracy only slightly increased to 99.5% during the recovery block (not significantly different from the APV block; $p > 0.4$, $n=23$) (Figure 2A, left panel) and we cannot exclude the possibility that the decrease in accuracy was caused by a small but systematic decrease in the animal’s motivation over time. APV did not influence the accuracy of monkey J. It was 98% in the pre-drug epoch, 97.9% during APV administration and 97.6% in the recovery
block (all p > 0.4; t test, n=23) (Figure 2A, right panel). Similarly, the AMPA-R antagonist CNQX application did not influence accuracy in monkeys J and E (both monkeys, p > 0.7) (Figure 2B).

**Figure 2 | Influence of APV and CNQX on the monkeys’ accuracy.** a, Accuracy in the ODR task, in the block of trials before APV delivery (black bar), during APV delivery (red bar), and in the recovery period (grey bar), for monkeys B (left panel) monkey J (right panel). * = p < 0.05 b, Accuracy before CNQX delivery (black bar), during CNQX delivery (red bar), and in the recovery period (grey bar), for monkeys J (left panel) monkey E (right panel).

**Effects of blocking NMDA-Rs on neuronal activity in the dlPFC**

To investigate the role of NMDA-Rs in persistent firing, we recorded the activity of single neurons in the dlPFC during the ODR task. We only selected well isolated single units that exhibited spatial selectivity for further
Contribution of AMPA-Rs and NMDA-Rs during working memory

We recorded activity from a total of 56 neurons (33 and 23 neurons in monkeys B and J, respectively) that were held long enough to compare activity before drug application to that during APV administration. We lost the isolation of ten neurons (6 in monkey B and 4 in J) after drug application before the recovery block, but we were able to record data for the other 46 neurons data during the recovery block. Most of the neurons (28 out of 33 for monkey B and 17 out of 23 for monkey J) exhibited persistent firing during the memory period.

Typical example recordings for both monkeys are illustrated in Figure 3A,B, and the population response obtained by averaging across all neurons is shown in Figure 3C,D. The neurons showed elevated firing during the full duration of the trial when the visual cue was presented at the preferred location of their RF, while showing a suppression of spiking activity in response to the presentation of the visual cue at the anti-preferred location of their RF (Figure 3C,D). In both monkeys, APV suppressed baseline activity before visual cue onset compared to pre-drug recordings ($p < 10^{-5}$; t test, $n = 33$ for monkey B, $p < 10^{-3}$; t test, $n = 23$ for monkey J) (Figure 3C,D). Administration of APV suppressed spiking activity in both monkeys at the preferred location during the response elicited by the visual cue (time window 50-250 ms; $p < 10^{-3}$; t test, $n = 33$ for monkey B, $p < 10^{-3}$; t test, $n = 23$ for monkey J), during persistent activity (time window 300-1150 ms; $p < 10^{-5}$; t test, $n = 33$ for monkey B, $p < 10^{-3}$; t test, $n = 23$ for monkey J) and also in the saccade window (time window 200 ms before saccade onset to the onset of the saccade; $p < 10^{-3}$; t test, $n = 33$ for monkey B, $p < 10^{-3}$; t test, $n = 23$ for monkey J) (Figure 3C,D). In both monkeys the suppression was much larger at the preferred location than at the anti-preferred location, and in monkey J this suppression for cue presentation at the anti-preferred location was even absent. Blocking the NMDA-Rs therefore weakened the spatial tuning of the
cells by reducing the difference in spiking activity between the preferred and anti-preferred cue. To measure the spatial tuning we computed d’, which measures how well a single neuron distinguishes between the memory for the two locations in single trials. APV decreased the d’ for most cells in both monkeys (Figure 3E,F) during the memory epoch (300-1150 ms after the onset of the cue, i.e. starting 150 ms after cue offset). In monkey B, the d’ decreased from an average of 2.42 to a value of 1.69 (p < 10^{-4}; paired t-test, n = 33) and in monkey J d’ decreased from 2.17 to 1.64 (p < 10^{-3}; paired t-test, n = 23).

Spiking activity gradually restored to pre-drug levels when drug administration was stopped. Although recovery was not complete in all our recordings, the activity of all the cells changed back into the direction of pre-drug recordings, both for baseline spiking activity before visual cue onset (p < 10^{-5}; t test, n = 54 for monkey B, p < 10^{-3}; t test, n = 38 for monkey J) as well as for spiking activity for the remainder of the trial when the visual cue was presented at the preferred (50-1150 ms after the onset of the cue; p < 10^{-3}; t test, n = 27 for monkey B, p < 0.05; t test, n = 19 for monkey J) and anti-preferred location for monkey B (p < 0.05) (Figure 3G). In monkey J the suppression of spiking activity following APV delivery was absent at the anti-preferred location, and therefore no restoration of spiking activity was observed at the anti-preferred direction for monkey J (p > 0.05) (Figure 3H).
Contribution of AMPA-Rs and NMDA-Rs during working memory

Figure 3 | Effect of APV on neuronal activity during ODR task. a,b, Example single units of monkeys B (A) and J (B), illustrating the effect of APV on neuronal activity. Black
trace, activity in the pre-drug period. Red trace, activity after application of APV. c,d Average activity of neurons of monkey B (c) (N=34) and J monkey J (d) (N=23) before and after drug application. e,f, Abscissa, d’ before APV delivery; ordinate, d’ during APV delivery. Every data point represents a well-isolated neuron. g,h, The effect of APV on population response of the ODR and recovery for (g) monkey B (N=33) and (h) monkey J (N=23). The preferred location (continuous lines) and anti-preferred location (dashed lines), before (black lines) and during APV delivery (red lines). Grey lines illustrate activity in the recovery episode.

To investigate if a change in the variability of the neuronal response across trials contributed to this decrease in the d’ prime we also computed the influence of APV on the Fano-factor. In monkey B, APV did not have a significant effect on the Fano factor during the delay period, both for the preferred direction (p > 0.05; t test, n = 33) and the anti-preferred direction (p > 0.05; t test, n = 33). In monkey J, APV significantly increased the Fano factor during the delay period for the preferred direction (p < 0.01; t test, n = 23), but not for the anti-preferred direction (p > 0.05; t test, n = 23).

To further examine the time-course of the drug effect, we plotted the difference between spiking activity before and during the administration of APV (Figure 4).
Contribution of AMPA-Rs and NMDA-Rs during working memory

Figure 4 | Effect of APV on neuronal activity for the preferred and non-preferred direction. a,b, Absolute difference of spiking activity during the ODR task. We subtracted the activity in the pre-drug period from that during APV delivery, elicited in trials with a cue at the preferred (green trace) and anti-preferred location (red trace). The blue rectangles illustrate the time-windows used for quantification.

For the preferred location we compared the effects of APV during different epochs of the task (spontaneous-, visual-, delay- and saccade activity) for the two monkeys using a three-way repeated-measures ANOVAs with the factors drug (2 levels), epoch (4 levels) and monkey (2 levels). The influence of APV on the firing rate during the cue, memory and saccade epochs was higher than the effect on spontaneous activity ($F_{3,448} = 15.78$, $p < 0.001$). However, the effect on cue-driven, persistent and saccade related activity was similar.

This contribution is in accordance with a general multiplicative effect of NMDA-Rs on spiking activity but appear to be at odds with the hypothesis that NMDA-Rs have a specific role in the generation of persistent activity. We therefore also examined the small subset of neurons with a visual response without delay activity (Figure 5). APV suppressed the visually driven activity of these neurons in both monkeys (cue-epoch, $p < 0.05$; t test, $n = 6$ for
monkey B and $p < 0.05$; t test, $n = 6$ for monkey J), in accordance with a more general role of NMDA-Rs in both cue-driven and persistent activity.

**Figure 5 | Effect of APV on the activity of visual cells.** a,b, Average response of visual neurons without delay activity in monkeys B ($N=7$) (A) and J ($N=7$) (B), elicited in trials with a cue at the preferred (continuous) and anti-preferred location (dashed), before (black traces) and during APV delivery (red traces). c,d, Abscissa, $d'$ before APV delivery. Ordinate, $d'$ during APV delivery. Every data point represents an individual visual cell.
Contribution of AMPA-Rs and NMDA-Rs during working memory

Contribution of AMPA-Rs to activity in dLPFC

We recorded a total of 41 neurons during CNQX application (27 in monkey J and 14 in monkey E), and more than half of them exhibited sustained firing during the memory period (15 out of 27 for monkey J and 14 out of 14 for monkey E). Of these neurons 29 were kept long enough to examine activity in the recovery period, after drug application. Typical example recordings are illustrated in Figure 6A,B, and the population response is shown in Figure 6C,D. In both monkeys, baseline spiking activity before visual cue onset was suppressed during the CNQX administration (p < 10\(^{-3}\); t test, n = 27 for monkey J, p < 0.05; t test, n = 14 for monkey E) (Figure 6C,D). For both monkeys, administration of CNQX suppressed spiking activity during the full duration of the trial when the visual cue was presented at the preferred location, in the cue window (p < 10\(^{-3}\); t test, n = 27 for monkey J, p < 10\(^{-3}\); t test, n = 14 for monkey E), memory window (p < 10\(^{-3}\); t test, n = 27 for monkey J, p < 10\(^{-3}\); t test, n = 14 for monkey E), as well as in the saccade window (p < 10\(^{-3}\); t test, n = 27 for monkey J, p < 0.05; t test, n = 14 for monkey E). When the visual cue was presented at the anti-preferred location there was a significant reduction of activity in the cue window (p < 0.05; t test, n = 27 for monkey J, p < 0.05; t test, n = 14 for monkey E), memory window (p < 0.05; t test, n = 27 for monkey J, p > 0.05; t test, n = 14 for monkey E) and the saccade window (p < 10\(^{-3}\); t test, n = 27 for monkey J, p < 0.05; t test, n = 14 for monkey E) (Figure 6C,D). To examine the influence of CNQX on the tuning, we calculated d’s. In monkey J CNQX caused a decrease in d’ from an average of 1.01 to a value of 0.83 (p < 10\(^{-3}\); t test, n = 27) and in monkey E the d’ decreased from 1.32 to 0.82 (p < 0.05; t test, n = 14) (Figure 6E,F).
Figure 6 | Effect of CNQX on neurons activity. a,b, The influence CNQX on neuronal activity in an example neurons in monkeys J (A) and E (B). Black curves, pre-drug period. Red curves, activity after CNQX application. Continuous curves, preferred cue. Dashed curves, non-preferred cue. c,d, Average activity of neurons in monkeys J (N=27) (C) and E (N=14) (D). e,f, Abscissa, d’ before CNQX delivery. Ordinate, d’ during CNQX delivery.
Contribution of AMPA-Rs and NMDA-Rs during working memory

We next investigated the influence of CNQX on the Fano-factor. In both monkeys, CNQX did not have a significant effect on the Fano factor during the delay period, both for the preferred direction (p > 0.05; t test, n = 27 for monkey J, p > 0.05; t test, n = 14 for monkey E) and the anti-preferred direction (p > 0.05; t test, n = 27 for monkey J, p > 0.05; t test, n = 14 for monkey E). Spiking activity did not restore to pre-drug levels when CNQX administration ceased. Some single units (9 out of 19 for monkey J, 4 out of 14 for monkey E) did not even show a trend of recovery, in accordance with previous studies showing that CNQX has longer-lasting effects.\(^{35,39}\)

To examine the time-course of the AMPA-R contribution, we determined the difference between spiking activity before and during the administration of CNQX (Figure 7). A three-way ANOVA with factors epoch, drug/no-drug and monkey for the preferred cue location revealed that the influence of CNQX on the firing rate during the cue, memory and saccade epochs was higher than the effect on spontaneous activity ($F_{3,328} = 8.87, p < 0.001$). While for monkey J the contribution of AMPA-Rs to spiking activity increased progressively during the trial (Figure 7A), this was not that evident for monkey E (Figure 7B).
Figure 7 | Effect of CNQX on neuronal activity elicited by cues at the preferred and non-preferred location. The influence of CNQX on the neuronal responses was determined by subtracting neuronal activity in the pre-drug period from that elicited when CNQX was applied, in monkeys J (a) and E (b). Green trace, activity elicited by the preferred cue. Red trace, activity elicited by the anti-preferred cue.

Comparison of the effect of APV and CNQX on delay activity
In addition to the ejection current, the magnitude of the effects of APV and CNQX also depend on a number of poorly controlled factors, including the efficiency of the drug, the distance between the neuron and pipette and the diffusion and clearance of the drugs. It is therefore difficult to directly compare the magnitude of the effects of these drugs. However, it is possible to determine how well the influence on delay activity for the preferred cue predicts the influence for the non-preferred cue, by computing the correlation (Figure 8). We restricted this analysis to neurons with a minimal delay activity of 5Hz in the anti-preferred direction before drug application, so that there was some room for a decrease in activity.
Contribution of AMPA-Rs and NMDA-Rs during working memory

Figure 8 | Comparison of effects on delay activity at the preferred and non-preferred location for APV and CNQX. Effects of drug application during the delay period at the preferred location (x-axis) and anti-preferred location (y-axis) during the application of APV (left graph) in monkey B (red) and monkey J (blue) and during the application of CNQX (right graph) in monkey J (blue) and monkey E (green).

The correlation coefficient for APV was 0.43, which was not significant (N=15 cells with sufficient delay activity; p > 0.1). Thus, the activity decrease in the preferred direction was a relatively poor predictor of the activity decrease in the anti-preferred direction. The correlation coefficient for CNQX was 0.94, which was significant (N=22 cells; P < 10^{-4}) so that the prediction worked much better for CNQX. Indeed, the difference between the magnitude of correlation coefficients for APV and CNQX was also significant (p < 5·10^{-4}; see Methods for how the p-value was determined). It seems likely that this difference between APV and CNQX is caused by the distinct actions of AMPA-Rs and NDMA-Rs. When glutamate binds to an AMPA-R, the channel opens and the cell is activated. In contrast, magnesium blocks NMDA-Rs when the cell is not sufficiently depolarized. This magnesium block may explain why a
decrease in delay activity in the preferred direction was not always accompanied by a comparable decrease in the anti-preferred direction. The NMDA-R may already have been blocked by magnesium in the anti-preferred direction, so that APV could not exert its effect. However, we do note one caveat with this analysis. The minimal delay activity of 5Hz eliminated many cells of monkeys B and E from the analysis, so that this result was largely based on neurons of monkey J (N=13 of 15 cells for APV and 19 of 22 cells for CNQX).

Discussion

In this study, we investigated the contribution of AMPA-Rs and NMDA-Rs to persistent firing in dlPFC. We iontophoretically applied antagonists of AMPA-Rs and NMDA-Rs using relatively small ejection currents to perturb activity without entirely blocking it to obtain sensitive measures of the role of the receptors during different epochs of a delayed saccade task. Although the blockade of glutamate receptors had substantial effects on spiking activity, we did not find consistent effects on the monkeys’ accuracy (Figure 2), similar to many previous studies using iontophoretic drug application. This absence of an effect on accuracy is expected as iontophoretically applied drugs do not spread far, and we caused a relatively weak perturbation in the activity of a small population of neurons.

We found that AMPA-Rs and NMDA-Rs contribute to neuronal activity during all phases of the ODR task. The similarity of the of the effects AMPA and NMDA blockers on persistent activity differs from a previous study examining texture segregation in area V1, where AMPA blockers mainly decreased the visually driven activity, whereas NMDA interfered specifically with the enhanced representation of figural texture-elements.
Contribution of AMPA-Rs and NMDA-Rs during working memory

over the background. In the present working memory task, the effects of AMPA-Rs were largely additive, because the decrease in spiking activity caused by CNQX was relatively independent of the pre-drug activity level (Figure 6C,D). Furthermore, the reduction of activity elicited by the cue in the neurons’ preferred direction predicted the decrease in activity in the anti-preferred directions relatively well (Figure 8). The effects of blocking AMPA-Rs were prominent in the baseline epoch, during the cue-period, the memory delay and also around the time of the saccade. In contrast, NMDA-Rs contributed strongly to the activity of well-driven neurons and less to the firing rate of the activity of weakly activated cells (Figure 3C,D).

Accordingly, the decrease of activity elicited by the preferred cue caused by APV was a poor predictor for the decrease in activity for the non-preferred direction. Our finding that NMDA-Rs amplify the activity of well-driven neurons whereas the influence of AMPA-Rs tend to be additive is compatible with previous results in the cat visual cortex.

These differential effects of AMPA-Rs and NMDA-Rs also explained the difference of the effects of APV an CNQX on the reliability of the spatial tuning, as quantified with the d’. Blocking AMPA-Rs caused a relative moderate decrease of the d’ because it decreased activity elicited by cues at the preferred and anti-preferred locations similarly so that the d’ decreased only slightly (Figure 6E,F). In contrast, blocking NMDA-Rs strongly reduced the d’ because the decrease in activity elicited by the neurons’ preferred cue was more pronounced than that elicited by the non-preferred cue (Figure 3E,F).

A previous study by Wang et al. suggested that persistent activity relies on the unique properties of the NMDA-R, with its voltage dependent gating due to the magnesium block and its relatively long time constant. The idea is that NMDA-Rs cause a positive feedback loop between
membrane depolarization and the release of the magnesium block, so that the excitatory currents can outlast a transient input onto the cell. However, our results do not support a specific role of NMDA-Rs in persistent firing, for a number of reasons. First, we found that AMPA-Rs and NMDA-Rs both contributed to persistent activity, and that the contribution of AMPA-Rs (Figure 6C,D) was comparable to that of NMDA-Rs (Figure 3C,D). Second, the contribution of NMDA-Rs to the initial visual response and to saccade-related activity was comparable to the contribution to the delay activity (Figure 3C,D). Third, blocking of NMDA-Rs also reduced spiking activity and weakened the spatial tuning of visual cells without persistent activity (Figure 5).

At first sight, our results are therefore at odds with the results of Wang et al. One difference between studies was in the choice of antagonists. We used the competitive NMDA-R antagonist APV, whereas Wang et al. used MK801, which is a non-competitive antagonist, and the NMDA subunit antagonist Ro25-6981, which blocks NMDA-Rs with the NR2B-subunit. Furthermore, we used CNQX to block AMPA-Rs, whereas Wang et al. used both NBQX and CNQX. However, we believe that the most important difference between studies is in the dosage of the drugs. Wang et al. almost completely abolished delay activity with the application of their NMDA antagonists but observed weaker effects with AMPA-R antagonists. It is likely that they would have seen a more complete suppression of delay activity with higher dosages of AMPA-blockers, because in our experience higher dosages of CNQX can also completely block neuronal activity. In the present study, we rather chose to use iontophoresis currents that perturb but do not abolish activity, and we did not observe a specific effect of NMDA-Rs on persistent neuronal activity. It is also of interest to compare the present results to a study by Skoblenick and Everling who recorded from the dIPFC during
Contribution of AMPA-Rs and NMDA-Rs during working memory

systematic dose of the NMDA antagonist ketamine. Ketamine increased, rather than decreased the activity of most dLPFC neurons, in accordance with studies in the frontal cortex of rodents. This discrepancy is most likely related to the systematic application of ketamine, which influences neuronal activity in many brain regions that can indirectly impact on the activity of neurons in the dLPFC. In the present and previous studies, the local, iontophoretic application of NMDA blockers invariably decreased neuronal activity in dLPFC.

AMPA and NMDA receptors are not the only receptors that have been implicated in the mechanisms for persistent firing. Blocking dopamine receptor D1 for instance revealed an ‘inverted-U’ dose-response relationship, because too little or too much receptor activity reduces persistent firing, and both D1 and D2-receptors influence the representation of task-rules during a delay. Similarly, acetylcholine has been implicated in the maintenance of persistent activity through its action on nicotinic and muscarinic receptors, although a recent study demonstrated that the decrease in activity caused by muscarinic blockers are not specific to delay activity, just as we observed for NMDA-Rs. Furthermore, a2A-adrenoceptors also impact on persistent firing, in part by acting on non-selective cation permeable transient receptor potential channels (TRPC) and hyperpolarization activated cyclic nucleotide-gated potassium channels. Thus, many receptors contribute to persistent firing, implying that it relies on a complex interplay between many receptors, including NMDA-Rs and AMPA-Rs.

The activation of the receptors that cause persistent firing requires synaptic input that might be provided by other neurons with persistent activity within the same area and from reciprocal excitatory loops between cortical and/or subcortical areas. In the first, local scenario, the persistent firing would be generated by reciprocal excitation between
Chapter 3

pyramidal neurons with similar tuning in the same area. In the second, more
global scenario, persistent firing is maintained by reciprocal excitation
between cortical areas or by loops through subcortical structures, including
the basal ganglia. Persistent activity during memory delays is indeed
observed in many other cortical areas, including the parietal cortex, medial superior temporal cortex, the inferotemporal cortex and even in
the primary visual cortex. A study by Chafee and Goldman-Rakic in
monkeys combined local cooling of either parietal and prefrontal cortex with
recording in the other area during a working memory task. The inactivation of
one area decreased the activity of some neurons in the other region but
increased the activity of others, without a clear effect on behavior. Recent
studies that approached the same questions in mice revealed an important
role of frontal cortex in the maintenance of information during memory delays.
Optogenetic silencing of neuronal activity in the frontal cortex during
memory delays was able to delete working memories. Interestingly, a brief
unilateral blockade of persistent activity in frontal cortex could be later
restored by activity of the contralateral frontal cortex, in accordance with the
hypothesis that persistent activity relies on the reverberation of activity
between areas of the cerebral cortex.

In combination with these previous studies, the present results
contribute to our understanding of how working memories are maintained in
the frontal cortex, revealing that both AMPA-Rs and NMDA-Rs sustain
persistent spiking activity, while the relative contribution of NMDA-Rs
increases for neurons that are strongly active. This is a relevant finding, both
for models on the neural mechanisms during working memory and
for clinical conditions in which working memory is impaired. Future studies
can now investigate how the dynamics of these receptors, in combination with
Contribution of AMPA-Rs and NMDA-Rs during working memory

recurrent excitation within and between brain regions, explains how task-relevant information is kept online during memory delays.
References


Contribution of AMPA-Rs and NMDA-Rs during working memory


Chapter 3


Contribution of AMPA-Rs and NMDA-Rs during working memory


Contribution of AMPA-Rs and NMDA-Rs during working memory


Chapter 4

Texture segregation causes early figure enhancement and later ground suppression in areas V1 and V4 of visual cortex

Jasper Poort, Matthew W. Self, Bram van Vugt, Hemi Malkki and Pieter R. Roelfsema

Chapter 4

**Introduction**

The assignment of image elements to figure or background is an elementary step in visual perception. A powerful illustration of this process is the face–vase illusion (Fig. 1A), where our interpretation of the image alternates\(^1\). The assignment of image regions to figure or ground has a profound influence on perception, because image elements that are part of figures receive preferential processing and leave stronger memory traces\(^2,3\). The perceptual status of the ground regions is less clear. One study suggested that background features are not processed up to a perceptual level\(^3\), but others suggested that background regions are actively suppressed\(^4-6\). This question can also be formulated at the level of neuronal processing: does figure-ground segregation enhance the neuronal representation of the figure, suppress the background or both (Fig. 1D)?
**Figure 1** | Illustration of figure-ground organization and the behavioural paradigm. **a**, Face-vase illusion. This classic ambiguous figure-ground display causes perception to alternate between two faces or a vase on a formless background. **b**, Illustration of the three types of stimuli. The grey circular outline illustrates the position of a receptive field (RF), the central grey circle indicates the fixation point. Left, RF is on a figure; Middle, RF is on the background; Right, RF is on a uniform texture. Note that the image elements in the RF are the same in all conditions. **c**, In Experiment 1 each trial started when the monkey directed gaze to the fixation point (FP), and kept fixation for an additional 300 ms. Fixation was followed by two successive stimuli. The first stimulus consisted of one of the three configurations in panel A and was presented for 400 ms. The second stimulus was again one of three configurations, which was presented for an additional 400 ms. Next, the fixation dot disappeared and the monkey was rewarded after making a saccade to the location of the figure, if one was present. If no figure was present (uniform texture), the monkey had to maintain fixation for an additional 250 ms, after
which the animal was rewarded. \textbf{d}, Different scenarios for figure-ground modulation. Figure-ground modulation can be the result of figure enhancement or ground suppression, or a combination of these two processes.

Previous studies of neuronal activity in primary visual cortex (V1) and mid-level area V4\textsuperscript{7,8} during texture segregation found that neurons respond more vigorously to image elements of a figure than to elements of the background (Fig. 1B). This response difference is called figure-ground modulation (FGM). Interestingly, FGM is strongest in the superficial and deep layers of V1, and weakest in input layer 4 and it is associated with a pattern of synaptic activity that suggests an important role for feedback from higher visual areas\textsuperscript{9}. However, the precise contributions of figure enhancement and ground suppression to FGM in the texture segregation task remain unknown. Strong suppressive effects were observed in Landman et al. (2003), who demonstrated that the activity elicited by background elements in V1 decreases with the number figures present in a display\textsuperscript{10}. In contrast, one fMRI study has demonstrated that responses elicited by figures are enhanced in visual cortex (V1–V4)\textsuperscript{11}. However, another fMRI study\textsuperscript{12} did not find figure enhancement but only background suppression. Both studies also found effects of figure-ground perception in extra-striate cortex, but they could not resolve activity related to figure and background regions. In a related study on contour-grouping, Chen et al. (2014) examined activity in areas V1 and V4 of monkeys trained to perceive an elongated contour formed by collinear line elements among randomly oriented distractor elements\textsuperscript{13}. The representation of elements of the contour was enhanced in V1, and the activity elicited by the randomly oriented line elements was suppressed (see Gilad et al. 2013, for similar results using voltage-sensitive dye imaging in V1\textsuperscript{14}). A recent fMRI study also reported an enhanced contour representation combined with a suppression of randomly
oriented contours\textsuperscript{15}. Both studies\textsuperscript{15,13} also revealed effects in extra-striate cortex, but again, contour and background responses could not be measured separately.

In the texture-segregation task, the contributions of figure enhancement and ground suppression to texture-segregation and their timing\textsuperscript{16} remain unclear and the role of these two processes in extra-striate cortex is generally unknown. Furthermore, previous studies did not address the influence of activity enhancement and suppression in the different cortical layers. We therefore recorded from V1 and V4, a higher area that plays an important role in texture-segregation\textsuperscript{17,18}, to address the following questions: (1) how do figure enhancement and ground suppression contribute to FGM in V1 and extra-striate cortex during texture segregation? (2) What is the profile of figure enhancement and ground suppression across the layers in area V1? (3) Do neurons with figure enhancement also exhibit ground suppression or do these two processes influence different neuronal circuits?

\section*{Materials and Methods}

\textit{Visual stimulus and behavioural paradigm}

We conducted two experiments. In Experiment 1 we investigated the contribution of figure enhancement and ground suppression in a texture segregation task in areas V1 and V4 and in Experiment 2 we investigated the laminar profile of suppression and enhancement in area V1. The general aim of the experiments was to measure figure enhancement and ground suppression. We isolated the contribution of figure enhancement by comparing activity elicited by a figure in the neurons’ receptive field to that elicited by a homogeneous texture (Fig. 1B, left vs. right panel). We isolated ground
suppression by comparing activity elicited by the homogeneous texture to that elicited by the ground condition in which there was a figure remote from the neurons’ receptive field (Fig. 1B, right vs. middle panel).

Three adult male monkeys participated in Experiment 1 and two adult male monkeys in Experiment 2. They were seated at a distance of 0.75 m from a monitor (width: 0.4 m) with a resolution of 1024×768 pixels and a frame rate of 110 Hz (85 Hz in Experiment 2).

In Experiment 1, the visual stimulus consisted either of a square figure with oriented line elements (32 pixels long, 0.93°, and 2 pixels wide) on a background with an orthogonal orientation, or it consisted of a homogeneous texture (Fig. 1B). To construct the stimulus, we first made four full-screen base textures, two with an orientation of 45° and two with an orientation of 135°. A base texture was made by randomly placing 13,000 black line elements (luminance 2.8 cd m⁻²) with a given orientation on a white (luminance 94 cd m⁻²) background. We then created full screen stimuli for the figure and ground conditions by copying a square 4°×4° region of a 45° base-texture onto a 135° base texture or by copying the same square region of a 135° base-texture onto a 45° base-texture. In the uniform condition, we presented only a 45° or 135° base texture that covered the full screen. To analyze the figure–ground modulation, we averaged neuronal responses to the two complimentary stimuli, thereby ensuring that the RF was stimulated on average by the same set of local features, regardless of whether the RF was on the figure, background or homogeneous texture (see Fig. 1B).

A trial started as soon as the monkey's eye position was within a 1°×1° window centred on the fixation point (0.58°), presented on a grey background (luminance 34 cd m⁻²). The monkey had to maintain fixation within the fixation window until cued to make a saccade by the disappearance of the fixation point.
In Experiment 1, the monkey saw two figure-ground stimuli that were presented successively. When the monkey had kept fixation for 300 ms, the first stimulus was presented (period 1, 400 ms). It consisted of either a figure in the RF (figure condition), a figure that was not in the RF (ground condition) or a uniform texture. The figure could appear at one of four locations: in three monkeys, we recorded data from two V1 electrode arrays (see below) so that RFs were clustered at two positions in the visual field. Therefore, we used two figure positions that were centred on one of the RF clusters, and two corresponding positions at the same eccentricity as the RF clusters, rotated by 180°. After period 1, the second stimulus was presented that was again a figure, a ground or uniform condition (period 2, 400 ms). After period 2, the fixation point disappeared, and the monkey had to make a saccade to the target window of 4°×4° centred on the location of figure to obtain a drop of apple juice as a reward (Fig. 1C). If there was no figure present in period 2 (uniform condition), the monkey was rewarded if he maintained fixation for an additional 250 ms. The monkeys detected figures with high accuracy (98% correct for monkey 1, 94% for monkey 2, and 96% for monkey 3). The accuracy was lower in catch trials without a figure (92% for monkey 1, 63% for monkey 2, and 74% for monkey 3) because the monkeys had to maintain fixation for a longer duration. We only included correct trials in all of our analyses.

In Experiment 2 there was only one epoch with a full-screen texture (5,345 line elements per texture with a width of 1 pixel and a length of 16 pixels). In 75% of trials, the texture contained a figure (4°×4°). The figure was placed in the RF (figure condition), at one of two locations situated at the same eccentricity but at 120° away from the RF (ground condition). The animal had to maintain fixation for 300 ms, after which the fixation dot disappeared and the monkey had to make an eye movement to a 4°×4° window centered on the figure. On the other 25% of trials a uniform texture was presented and the
animal was rewarded for maintaining fixation for an additional 400 ms after
the fixation dot was extinguished. The performance in detecting figures was
above 95% correct for both monkeys. The accuracy in catch trials was 77% for
monkey 4 and 89% for monkey 5.

Surgical procedures

We used the same surgical protocol as described previously\textsuperscript{8,19}. The monkeys
underwent two surgeries under general anaesthesia that was induced with
ketamine (15 mg kg\textsuperscript{-1} injected intramuscularly) and maintained after intubation
by ventilation with a mixture of 70\% N\textsubscript{2}O and 30\% O\textsubscript{2}, and supplemented with
0.8\% isoflurane, fentanyl (0.005 mg kg\textsuperscript{-1} intravenously), and midazolam (0.5 mg
kg\textsuperscript{-1} h\textsuperscript{-1} intravenously). In the first surgery, we implanted a head holder. In the
second surgery, we implanted arrays of 4×5 electrodes (Cyberkinetics
Neurotechnology Systems Inc.) in areas V1 and V4 for Experiment 1, and a
chamber above V1 over a small craniotomy for the laminar recordings of
Experiment 2. All procedures complied with the NIH Guide for Care and Use of
Laboratory Animals (National Institutes of Health, Bethesda, Maryland), and
were approved by the institutional animal care and use committee of the Royal
Netherlands Academy of Arts and Sciences.

Recording of neuronal activity

In Experiment 1 we recorded multi-unit activity in two monkeys that were
chronically implanted with electrode arrays in V1 and V4, and one monkey with
arrays only in V1. In Experiment 2 we recorded from V1 of two monkeys (they
did not take part in Experiment 1) using a multicontact laminar probe (‘U-
probe’, Plexon Inc) that was inserted into V1, as described previously\textsuperscript{19}. In both
experiments, multi-unit spiking activity (MUA) was recorded with a TDT
Figure enhancement and ground suppression during FG segregation

(Tucker Davis Technologies) data acquisition system. As in previous studies\textsuperscript{20-22}, MUA signals were amplified, band-pass filtered (500-5000 Hz), full-wave rectified and then low-pass filtered at 500 Hz and sampled at a rate of 763 Hz. The MUA signal contains spikes from neurons within \( \sim 150 \mu m \) of the electrode tip\textsuperscript{9}, which corresponds to the distance over which a V1 cell can be recorded with single-unit recording. Accordingly, the MUA represents the pooled activity of a number of single units in the vicinity of the tip of the electrode and the population response obtained with this method is therefore similar to the population response obtained by pooling across single units\textsuperscript{22,23}. The eye position was measured with an eye tracker camera system (Thomas Recording, Germany) and sampled at a rate of 250 Hz.

For Experiment 2, we also computed the current-source density (CSD) from the local field potential. The LFP at each recording site was obtained by low-pass filtering the signal from the electrode below 200 Hz. The CSD was then calculated as:

\[
CSD(x) = -\sigma \cdot \frac{\varphi(x - h) - 2\varphi(x) + \varphi(x + h)}{h^2},
\]

where \( \varphi \) is the LFP voltage (in V), \( x \) is the point at which the CSD is calculated, \( h \) is the spacing of recording sites for the computation (here 200\( \mu m \)) and \( \sigma \) is tissue conductivity (we used 0.4 S.m\textsuperscript{-1}). This equation yields the CSD in units of A.m\textsuperscript{-3} but we here report the CSD in physiologically more relevant units; \( \mu A.mm\textsuperscript{-3} \).

For each V1 recording site we measured the receptive field by determining the onset and offset of the response to a slowly moving light bar to eight movement directions\textsuperscript{24}. In Experiment 1, the median V1 receptive field area was 1.6 deg\(^2\) (range 0.08 to 7.6 deg\(^2\)), and the median eccentricity was
4.02° (range 2.5 to 6.9°). In Experiment 2 the median receptive field area was 2.2 deg² (range 0.39 to 15.8 deg²) and the median eccentricity was 4.12° (range 1.8° to 12°). In V4 (Experiment 1) we mapped RFs by presenting white dots (0.5 deg, luminance 82 cd m⁻²) on a grey background (luminance 14 cd m⁻²) at different positions of a grid (0.5 deg spacing). The hotspot of the V4 RF was defined as the position with the maximum response (median eccentricity 4.04°, range 0.79° to 7.43°) and the RF borders as the locations where activity fell below 50% of the maximum²⁵. Using this criterion, the median V4 RF area was 19.7 deg² (range 6.5 to 38 deg²).

Data analysis

We quantified the visual responsiveness of neurons at each recording site by calculating the mean spontaneous activity level across all conditions $Sp$ and the standard deviation $s$ across trials in a 200 ms time window preceding stimulus onset. We then computed the peak response, $Pe$, by smoothing the average response across conditions with a moving window of 25 ms and taking the maximum during the stimulus period (0–300 ms after stimulus onset). The visual responsiveness index was computed as $VR=(Pe-Sp)/s$. Only recording sites with a good visual response ($VR>3$) were included in the analyses. In Experiment 1 we included 102 V1 recording sites (40 in monkey 1, 33 in monkey 2, and 29 in monkey 3) and 36 in V4 (14 in monkey 1 and 22 in monkey 2). The number of recording sites in Experiment 2 will be specified below. MUA data from each recording site were normalized by subtracting $Sp$ and subsequently dividing by $(Pe-Sp)$.

FGM was computed as the difference between the responses evoked by the figure and background. To quantify the amount of figure enhancement, we
computed the difference between the response evoked by the figure and the response evoked by the uniform texture (figure-uniform modulation, FUM). To quantify the amount of ground suppression, we computed the difference between the response elicited by the uniform texture and the background (with the figure at another location, outside the receptive field) (uniform-ground modulation, UGM).

We determined the latency of the visual responses, FUM and UGM by fitting a function \( f(t) \) to the neural response (or response difference)\(^{26,27} \). The function was derived from the assumptions that the onset of the response has a Gaussian distribution and that a fraction of the response dissipates exponentially which yields the following equation:

\[
f(t) = d \cdot \exp(\mu \alpha + 0.5 \sigma^2 \alpha^2 - \alpha t) \cdot (G(t, \mu + \sigma^2 \alpha, \sigma) + c \cdot G(t, \mu, \sigma)),
\]

where \( G(t, \mu, \sigma) \) is a cumulative Gaussian density with mean \( \mu \) and standard deviation \( \sigma \), \( \alpha \) is the time constant of the dissipation, and \( c \) and \( d \) represent the contribution the non-dissipating and dissipating component (see Roelfsema et al. (2003) for details\(^{27} \)). We fitted \( f(t) \) to the responses using the curve fitting toolbox in MATLAB (MathWorks) and defined the latency as the time point where the fitted function reached 33% of its maximum. To determine the significance of a latency difference between two conditions, we used a bootstrapping procedure. We randomly selected a number of recording sites equal to the original sample with replacement and fitted the latency in each condition and subtracted these latencies to obtain a distribution of the latency difference across 1000 repeats. The latency of the figure-uniform modulation was measured by fitting the curve to the difference between the response evoked by the figure and uniform texture, and the latency of the
uniform–ground modulation by fitting a curve to the difference between the response evoked by the uniform texture and background.

To quantify how reliably individual recording sites discriminated between the different stimulus conditions we computed the d-prime: 

\[ d_{AB} = \frac{(m_A - m_B)}{s} \]

where \( m_A \) and \( m_B \) are the mean responses in stimulus conditions \( A \) and \( B \), and \( s \) is the pooled standard deviation. \( d_{FU} \) is a measure for the discrimination between a figure and a uniform texture, and \( d_{UG} \) for the discrimination between a uniform texture and the background. We quantified the correlation between the d-prime in different conditions with Pearson's correlation coefficient, and used the Student's t distribution to assess significance.

**Laminar analysis**

In Experiment 2 we recorded from 30 penetrations in monkey 4 and 14 penetrations in monkey 5 with laminar electrodes with a spacing between neighbouring electrodes of 100 µm. Part of the data of Experiment 2 have been used in a previous study\(^9\), but that study did not analyse the responses elicited by the homogeneous texture, which allowed us to separately determine the contribution of figure enhancement and ground suppression to FGM. We identified the depth of each recording site relative to the layer 4c/layer 5 boundary using the CSD as described previously\(^9\). We then assigned each recording site to one of three laminar compartments based on the distance of the recording site to the boundary. Recording sites between -0.7 and -0.1mm (i.e. below the boundary) were assigned to the deep layers, those between 0 and 0.5mm (above the boundary) were assigned to layer 4 and those between 0.6 and 1.0mm to the superficial layers. Sites below -0.7mm and above 1.0mm were excluded from the analysis. In this experiment, we also excluded
recording sites with a VR less than 3. The number of remaining MUA recording sites per compartment were as follows: monkey 4: \( N_{\text{deep}} = 76, N_{\text{layer 4}} = 97, N_{\text{superficial}} = 33 \); monkey 5: \( N_{\text{deep}} = 84, N_{\text{layer 4}} = 87, N_{\text{superficial}} = 31 \). Recordings from different penetrations were aligned on the basis of the layer 4c boundary location before averaging across penetrations. To estimate the latency of the CSD modulation, we used the current sink in layer 5, because it was a reliable feature of both the figure enhancement and ground suppression. The current sink was well fit by a Gaussian density function: 
\[
g(t) = a \cdot g(t, \mu, \sigma)
\]
with mean \( \mu \) and standard deviation \( \sigma \), and amplitude \( a \). As a measure for the latency of the sink, we took the time point at which the fitted curve reached 33% of its maximum. To quantify the reliability of figure enhancement and ground suppression across the different laminae we computed \( d \)-primes: \( d_{\text{FU}} \) and \( d_{\text{UG}} \), as described above. As we were particularly interested in the laminar profile of ground suppression we only included penetrations in the laminar analyses with significant UGM when averaging across the entire penetration \( (P<0.05, \text{Wilcoxon signed-rank test}) \). Note that this two-tailed test cannot cause a bias in the results.

Statistical significance of CSD sinks/sources was assessed using a non-parametric bootstrap cluster statistic. The full details are given in Self et al. (2013)\(^9\). Briefly, 2 dimensional (time \( \times \) depth) \( t \)-statistic maps were calculated for each penetration for the difference between figure and uniform, or uniform and ground. These \( t \)-maps were thresholded at \( P<0.05 \) (two-tailed) and adjacent \( t \)-scores above threshold were clustered and the absolute values summed to produce a cluster statistic. Bootstrapping was used to assess the significance of these clusters.
Eye movement analysis

The monkeys had to maintain their eye position within a 1° diameter fixation window. We carried out a stratification analysis to investigate the potential effect of small differences between the eye-positions in the figure, ground and uniform stimulus conditions\textsuperscript{89,28}. We computed the average horizontal and vertical eye-position in each trial. We then divided the fixation window in 4 x 4 bins of 0.25° × 0.25° and assigned every trial to one of these bins based on the average eye-position. We equated the number of trials in each bin across conditions (figure, ground, uniform) by randomly removing surplus trials to ensure that the distribution of eye movements was similar across these conditions and reanalyzed the data of the trials that remained after stratification.

Results

Experiment 1: Behavioural task

We trained three monkeys to perform a figure-detection task with two epochs (see Fig. 1B,C and Materials and Methods). After the monkey directed gaze to the fixation point, we presented the first stimulus which was either a figure during a period of 400 ms at one of four possible locations, or no figure was presented (uniform condition). This was followed by a second period of 400 ms in which a second stimulus was presented, which could again contain a figure or no figure. At the end of period 2 the fixation point disappeared. If a figure was present in period 2, the monkey had to make a saccade to its centre. If no figure was present, he had to maintain fixation to obtain a reward (catch trial). Note that the stimulus during the first period was uninformative about the
required saccadic eye movement, although we cannot entirely rule out the possibility of covert eye movement planning during this epoch.

**Figure enhancement and ground suppression in V1 and V4**

Fig. 2A,E shows the activity elicited by the figure, the background and the uniform texture in V1 and V4 during the first stimulus epoch, averaged across three monkeys. Before pooling the neuronal responses across the recording sites, we first normalized the activity to the peak response, which is elicited after ~40 ms. This initial response in V1 was similar in the three conditions, but after a delay the responses to the figure became enhanced relative to responses to the background and uniform texture (Fig. 2A, blue trace in the lower panel shows the difference between figure and uniform texture, FUM; figure-uniform modulation). The modulation of neuronal activity may appear small if it is compared with the initial peak response, but it is in fact quite strong in the later period, when these transients have subsided. The V1 population response elicited by the figure was enhanced by 106% relative to the response evoked by the background (time window 150–300 ms). After an additional delay, the responses to background became suppressed relative to the uniform texture (UGM; uniform texture minus background response, green in lower panel of Fig. 2A). Note that this later suppression is induced by a figure in the opposite hemifield. Compared to the uniform texture, V1 activity elicited by the figure was enhanced by 42% relative to the response evoked by the uniform texture (Wilcoxon signed-rank test, all monkeys $P<0.001$), and the response evoked by the ground was reduced by 31% relative to uniform texture (all monkeys $P<0.01$).
In V4, the RFs were much larger than in V1 and in most cases the V4 RF overlapped with both the interior and the edges of the figure so that the figure can act as a pop-out stimulus at this spatial scale. As a result, there was a relatively early enhancement of V4 responses to the figure compared with the responses to the background and uniform texture (Fig. 2E). As in V1, this early enhancement was followed by a delayed suppression of the response to the background relative to the uniform texture, caused by the presence of a figure in the opposite hemifield. When compared to the response elicited by the uniform texture, V4 activity evoked by the figure was enhanced by 30% and the response to the background was reduced by 18% (both $P<10^{-6}$, both monkeys $P<0.01$). We determined the latency of these effects by fitting curves (see Methods). The visual response latency in V1 was 39 ms with a 95%-confidence interval (CI) of 38–40 ms, and it was followed by FUM at 82 ms (CI 68–102 ms)—significantly later ($P<0.001$; bootstrap analysis)—which was, in turn, followed by UGM at 137 ms (CI 136–141 ms), which was significantly later ($P=0.02$). The latency of the visual response in V4 was 49 ms (CI 48–50 ms), followed by FUM at 57 ms (53–62 ms), which was in turn followed by UGM at 133 ms (CI 132–140 ms) (latency differences, both $P<0.001$). FUM in V1 was later than FUM in V4 ($P=0.002$), as shown previously, but we found that the timing of UGM was similar in areas V1 and V4 (137 ms in V1 vs. 133 ms in V4; $P=0.88$). Thus, the suppressive effect of the figure in the opposite hemisphere has a similar timing in the two areas.
Figure 2 | Figure-enhancement and ground suppression in V1 and V4. a,b Average neuronal activity in V1 in first (A) and second stimulus period (B) when the receptive field was on the figure (F, blue), background (G, red) or on the uniform texture (U, green). N=102 recording sites. The lower panels illustrate the time-course of figure enhancement (blue, figure minus homogeneous) and ground suppression (green, homogeneous minus ground). The arrows show the latency of figure enhancement and ground suppression and
the bars denote the 95%-confidence intervals. **c, d** d-primes of figure enhancement (x-axis, $d_{FU}$) and ground suppression (y-axis, $d_{UG}$) for stimulus 1 (C) and stimulus 2 (D). Data from the three monkeys are shown in different colours. **e, f**, Activity in area V4. $N=36$ recording sites. **g, h**, d-primes in V4.

We computed d-primes (see Methods) to quantify how reliably individual recording sites discriminated between a figure and a uniform texture ($d_{FU}$) and between a uniform texture and the background ($d_{UG}$). Most of the V1 recording sites exhibited an increased response to the figure relative to uniform textures as well as a reduced response to the background (Fig. 2C) ($d_{FU}$, mean 0.16, $d_{UG}$, mean 0.10, Wilcoxon signed-rank test, both $P<10^{-10}$). In V4 the results were similar because the figure elicited a greater response than the uniform textures, and responses to the background were suppressed relative to those evoked by uniform textures (Fig. 2G) ($d_{FU}=0.90$, $d_{UG}=0.44$, Wilcoxon signed-rank test, both $P<10^{-6}$).

The discrimination between figure and uniform textures in V4 was stronger than in V1 (V1 $d_{FU}=0.16$, V4 $d_{FU}=0.90$, Wilcoxon rank-sum test, $P<10^{-9}$) and the same was true for the discrimination between uniform textures and background. We computed the correlation between $d_{FU}$ and $d_{UG}$ across recording sites to investigate whether neurons tended to co-express both effects. Interestingly, the correlation between figure enhancement and ground suppression d-primes was not significant in V1 and V4 (V1, $r=0.16$, $P=0.10$, V4, $r=-0.31$, $P=0.06$). This result indicates that figure enhancement and ground suppression are separate processes that influence different circuits, as is also evident from the difference in their timing.

In the first stimulus period we presented a figure-ground display but the monkey was not required to make an eye-movement. After 400 ms the
second stimulus appeared. If a figure was present in the second phase, it served as target for an eye movement. We pooled across all conditions with a particular stimulus in the second period, allowing the stimulus in the first period to vary (Fig. 1C). We ensured that the stimulus history was balanced so that the first stimulus did not predict the second stimulus. When we corrected for the onset time of the second stimulus (at 400 ms), we found that the latency of figure enhancement in V1 (Fig. 2B) was 76 ms (CI 66–91 ms) and that it was followed by ground suppression at 141 ms (135–146 ms), significantly later (P<0.001). In V4 (Fig. 2F), the latency of figure enhancement was 76 ms (CI 71–83 ms), which was followed by ground suppression at 137 ms (133–146 ms) (P<0.001).

Figure enhancement and ground suppression were also highly consistent across the population of recording sites in the second stimulus period. In V1, the average response elicited by the figure was enhanced by 40% relative to the response evoked by the uniform texture (Fig. 2B, all monkeys P<10^{-4}), and the response evoked by the ground was reduced by 31% (all monkeys P<0.01). In V4, figure enhancement was 43%, on average, and ground suppression 16% (Fig. 2F, all Ps<0.001). The same result held up when we examined the d-primes. Our measure for figure enhancement, d_{FU}, had a mean value of 1.08 in V4, higher than the value of 0.15 in V1 (P<10^{-11}). Similarly, ground suppression in V4 with a mean d_{UG} of 0.31 was stronger than that in V1 with a mean of 0.09 (P<10^{-10}, Fig. 2D,H). As in period one, the correlation between figure enhancement and ground suppression d-primes was not significant in V1 (V1, r=0.16, P=0.11) and there was even a significant negative correlation in V4 (r=-0.53, P<0.01). However, this correlation failed to reach significance when the data of two monkeys was analysed separately (both Ps>0.16). We conclude that neuronal activity in period 2 was remarkably similar to that in period 1, and that the findings therefore do not depend strongly on eye movement planning. In both periods figure enhancement in V1
and V4 occurred before ground suppression. The strength of figure enhancement was a poor predictor for the strength of ground suppression across neurons, which confirms that figure enhancement and ground suppression are different processes.

Eye movements do not account for figure enhancement or ground suppression

Small differences between the average eye position in the figure, uniform and background stimulus conditions (within the 1° fixation window) could in principle contribute to the response differences that we observed. We therefore carried out a stratification control analysis in which we first made the distribution of eye position the same across stimulus conditions (see Methods) and repeated our analysis. We found that the neural d-prime values after stratification (period 1, V1 $d_{FUstrat}=0.17$, $d_{UGstrat}=0.11$, V4 $d_{FUstrat}=0.94$ $d_{UGstrat}=0.43$; period 2, V1 $d_{FUstrat}=0.16$, $d_{UGstrat}=0.09$, V4 $d_{FUstrat}=1.03$ $d_{UGstrat}=0.33$) were similar to the original d-prime values without stratification (all Wilcoxon signed-rank test comparing neural d-primes before and after stratification $Ps>0.09$). Thus, small differences in eye position between the conditions cannot account for figure enhancement or ground suppression.

The profile of figure enhancement and ground suppression across the cortical layers

Next, we studied the strength of figure enhancement and ground suppression across the cortical layers of V1 using laminar electrodes in two different monkeys (monkey 4 and 5). We presented textures containing a figure to create the figure and background conditions and also uniform textures (Fig. 1B). As in
Experiment 1, the animals performed a figure-detection task, but now there was only a single epoch. The monkeys either made an eye movement to the figure (on figure/ground trials) or maintained fixation if there was no figure (uniform trials). Relative to uniform textures, figure responses were enhanced (Wilcoxon signed-rank test, both monkeys, \( P<0.001 \)) and background responses were suppressed (both monkeys \( P<0.001 \)) (Fig. 3A). The magnitude and latency of the ground suppression were similar to that in V1 of the monkeys that participated in Experiment 1. Averaged across the layers, the latency of figure enhancement was 84 ms (CI 68–95 ms) and the latency of ground suppression was 171 ms (127–202 ms). In a previous study\(^9\), we found that FGM had the strongest influence on neuronal activity in the deep and superficial layers, and the weakest influence on activity in input layer 4. This previous study compared the ground condition to the figure condition and it did therefore not separate the contributions of figure enhancement and ground suppression.

To isolate figure enhancement, we here compared the responses elicited by the figure to those elicited by the uniform texture (Fig. 3B). Figure enhancement was considerably stronger in the superficial and deep layers than in layer 4. For the quantification of figure enhancement, we grouped recording sites into three laminar compartments (deep, layer 4 and superficial) and calculated \( d_{FU} \) (figure vs. uniform). The level of \( d_{FU} \) varied significantly across these laminar compartments (Friedman test, \( P=0.007 \)). Post-hoc tests revealed that the difference between the deep layers and layer 4 was significant (Wilcoxon signed-rank test, \( P=0.02 \), with Bonferroni correction for multiple comparisons) and the difference between the superficial layers and layer 4 was significant too (\( P<0.03 \)). There was no significant difference in figure enhancement between the superficial and deep layers (\( P=0.69 \)). We then examined the laminar profile of ground suppression by comparing the uniform
Chapter 4

and ground conditions (Fig. 3C). The laminar profile of ground suppression was similar to that of figure enhancement. The values of $d_{UG}$ differed significantly between laminar compartments (Friedman test, $P<0.001$). Ground suppression was significantly stronger in the deep and superficial layers than in layer 4, and suppression was also slightly stronger in the superficial layers than in the deep layers (deep vs. layer 4: $P=0.03$, superficial vs. layer 4: $P=0.004$; deep vs. superficial: $P=0.04$). We next examined the correlation between figure enhancement ($d_{FU}$) and ground suppression ($d_{UG}$) across recording sites, but it was not significant ($r=0.02$, $P=0.47$).

To investigate the synaptic contributions underlying these changes in spiking activity we studied the laminar CSD profile. Sinks in the CSD represent the laminar locations where currents flow into the neurons and they therefore represent putative excitatory inputs, whereas sources represent the laminar locations where the currents flow out of the neurons. The appearance of a full-screen uniform texture produced a typical laminar pattern of current flow with current sinks beginning in layer 4 and then spreading into the superficial and deep layers (Fig. 3D). The earliest sinks in layer 4 are thought to represent excitatory feedforward input from the LGN. We next examined the differences in current flow between the figure and uniform conditions, which provides insight into the connections that contribute to figure enhancement (Fig. 3E). If the figure fell in the neurons’ receptive field, we observed an extra sink in the upper layers (most likely in layers 1 and 2) and layer 5, at a latency of 97 ms (CI 76–103 ms). This pattern resembles the difference in current flow when we compared the figure condition to the background. Interestingly, layers 1, 2 and 5 are targeted by feedback connections from higher visual areas, which suggests that figure enhancement is caused by excitatory feedback from higher visual areas. To examine the currents underlying ground suppression, we subtracted the CSD when the receptive fields fell on the ground from the CSD.
elicited by a homogeneous texture, because the homogeneous texture elicited the strongest MUA response. The laminar profile of this CSD difference was very similar to that underlying figure enhancement, with stronger current sinks in the upper layers and layer 5. Thus, the sinks in the ground condition in layers 1, 2 and 5 were weaker than those elicited by a homogeneous texture, which suggest that ground suppression is associated with a decreased synaptic drive into these layers. The influence of ground suppression on the CSD occurred at a latency of 181 ms (CI 144–194 ms) after stimulus onset, at approximately the same time as the suppression of spiking activity caused by the presence of a figure far from the receptive field of the neurons.

**Figure 3 | The laminar profile of figure enhancement and ground suppression.** a, Average neuronal activity in V1 when the receptive field was on the figure (blue), background (red) or on the uniform texture (green) (408 recording sites). b, Laminar profile of the difference between MUA evoked by the figure and uniform texture. The boundary between layer 4 and 5 is at 0mm. c, Laminar profile of the difference between
activity evoked by the uniform texture and background. d, Current-source density (CSD) at different cortical depths elicited by the appearance of a uniform texture. e, Difference in the CSD evoked by the figure centre and a uniform texture. Warm colours show stronger sinks in the figure condition (and/or stronger sources in the uniform condition). The white asterisks mark significant sinks and sources as assessed by a bootstrap cluster statistic (Methods). f, Difference in normalized CSD evoked by the uniform texture and the background. Warm colours show stronger sinks in the uniform condition (and/or stronger sources in the ground condition).

Discussion

Perceptual organization enhances the representation of figures relative to the background. 

Researchers call the enhanced representation of figures over the background FGM. Here, we studied the neuronal correlates of perceptual organization with electrophysiology in V1 and V4, using a homogenous texture as the neutral condition. We found, for the first time, that both figure enhancement and ground suppression contribute to FGM in both cortical areas. Figure enhancement occurred first in V4 and in V1, and after an additional delay the representation of the background was suppressed in both areas. The difference in the timing between figure enhancement and ground suppression implies that these mechanisms are at least partially independent, and our finding that enhancement and suppression were largely uncorrelated across recording sites in V1 and V4 supported this notion of independence.

Yet, figure enhancement and ground suppression were not dissimilar in all respects. We found that these processes had similar profiles across the cortical layers, with strongest effects on spiking activity in the superficial and deep layers and weakest effects in layer 4. Furthermore, figures led to increased sinks in layers 1, 2 and 5 and a stronger source in layer 6 than the uniform
texture, and similarly, uniform textures lead to increased sinks/sources in these same layers when compared to backgrounds. Layers 1, 2 and 5 are the targets of feedback connections from higher visual areas, in particular V2\textsuperscript{32-34}. This result, therefore, suggests that feedback projections are most active in the figure condition, less active in the uniform condition and least active in the background condition. We note, however, that these laminar profiles are the result of subtracting the CSD in one condition from that in another condition. Thus, these data are also consistent with the alternative hypothesis that the background causes strong sources in layers 1, 2 and 5, combined with a sink in layer 6. Yet, we do favour the first hypothesis because the background suppression requires the integration of information across large regions of the visual scene\textsuperscript{35}. Neurons in higher visual areas seem a likely source for these suppressive feedback effects, because they have large RFs and they send feedback to layers 1, 2 and 5 where we found stronger sinks in responses to figures.

The role of response enhancement and suppression in perceptual organization

A number of previous studies investigated the influence of perceptual organization on neuronal activity in the visual cortex. Previous fMRI studies reported that the representations of figures are enhanced\textsuperscript{11}, that the representation of the background is suppressed\textsuperscript{12} or a combination of both effects\textsuperscript{15}. Important questions were left open by these fMRI studies because they could not separate the representation of figures and background in the higher visual areas and fMRI may not distinguish between variations in firing rate of the neurons and changes in synaptic input\textsuperscript{21,36}. 
Chapter 4

Of particular relevance is a previous study that studied spiking activity in V1 and V4 of monkeys during perceptual organization\textsuperscript{13}. The monkeys had to identify a target string of colinear line elements among irrelevant background elements. The string evoked enhanced V1 activity with a latency of around 95 ms and the activity elicited by background elements was suppressed approximately 20 ms later. This study also demonstrated that V4 responses elicited by the string were enhanced after 59 ms, but again, the V4 responses to the background elements were not measured separately. Gilad et al. (2013) used a similar task design and monitored neuronal activity in V1 with voltage-sensitive dye imaging\textsuperscript{14}. Again, neuronal activity elicited by the string was enhanced and activity elicited by the background elements was suppressed, but this study did not report a significant difference in latency between enhancement and suppression.

Different processing phases during texture segregation

To enhance our understanding of the processes responsible for perceptual organization, we here capitalized on the texture-segregation task (Fig. 4). We obtained evidence for a rule of thumb where the latency of an effect on the activity of a V1 cell depends on the relevant spatial scale\textsuperscript{16}. The neuron’s first spikes code the features in its receptive field, including local line orientation (phase 1 in Fig. 4). Early contextual effects near the boundaries between figure and ground follow, and they cause a local enhancement of activity (phase 2). The next phase is the enhancement of the representation of the figure centre, involving the integration of features across a few degrees of visual angle (phase 3). In the last phase, figures that are many degrees away from the receptive field and that can even be in the opposite hemifield suppress neuronal activity (phase 4).
Figure 4 | Time-course of figure-ground segregation in V1 and V4. First, image features (the local orientation of line elements) are registered. In the second phase, boundaries are detected through a local inhibitory interaction. V4, with its large RFs, represents the figure with a lower spatial resolution than V1 so that edges are more diffuse and the representation of the centre of the figure is enhanced at an early point in time (unlike in V1). Third, responses elicited by the centre of the figure are now also enhanced in V1. The laminar V1 profile is consistent with a feedback influence from higher visual areas (black arrow). Fourth, the presence of a figure elsewhere appears to reduce feedback from higher visual areas into layers 1, 2 and 5, resulting in a relatively global suppression of the background representation (white arrows).

The phases of boundary detection, region filling and late suppression (phases 2-4 in Fig. 4) require different computations and thus rely on different neuronal mechanisms. Figure boundaries can be detected by local inhibition between neurons with nearby RFs tuned to the same orientation\textsuperscript{37-39}. This suppression is present at an early phase of the response\textsuperscript{40-43}, and is strong in image regions with a homogeneous orientation and weaker at figure boundaries. It can, therefore, explain the early response enhancement at figure boundaries in V1\textsuperscript{44} and V4\textsuperscript{48} as the relative lack of suppressive influences from neurons tuned to the same orientation. Higher areas represent the figure and
its boundaries at a coarser resolution (Fig. 4). Pop-out can occur in these areas when the neurons’ receptive field covers the figure so that neurons in the surround tuned to the same orientation are not well driven and provide only little inhibition. It seems likely that this early enhancement of the representation of boundaries is related to ‘border-ownership’ signals in V1, V2, and V4, which code the side of edges that belong to the figure\textsuperscript{45,46}.

The next phase is region-filling (figure-enhancement in Fig. 4). Now also image elements that are in the centre of the figure are labelled with enhanced neuronal activity\textsuperscript{8}. It is likely that this phase relies on an excitatory top-down effect, from neurons in higher areas that represent the figure with extra activity to neurons in lower areas tuned to the same orientation\textsuperscript{8}. This feedback scenario is in accordance with the earlier emergence of figure enhancement in V4 than in V1 during texture segregation\textsuperscript{8} and contour detection\textsuperscript{13}. The present results confirm that region filling increases neuronal activity over the level elicited by homogenous textures.

After yet an additional delay of about 50 ms, neuronal activity elicited by the background in V1 and V4 neurons is suppressed by figures that are far from the receptive field. This 50 ms delay is longer than the 20 ms delay observed in V1 by Chen et al. (2014)\textsuperscript{13}. This difference is in accordance with the rule of thumb mentioned above, because the neurons’ receptive fields were farther from the figure in the background condition of the present study than in Chen et al. (2014)\textsuperscript{13}. Furthermore, we here report that the pattern of enhancement followed by suppression also occurs in V4 and that the suppression in V1 and V4 occurs at similar time points. This initial focal response enhancement (Fig. 4, phase 3) followed by delayed global inhibition (Fig. 4, phase 4) could be a general principle that appears to hold true across visual tasks and visual cortical areas. On the one hand, we measured
enhancement and suppression for the same recording sites and observed that the strengths of these two effects are independent. On the other hand, the laminar profile of MUA and the CSD was similar for enhancement and suppression and suggested that both effects represent influences of feedback from higher visual areas. We mentioned above that we favour the interpretation that excitatory feedback from higher areas is highest for figural image elements, weaker for elements of a homogeneous texture and weakest for the background. In this view, figure detection would boost representations in higher visual areas and cause extra excitatory feedback at the figure location in early visual cortex, while reducing excitation at other locations, thereby causing ground suppression.

Attention and figure–ground modulation

Some of the processes for texture segregation are related to selection by visual attention, although we did not explicitly test the distribution of attention in the present study. In a previous study, we demonstrated that the early phase of texture-segregation that gives rise to pop-out and boundary detection is largely stimulus-driven, but that the later region filling process that labels the centre of the figure with enhanced activity is reduced if the animal directs attention elsewhere. This labelling process appears to correspond to object-based attention that is directed to all image elements of the figure. It is therefore conceivable that the late suppression caused by a figure far from the RF is the consequence of a shift of attention away from the ground region and towards the figure. Such a sequence of events would be compatible with studies showing that shifts of attention start with increased activity for the newly attended item followed by a decrease in activity for non-attended items. It is therefore of interest that FGM in Experiment 1 also occurred in the first period when the
monkeys could ignore the stimulus. We note, however, that we cannot exclude the possibility that the animals directed attention to the figure, because a similar figure had to be selected for an eye movement at a later point in time.

Conclusion and outlook

The present results combined with previous work demonstrate that texture segregation relies on a number of different processes that unfold at characteristic time scales. An important goal for future research will be to delineate these distinct processes at the columnar and cellular level. Work in mouse visual cortex has begun to provide insight how different cell types – in particular interneurons – provide a specific contribution to some of these processes. For example, surround suppression is mediated by SOM cells\textsuperscript{51}, and feedback connections target SOM cells to increase this suppression and VIP cells to cause disinhibition\textsuperscript{52}. It is therefore tempting to speculate that boundary detection in the present texture-segregation task depends on SOM cells, with a later top-down input to VIP neurons causing disinhibition for region filling and an even later top-down input to the SOM cells for ground suppression. These separate contributions of different interneuron circuits might also account for the independence of the strength of enhancement and suppression across neurons. Unfortunately, the specific contributions of the different interneuron types in the primate system are less well understood. We anticipate that important progress in this domain can be made with the design of new behavioural paradigms for mice and with the development of transgenic monkeys where the role of the different cell types can be tested during visual perception\textsuperscript{53}. 
References

Chapter 4


Figure enhancement and ground suppression during FG segregation

Chapter 4


Chapter 5

The segmentation of proto-objects in monkey primary visual cortex

Matthew W. Self, Danique Jeurissen, Anne van Ham, Bram van Vugt, Jasper Poort, Pieter R. Roelfsema

This chapter has been submitted to Current Biology
Abstract

To correctly perceive the visual scene, the brain must enhance image regions that belong to figures and suppress those that belong to the background. Natural images contain many regions that appear to be part of a figure when analyzed locally (proto-objects), but which are actually part of the background if the whole image is considered. To understand how the brain resolves conflicts between local and global processing we recorded neuronal-activity from V1 of macaque monkeys while they discriminated between N/U shapes that have a central ‘proto-ground’ region. V1 activity was enhanced on the figure surface, but was suppressed on all background regions, including the proto-ground, within ~115 ms. Suppression of the proto-ground was only present in animals that had been trained to perform the shape-discrimination task and activity on the proto-ground could be used to predict the choice of the animal, suggesting that proto-ground suppression in V1 contributes to shape-discrimination performance.
Introduction

The primate visual system is able to rapidly segregate the visual scene into figure and background, a critical step in the recognition of objects. This complex process involves combining local cues about the locations of edges with Gestalt and contextual cues from the entire visual scene\textsuperscript{1,2}. In natural scenes this process is complicated by the presence of many local regions which have the properties of objects but may actually belong to the background when the global scene is considered (Fig. 1).

\textbf{Figure 1 |} Natural scenes contain many convex boundaries which, locally, indicate the presence of objects. The regions in red show examples of convex boundaries which do belong to objects, we refer to the regions enclosed by these boundaries as ‘proto-figures’. The regions enclosed in blue show examples of convex boundaries which actually enclose background regions. We refer to these regions as ‘proto-grounds’. Correctly grouping the proto-figures into a single figure-representation and separating this from the background representation is critical for the recognition of (and localization of) objects.
These ‘proto-objects’ must be correctly assigned to figure or background to form an accurate representation of the scene, but the underlying neural mechanisms of this process are unknown. Studies of the primary visual cortex (V1) of macaque monkeys have suggested that the activity of neurons representing the surfaces of figures becomes enhanced relative to those representing the background\textsuperscript{3,4} after approximately 100 ms after stimulus onset. This enhanced activity, known as figure-ground modulation (FGM), suggests an important role for V1 in the segmentation of the scene and in our perception of figure surfaces, however these previous studies of FGM were typically conducted using small square figures for which the assignment of figure and ground is relatively unambiguous. This raises the question of whether V1 represents the global perceived figure-ground structure or local proto-object assignments which may not be reflected the global percept. In this study we addressed whether V1 represents global or local scene structure using complex textured figures (Fig. 2a), which form N or U shapes, along with checkerboard stimuli. These shapes cannot be discriminated through local mechanisms (Supp. Fig. 1) and contain a proto-ground region which has many object-like Gestalt properties but actually belongs to the background (Fig. 2b).

We studied the neural responses in naïve animals who had not performed shape-discrimination tasks before and then trained these animals to discriminate between N and U forms. Our results suggest that V1 activity is still modulated by figure-ground assignment when using more complex shapes and that proto-ground regions in V1 become suppressed down to the level of the background. Training and attention enhanced this modulation through different mechanisms leading to a spatially precise modulation of neural activity in V1 that strongly resembles our enhanced perception of figure surfaces\textsuperscript{5–7}.
**Figure ground modulation during proto-object segmentation**

**Figure 2 | The stimuli and task.** a, A screen-shot showing a full-screen texture containing the red fixation dot and a texture-defined N form. The form is outlined for clarity here (yellow lines not visible for the monkey). b, When viewed through an aperture of 8° diameter it is impossible to determine if the blue or red asterisk is located on the figure. Only when viewed through larger apertures is the true figure-ground assignment apparent. c, Animals were trained on an N/U discrimination task. They first attained fixation on the red fixation dot, after 300 ms the full-screen texture was shown containing
either an N, U, or checkerboard. After a further 400 ms fixation the fixation dot was extinguished and the animal was required to make an eye-movement to one of the two magenta targets (the shape/target association depended on the monkey), or to maintain fixation in checkerboard trials. d, The different conditions in the experiment were created by shifting the N/U form or checkerboard pattern relative to the receptive-fields of the V1 units to create the 5 main conditions. Each condition (except the proto-ground) consisted of a pair of locations on the left and right side of the figure. The checkerboard stimulus was used to control for local boundary processing and to extract an estimation of the response to the figure surface. e, After training both animals attained high levels of performance on the N/U discrimination task with performance on all conditions being significantly greater than chance (p < 0.05, binomial test). Both animals were worse at discriminating between the N/U for the background and outer-edge conditions due to the fact that these conditions required the N or U to be placed at greater eccentricities.

Results

Two macaque monkeys were trained on a shape discrimination task. The animals viewed a full-screen texture, which contained either a texture-defined ‘N’ or ‘U’ form or a checkerboard (Fig. 2a). N and U shapes contain proto-object regions which appear to be objects when viewed through apertures of greater than 4° and less than 14.4° in diameter, similar to the ‘field of view’ of neurons in early- and mid-level visual cortex (Fig. 2b, Supp. Fig. 1). We refer to the legs of the N/U as ‘proto-figures’ and the central background region as ‘proto-ground’ because these regions ultimately become assigned to different figure-ground compartments in perception. The animals were trained to make a saccade to a target depending on the perceived form (Fig. 2c). We examined multi-unit neural activity from chronically implanted arrays of electrodes in V1. We shifted the N/U form to different locations relative to the multi-unit receptive fields (RFs) to study the neural response to different parts of the
Figure ground modulation during proto-object segmentation

shape, or to the background regions. We created 5 main conditions in which the RF fell on the background, outer edge, proto-figure, inner edge, or proto-ground (Fig. 2d). Performance on the task was well above chance at all positions for all monkeys (Average N/U performance = 89.7% for monkey D, and 90.3% for monkey J, Fig. 2e). Figure 3a shows multi-unit activity from an example recording site for each of these 5 conditions. Neural activity at this site was modulated by the position of the figure in two clearly defined phases. In the early ‘peak’ phase (30-80 ms) responses were significantly higher when a textured boundary was present in the RF compared to the non-boundary conditions (two-sample t-test, p < 0.001). This significant early boundary modulation was present at almost all recording sites (two-sample t-test, p < 0.05, 77 out of 92 sites, Fig. 3b) and was significant when averaged across the population of recording sites (paired t-test, both monkeys p < 0.001). This early boundary modulation has been described previously⁸–¹¹ and is thought to arise through competition between orientation tuned cells within the superficial layers of V1¹⁰. Of particular interest to this study, in a later sustained phase (100-250 ms), responses at the example site were significantly higher on the proto-figure compared to the proto-ground (two sample t-test, p < 0.001) an effect we refer to here as proto-object modulation (POM) to distinguish it from standard figure-ground modulation. The relative increase of activity on the proto-figure was highly consistent across the population of recording sites in both monkeys. POM was significant in 73 out of 92 sites (two-sample t-test, p < 0.05) and was also significant across the population in each monkey individually (paired t-test, both monkeys, p < 0.001, Fig. 3c). The enhancement of proto-figures relative to proto-grounds was not related to differences in the position of the eye (Supp. Fig. 2). We examined the spatial properties of POM in more detail by taking advantage of the scatter of RF positions (Fig. 3d, Supp. Fig 3a-b) to produce interpolated maps of modulatory activity (Fig. 3e). These
maps show the change in V1 activity relative to the background condition (see Supp. Fig. 3c for further details). During the early phase, clear peaks were visible at the boundaries of the N/U figure. At later time-points the figure-surface gradually became filled in with increased modulation while responses on the proto-ground region were reduced back down to the level of the background leading to a clear neural segregation of proto-figures from proto-ground. The spatial precision of the neural modulation during the sustained period was remarkable (Fig. 3f) considering that this profile is unavoidably blurred by differences in the center of fixation on different trials (Supp. Fig. 2) and imprecisions inherent in measuring RF positions.
Figure ground modulation during proto-object segmentation
Figure 3 | MUA responses from V1 to the N/U stimuli during the shape-discrimination task. a, The response to the five main conditions from an example recording site in V1. The edge-conditions (magenta and orange lines) modulated during the peak phase (30-80 ms) whereas the proto-figure (red line) became enhanced relative to the proto-ground (cyan line) during the later sustained period (100-250 ms). b, The x-axis shows the responses during the peak phase from all V1 electrodes in conditions containing a boundary (IE and OE), the y-axis shows responses from all non-boundary conditions (BG, PF, PG). Solid symbols show significant recording sites (t-test, p < 0.05), almost all sites gave stronger responses to boundaries. c, Responses during the sustained phase for the proto-ground (x-axis) and the proto-figure condition (y-axis). All sites had stronger responses during the proto-figure condition. d, The scatter of the RF centers (black dots) from all V1 recording sites relative to the figure (outlined in grey). For monkey D we used a slightly larger vertical spacing of the figures, as there was more vertical scatter of the V1 RF centers. e, Using this scatter we created spatiotemporal maps of modulation. The 3D surface shows the enhancement (red colors) or suppression (blue colors) relative to the background. Stimulus onset is marked by the blue-yellow bars, which also shows the z-axis scale (the height of the bar is 0.02 normalized units). The timescale in milliseconds is marked by the bar on the y-axis. f, The fine-grained spatial profile of activity averaged between 100-250 ms from the data shown in (e). The grey areas mark the location of the proto-figures.

Enhancement or suppression?

The relative increase in activity on the proto-figure compared to the proto-ground could be due to enhanced activity on the proto-figure, suppression of the proto-ground, or both. To discriminate between these alternatives, we created a matched checkerboard control for each N/U stimulus. Checkerboards have clear boundaries that result in perceptual segregation, but there is no global assignment of particular checks to figure or background making them ideal as a reference stimulus. A matched checkerboard was constructed for each
stimulus position so that the boundaries of the checkerboard exactly coincided with the boundaries of the N/U forms (Fig. 4a). This allowed us to subtract boundary related activity from the response to the N/U to isolate the modulatory figure-ground assignment signal. To keep task demands the same for checkerboards and N/U forms we used a passive fixation task in which the monkey had to maintain fixation for 400 ms for all conditions to gain a reward. Figure 4b shows a comparison of the spatial profile of modulation for N/U forms and checkerboards. Compared to the checkerboard, proto-figures were relatively enhanced (red areas in Fig. 4c) whereas the proto-ground was relatively suppressed (blue areas in Fig. 4c). The strength of these two effects was not correlated across neurons (Monkey D: Pearson's r = -0.03, Monkey J: r = 0.04, both p > 0.05) suggesting that enhancement and suppression may arise through separate processes. The time-course of enhancement/suppression (Fig. 4d) revealed the presence of a non-specific enhancement of both proto-figures and proto-ground relative to their checkerboard controls at around 100 ms (black arrow in Fig. 4d). This 'bump' occurred at the time that FGM typically arises in simple figure-ground tasks\textsuperscript{8,10} and appeared to be a coarse form of figure-ground segregation that did not distinguish between proto-figures and proto-grounds, but increased activity generally in the locality of the N/U. The difference in response between proto-figures and proto-grounds first arose slightly later at 116 ms (Fig. 4d).
Figure 4 | Comparisons with the checkerboard. 

a, We used a texture-defined checkerboard as a reference stimulus to judge figure-enhancement and ground-suppression. The checkerboard contained similar boundaries to the N/U shapes allowing us to subtract the boundary-related responses. 

b, The spatial profile of activity (100-250 ms) in V1 for N/U stimuli and checkerboards. Zero on the y-axis is the average activity on the background of the N/U stimulus. 

c, The difference between the spatial profiles for N/U stimuli and checkerboards revealed regions of figure-enhancement and ground suppression. 

d, The time-course of figure-enhancement and ground-suppression. The arrow marks a period during which responses were enhanced relative to the checkerboard for both the proto-figure and proto-ground regions. The asterisks mark samples for which there was a significant difference between the two effects (paired t-test, $p < 0.05$). Latency was calculated as the time of the first significant sample that was followed by 10 further significant samples. The shaded region is +/- 1 s.e.m.
The relationship to behavior

Suppression of proto-ground regions may be critical for accurately perceiving the shape of objects in the visual scene. It is unknown however, whether modulation of neural activity in V1 contributes to decisions on shape-discrimination tasks. We investigated whether it was possible to predict the accuracy of the monkey’s decision using activity from V1 multi-units on individual trials. Suppression of the proto-ground was the most critical factor in determining the monkey’s performance: neural activity on the proto-ground showed a significant negative relationship with the accuracy of the monkey’s choice (Fig. 5a, median choice probability = 0.46, p < 0.001, Wilcoxon signed-rank test). This suggests that trials on which the monkey failed to suppress the proto-ground in V1 were more likely to result in errors, accordingly neural responses to the proto-ground were significantly higher on error-trials (Fig 5b, p < 0.001, t-test). Neural activity on the inner edge also showed a weak negative relationship with performance (median choice probability = 0.49, p = 0.03). Interestingly, activity on other regions of the figure, or on the background, could not be used to predict performance (Fig. 5a, all p > 0.05) and no significant differences in neural activity were observed between correct and error-trials for these regions. Does this mean that the relative enhancement of activity on the proto-figure was irrelevant for the monkey’s performance? The only difference between an N and U stimulus is the position of the horizontal segment that links the two legs of the figure (the ‘discriminant segment’, Fig. 5c). In monkey J we examined whether the quality of segmentation of the discriminant segment in V1 was related to the behavior of the animal. We moved the N/U figures so that either the discriminant segment or the corresponding ground region was in the RF (Fig. 5c). Behavior at these positions was similar to behavior for the main conditions (Fig. 5d). We then compared figure-ground modulation signals on correct trials to incorrect trials.
where the animal reported seeing the wrong shape (Fig. 5e). We observed significantly stronger responses when the RF was on the discriminant segment than when it was the equivalent ground position on correct trials (average difference = 0.07 n.u., t-test, p < 0.001). On error trials we observed much weaker modulation (average level = 0.01 n.u.), which was not significantly different from zero (t-test, p = 0.23). As there were less error trials than correct trials we created a ‘matched’ condition in which we resampled correct trials so that the same number of trials were included as error trials for statistical comparison (see Materials and Methods). We then directly compared these conditions statistically and found significantly greater levels of modulation on matched correct trials than on error trials (paired t-test, p < 0.001) (Fig. 5e). Note that the modulation was weaker on error-trials, but it was not reversed. This suggests that the incorrect shape discrimination was due to a failure to correctly segregate the discriminant segment rather than perceiving an illusory segment at the wrong position. In support of this view, the animal’s performance could be predicted with above-chance levels of accuracy when the receptive-field was on the figure (median choice probability = 0.54, Wilcoxon signed-rank test: p = 0.002) but not when it was on the ground (median choice probability = 0.48, p = 0.6). These results suggest that accurate performance on this task contains contributions from both suppression of the proto-ground and enhancement of the discriminant segment.
Figure ground modulation during proto-object segmentation

**Figure 5** | Correlations of neural activity with performance. **a**, Measures of choice-probability, measured as the median area under the receiver-operator characteristic curve (auROC) from different parts of the N/U stimulus. auROC values were significantly lower than 0.5 for the proto-ground meaning that higher activity on this region was associated with more shape-discrimination errors. Higher activity on the discriminant segment (see below) was associated with more correct trials. Data from the 5 main conditions comes from both animals, the discriminant segment was only tested in monkey J. **b**, MUA responses from three different parts of the N/U stimulus on correct and error
trials. Responses were higher on error-trials for the proto-ground region. c, To discriminate between the N and U the monkeys needed to locate the discriminant segment, connecting the two legs of the figure. We studied V1 responses on this segment in monkey J by placing it (or the equivalent location from the background) in the RF. d, Performance for the discriminant segment conditions. e, MUA responses from the discriminant-figure and discriminant-ground when the animal correctly identified the shape, or made a mistake. The data in the right panel come from the same number of correct trials used to make the error averages in the middle panel (Materials and Methods).

The effect of learning on figure-ground segmentation

Does POM arise through extensive training with the N/U forms? Prior to training the animals on the shape-discrimination task we recorded neural activity in V1 while the animals were naïve and performed the passive fixation task. The animals had been trained in other visual tasks but they had never seen the N/U textured stimuli before or performed any shape-discrimination tasks. We then trained the animals (36 days, Monkey D; 65 days Monkey J) until they reached high levels of performance on the N/U discrimination task. After training we returned to the passive fixation task, allowing us to directly compare the effect of training on neural activity while holding the behavioral task constant.

Neural responses in V1 were significantly modulated by figure-ground compartment in the naïve animals. We observed significantly elevated responses on the proto-figure compared to the proto-ground in both animals (Fig. 6a, t-tests, both animals: p < 0.001). Comparisons with the checkerboard control (Fig. 6b, left) revealed that this difference was entirely driven by enhancement of the proto-figure (t-test, p <0.001), we saw no evidence for proto-ground suppression in the naïve animals (t-test, p = 0.06). The extensive
training with the N/U forms produced changes in the representation of the surface in V1 compared to the naïve animals (Fig. 6b, right). Notably, the level of proto-ground suppression was greatly enhanced in trained animals (arrowed in Fig. 6b) reducing the response to the proto-ground down to the level of the response to the background as an effect of learning. For statistical analysis we subtracted the response of the checkerboard and split the data into the five main conditions to examine the effect of training in a 2x5 repeated measures ANOVA (Fig. 6c). We observed a significant interaction between condition and training (ANOVA. $F_{2.9,266.6} = 20.5, p < 0.001$), this effect was driven by the decreased responses on the proto-ground (post-hoc t-test, $p < 0.001$) and background (post-hoc t-test, $p = 0.001$) after training. We therefore suggest that the visual system becomes better able to suppress regions of the background through learning.
Figure 6 | The effects of shape-discrimination training on the representation of the proto-object. a, Responses on the proto-figure and proto-ground in naïve animals, conventions as in Fig. 3c. b, Spatial profiles of the difference between the N/U figures and checkerboards in naïve and trained animals (100-250 ms). The arrow marks the increased proto-ground suppression observed in trained animals. c, Response differences between the N/U and checkerboard grouped into the five main conditions for statistical analysis. Training significantly reduced responses on the proto-ground and background. Error-bars show ±1 s.e.m. Asterisks mark significant differences between conditions (post-hoc t-tests, Bonferroni corrected, p < 0.001).
The role of attention in segmentation

We examined the role that attention plays in segmentation of proto-objects by presenting a second, highly visible, colored N/U stimulus in addition to the texture stimulus. The animals were retrained to ignore the textured stimulus and perform the N/U discrimination task on the colored stimulus (Fig. 7a). In this way, we could compare neural activity when the animals were attending to the textured form to when they performed exactly the same behavioral task but on a separate object. We verified that the animals could successfully perform the task; average performance for the two monkeys was 74.4% (monkey D) and 88.6% (monkey J), and there was no effect of the spatial position of the texture defined N/U on performance (Fig. 7b) indicating that the monkey used the luminance defined N/U for his decision. There remained a significant level of POM when the animals attended away from the textured form (Fig. 7c, t-test, p < 0.001, both monkeys). However, there were significant changes in the figure representation depending on the location of attention as revealed by a significant interaction between position and attention in an across-phases ANOVA (F_{1.5,137.3} = 23.1, p < 0.001), Fig. 7d. When the monkeys were attending to the textured figure, the representation of the proto-figure was enhanced compared to when attention was directed away from the figure (post-hoc t-test, p = 0.03), and the background was more strongly suppressed (post-hoc t-test, p < 0.001). These two effects combined to produce a greatly increased neural segmentation of the figure compared to the background under attention, however attention had no effect on the suppression of the proto-ground (post-hoc t-test, p > 0.5). These differences can be seen by comparing the spatial profile of modulation across the figure in the presence or absence of attention (Fig. 7e). The results indicate that attending to the figures enhances segmentation by enhancing responses to the figure surface and suppressing the
background, however, unlike the training process, attention does not affect suppression of the proto-ground.

**Figure 7** | **The effects of withdrawing attention from the figure.** a, In this version of the experiment animals were required to judge the shape of a color-defined N/U form that appeared in the upper-left hemisphere and ignore the texture-defined form presented at the location of the receptive field. b, Behavior from the two animals during this phase of the experiment, note that the position of the texture-defined figure had no effect on performance. c, Responses during the sustained period for the proto-figure and proto-ground, conventions as in Fig. 3c. d, Statistical analysis of the response difference between the N/U and checkerboard stimulus. Conventions as in Fig. 6c. * = p < 0.05. e, Detailed spatial profiles of the activity during the sustained period for the N/U stimulus under attended (left panel) or non-attended (right panel) conditions. Here, the data from each
behavioral epoch were re-referenced to the average response to the proto-ground (which was unaffected by the attentional manipulation) so that the enhanced response on the near background when attention is directed away from the figure can be seen.

Discussion

The visual system implements the Gestalt rules of perceptual organization to group together parts of the visual scene that belong to the same object and segregate these from their background\textsuperscript{6,12,13}. Local regions of the visual scene may contain evidence for the presence of an object, such as a convex contour, leading to the formation of a proto-object representation: an early neural representation of an object that is not yet fully organized. Correct segregation of the visual scene is complicated by the presence of proto-objects which, locally, contain many of the Gestalt properties that define figures, but which actually belong to the background. The visual system must therefore utilize more global contextual information from across the visual scene to determine whether a particular proto-object belongs to a figure or the background. Proto-objects which belong to the background (which we refer to as ‘proto-grounds’) must be segregated from proto-objects which belong to figures (‘proto-figures’) to accurately determine the shape of objects and guide behavior. We tested whether neural responses in primary visual cortex differed on proto-grounds and proto-figures using relatively large N/U shapes. The interior of the N/U forms a proto-ground, it is surrounded by boundaries on three sides, has a high feature contrast with surrounding regions and has convex corners, all cues which are typically associated with figures in natural scenes. Furthermore, the proto-ground was visually very similar to the proto-figures which formed the ‘legs’ of the N/U forms, and to disambiguate these two regions the visual system would have to integrate visual information over regions of approximately $11^\circ$-
15° (Fig. 2b, Supp. Fig. 1). All these facts combined mean that neurons in V1 cannot use local computations to solve the figure-ground segregation problem for N/U forms. Nevertheless, we observed a very reliable modulation of activity on the proto-figure compared to the proto-ground in V1. Our results are summarized in Figure 8. Neural activity within a trial progressed through three phases: an early increase of activity at the locations of the texture-defined boundaries (Fig. 8a), followed by a coarse form of figure-ground segregation at approximately 100 ms (Fig. 8b) in which activity was generally increased on the proto-figure and proto-ground. Finally, after ~115 ms, activity was enhanced on the proto-figure and suppressed on the proto-ground, leading to clear neural segmentation of the N/U form in V1 (Fig. 8c). This late modulation of activity was strengthened by training on the shape-discrimination task, notably the suppression of the proto-ground was absent in naïve animals (Fig. 8d). Shifting attention away from the N/U weakened the neural segmentation, mainly by releasing the background from suppression (Fig. 8e).
Figure 8 | Graphical summary of the results. a, During the peak phase neural activity is strongly driven by feedforward input and the responses on the textured boundaries are enhanced. b, During an intermediate phase at approximately 100 ms responses are increased on both the proto-figure and proto-ground leading to a coarse enhancement of the figure. c, In trained animals, activity in the sustained period is enhanced on the figure surface and suppressed on the background and proto-ground region producing robust neural segmentation of the figure surface in V1. In naïve animals d, the suppression of the proto-ground is not complete, producing a weaker neural segmentation. When attention is directed elsewhere e, the suppression of the near background is not complete, reducing the relative enhancement of the figure.

Possible mechanisms of proto-object modulation

Previous studies of figure-ground segregation in monkeys have revealed a modulation of neural activity in V1 depending on whether the cell is responding to figure or background texture elements. It is likely that these modulatory effects are due to feedback from higher visual areas. Lesions of extra-striate
areas lead to reductions in the level of figure-ground modulation and a decrease in performance in detection of texture-defined stimuli\textsuperscript{14–16}. Furthermore, figure-ground modulation is strongest in feedback-recipient layers of V1 and the current-flow associated with FGM in V1 occurs in the feedback-recipient layers of V1\textsuperscript{10}. Border-ownership tuned cells, present in large numbers in higher visual areas such as V2 and V4 (Refs \textsuperscript{2,17–19}), are a likely source of feedback to V1. These cells give a stronger response when a figure is in a particular location relative to the edge in their receptive field. These cells can encode which proto-objects ‘own’ a particular border and previous studies have shown that they correctly report border ownership using N/U forms\textsuperscript{18}. Such cells would be ideally placed to inform V1 about the figure-ground assignment of different proto-objects. If border-ownership tuned cells send excitatory feedback in the direction of their preferred figure side, and suppressive feedback on their non-preferred side, then this could help V1 to disambiguate between proto-objects belonging to figure and ground leading to correct labeling of the figure surface\textsuperscript{20}. In line with this view the latency of border-ownership modulation is faster than that of figure-ground modulation and proto-object modulation\textsuperscript{18,21}. Computational models of scene segregation have shown that a recurrently connected hierarchical network of neurons in which neurons in higher levels are tuned for convexities in the image can reproduce border-ownership\textsuperscript{22,23} and figure-ground assignment\textsuperscript{24} signals using N/U forms. In these models, neurons at low and mid-levels in the hierarchy have small RFs and detect local convexities. However, initially the local convexities belonging to the proto-ground are also erroneously detected. Critically, model neurons at high levels in the hierarchy have large receptive fields which can correctly detect global convexities and overrule neurons at lower levels leading to correct segmentation of the scene. The results of these models, with initial enhancement of both proto-figures and proto-grounds,
followed by later selective enhancement of only proto-figures strongly resembles the spatiotemporal pattern of modulation observed in this study.

*Link to behavior*

Why does the visual system label the surfaces of complex figures with enhanced firing rates? After all, to perform this task correctly the animal simply had to extract the shape of the textured boundary. We have previously used tasks where monkeys had to make eye movements towards the center of a square figure to show that the spatial pattern of modulation in V1 may be read out by oculomotor structures to guide saccadic eye-movements. Supporting this view, behavioral studies in humans have demonstrated that eye-movements are guided to the centers of objects in natural scenes suggesting that the eye-movement system has access to the detailed spatial structure of objects. In this task the eye movement was not directed towards the figure but was determined by the outcome of a perceptual decision: was the figure an N or a U? Nevertheless, we observed that the level of spiking activity on the discriminant segment of the N/U could be used to predict the monkeys’ choice with above chance levels of performance. While the choice probability was only 54% this is similar to values reported in V2 for disparity judgements or in MT for structure-from-motion or motion-direction discrimination and it is remarkable that firing in this area contributes at all towards the perceptual decision in this task. These results raise the possibility that the perceptual decision is not based purely on boundary information present from early on in the response, but also on the surface label that is present in V1 at a later stage. In addition, the level of spiking on the proto-ground region was predictive of the monkey’s performance: trials in which activity was higher on the proto-ground were more likely to result in incorrect shape discrimination. This
surprising result suggests that suppressing the representation of potential objects in V1 is equally important to enhancing the figure surface. We speculate that by allowing perceptual ambiguities to be resolved through recurrent processing between boundary sensitive cells in higher visual areas and surface labelled cells in lower visual areas, the visual system may produce a more accurate discrimination for these complex shapes.

**Effects of learning and attention on figure-ground modulation**

In some previous studies the response enhancements of perceptually segregated objects were not observed in naïve animals\(^{29,30}\). We were therefore surprised to find a weak, but significant difference between responses to the proto-figure and proto-ground in the naïve animals. Even responses from monkey D, who had never before been exposed to texture-defined stimuli or performed any figure-ground task, were modulated by the complex figure. It is unclear why the results of these studies differ. We speculate that the fact that we presented the N/U at many spatial locations to map out the response to the full surface may have allowed the monkey to gain many views of the figure at different eccentricities and to rapidly form a template of its shape. We observed strong effects of training on the response to the proto-ground region of the figure. In naïve animals we did not observe suppression of the proto-ground, whereas suppression in the trained animal was at the same level as the background. It therefore appears that familiarization with the N/U shape and intensive training on the shape discrimination task allowed the monkeys to actively suppress proto-grounds more effectively.

Shifting attention away from the figure also changed the response to the N/U. Interestingly, these effects were quite different to those of training.
Withdrawing attention lead to a decreased response on the figure, as reported previously\textsuperscript{11}, and an increased response on the background. However, there was no change in the response to the proto-ground, which was equally well suppressed in both cases. The effect of attention therefore appears to be to enhance the difference in neural activity between figures and their backgrounds, without being sensitive to the fine local structure of the figures. Together these findings demonstrate that proto-figure enhancement and proto-ground suppression have different susceptibilities to learning and attention and may therefore be due to two separate neural processes. This idea is supported by the fact that the strength of these processes was not correlated across neurons. Computational models of figure-ground segregation have not typically focused on suppression of the background\textsuperscript{22–24}, but these results indicate that it may an important aspect of this perceptual process and an important target for future models.

**Online Methods**

*Training history and surgical details*

All procedures complied with the NIH Guide for Care and Use of Laboratory Animals, and were approved by the institutional animal care and use committee of the Royal Netherlands Academy of Arts and Sciences. Two adult male macaque monkeys participated in the experiment. One animal (monkey D, 5 years old) was completely naïve, the other (monkey J, 7 years old) had previously been trained for a figure-ground study with square stimuli\textsuperscript{11}. We implanted with a titanium head-post (Crist instruments) under aseptic conditions and general anesthesia as reported previously\textsuperscript{31}. The monkeys were trained to fixate on a 0.5° diameter fixation dot and hold their eyes within a
small fixation window (1.1° diameter). They then underwent a second operation to implant 4x5 and 5x5 arrays of micro-electrodes (Blackrock Microsystems) over opercular V1 (5 arrays in each monkey).11.

**Electrophysiology**

We recorded the envelope of multi-unit activity by digitizing the raw signal referenced to a subdural electrode at 24.4 kHz. The raw signal was then band-pass filtered (500 Hz-5 KHz) to isolate high-frequency (spiking) activity. This signal was rectified (negative becomes positive) and low-pass filtered (corner frequency = 200 Hz) to produce the envelope of the high-frequency activity, which we refer to as MUA. The MUA signal reflects the population spiking of neurons within 100-150 µm of the electrode and the population responses are identical to those obtained by pooling across single units.32,33

**Stimuli**

Stimuli were presented on a CRT monitor at a refresh rate of 85 Hz and with a resolution of 1024x768 pixels viewed from a distance of 52 cm. All stimuli were created using the COGENT graphics toolbox (developed by John Romaya at the LON at the Wellcome Department of Imaging Neuroscience) running in MATLAB (Mathworks Inc.). The stimuli consisted of full-screen textures composed of 13,000 oriented white lines (1.3° in length) drawn on a black background. Each day eight randomly generated textures were generated, four at 45° orientation and four at 135°. To generate the N/U figures, a portion of one texture was copied over a background texture of the opposite orientation. We chose the textures to generate the figure and background pseudo-randomly so that on average each texture was presented an equal number of times. The
N/U form consisted of two ‘legs’ of 8° in height and 4° in width, connected with a connecting segment of 2° in height. The central region between the legs was 4° in width, making the total figure size 12° x 8°. The connecting segment could be positioned at the top of the legs to create an N or at the bottom to create a U, we therefore refer to it as the ‘discriminant segment’. We also created a checkerboard texture which consisted of checks of 4° (width) by 6° (height) with alternating orientations. Across trials, we shifted the position of the N/U forms and the checkerboards horizontally in steps of 2° so that the RFs fell on different parts of the stimulus (thus creating 5 main conditions; Fig. 2D). To study responses to the discriminant segment, we shifted the N/U forms vertically (only in monkey J).

In the passive fixation task, the monkeys viewed the textures and maintained fixation for the duration of the trial. In the active version of the task the animals had to discriminate between texture-defined ‘N’ and ‘U’ forms by making an eye movement to one of two purple targets positioned at the vertical midline at 8° eccentricity. The monkeys viewed the textures for 400 ms at which point the fixation point was removed. Monkey J then had to make an upwards saccade for ‘N’s and a downwards saccade for ‘U’s, in monkey D the response mapping was reversed. On trials in which a checkerboard was presented the animals were rewarded for maintaining fixation for a further 275 ms after the fixation dot was extinguished. To study the effect of attention on figure-ground segmentation signals we presented an additional N/U presented in magenta (i.e. not textured) with a size that was 25% smaller. The magenta N/U was placed in the upper-left quadrant, at 4° eccentricity. The monkeys report its shape by making an eye-movement to the appropriate target, as in the main version of the task, while ignoring the texture defined N/U.
Chapter 5

Receptive Field Mapping

We mapped the receptive-fields of each multi-unit site in V1 using a drifting luminance-defined bar that moved in one of four directions. The borders of the RF were then calculated as described previously\textsuperscript{33}. The median RF size, taken as the square-root of the RF area, was 1.4° in V1 (range 1.0° to 3.1°) and the median eccentricity was 4.3° (range = 2.9° to 6.1°).

Data Analysis

The MUA data from each recording site was normalized on each recording day. We first produced a running estimate of the baseline MUA level by taking the mean response in each pre-trial period in which the animal was fixating (-200 to 0 ms relative to stimulus onset). The baseline response on each trial was then smoothed using LOWESS regression and a 101 trial smoothing window (MATLAB, Mathworks inc.). After subtracting the baseline response, we normalized all responses to the maximum smoothed (26 ms Gaussian kernel) peak response (in the time period 30-80 ms after stimulus onset) of the average response across all texture conditions. The data are therefore expressed in normalized units, i.e. a value of 0.1 indicates a response of 10% of the difference between the peak and the baseline.

We calculated a signal-to-noise ratio for each site by dividing the average peak response by the standard deviation of the baseline activity across trials. Only recording sites with an SNR > 4 and a well-defined RF were included in the analysis. After applying these inclusion criteria we analyzed data from 56 (out of 72) recording sites in V1 in monkey D and 38 sites (out of 42) in V1 in monkey J. For each of these sites we calculated the mean response across all recording days for each of the main conditions in the experiment, only those
conditions in which the figure was positioned appropriately relative to the RF were used to construct these means. For statistical analyses we took the mean activity of each site in two time windows: a peak window (30-80 ms) and a sustained window (100-250 ms). Note that while the animal was required to retain fixation for 400 ms, due to an error, Monkey D was able to make eye-movements after 200 ms and we therefore only analyzed data from 0-250 ms (allowing for the visual latency of responses in V1), and we used the same analysis window for the data of monkey J so that we could compare and pool the data across monkeys (the results remained the same if the 400 ms time period was used for monkey J). For across-sites analyses of the effect of spatial position we used repeated-measures ANOVAs with the data split into the 5 main conditions (illustrated in Fig. 2d). We compared stimulus conditions using post-hoc t-tests, corrected for multiple comparisons using the Bonferroni method. We compared different epochs of the task (effect of training; effect of attention) using a two-way repeated-measures ANOVAs with the factors position (5 levels) and epoch (2 levels).

To compute the spatio-temporal profiles of the neural response we first calculated the horizontal distance between the center of the N/U/checkerboard and the center of each RF for all the figure positions used in the experiment. These distances were binned into 0.75° wide bins. We extracted the MUA time-course at each spatial location and subtracted off the mean response from the background condition (i.e. the mean response from all trials in which the center of the figure was located more than 8° away from the RF center). We averaged all the responses from different electrodes within each bin and linearly interpolated the resulting response on a spatio-temporal grid (spatial resolution: 0.25°, temporal resolution: 2 ms). Only locations where the vertical distance of the figure center from the RF center was less than 1.5° were included in the map to exclude responses elicited by the internal boundary of
the discriminant segment. We averaged responses across N and U forms. For more details, see Supplementary Figure 3. For statistical analysis of the spatio-temporal maps (Fig. 4d and 5e) we regrouped the interpolated-data into the five main conditions according to the x-position of the RF. Data from -9° to -7° and 7° to 9° were assigned to the ground, from -7° to -5° and 5° to 7° to the outer edge, from -5° to -3° and 3° to 5° to the proto-figure, from -3° to -1° and 1° to 3° to the inner edge and from -1° to 1° to the proto-ground.

To compare neural activity of correct and error trials we resampled the correct trials (without replacement) to create a matched distribution with as many trials as there were error trials. To calculate choice-probabilities for each recording site we calculated the area under the receiver-operator characteristic curve based on the MUA response on individual trials averaged between 100-250 ms as described previously^{28,34}.

Eye movements were recorded using a digital-camera (Thomas recordings, 250 Hz frame-rate). For analysis of the eye data the position of the pupil was digitized and recorded at 500 Hz by the TDT recording system, the eye positions were low-pass filtered at 100 Hz and stored for further analysis.
References


2. Von der Heydt, R. Figure-ground organization and the emergence of proto-objects in the visual cortex. Front. Psychol. 6, 1–10 (2015).


Chapter 5


Figure ground modulation during proto-object segmentation


Supplementary Information

Analysis of feature contrast

Previous experiments examining figure-ground modulation with square figures could not rule out a possible contribution of feature-contrast to the modulatory signal. It is well-known that cells in V1 are suppressed by stimuli in their surround, and that this suppression is strongest when the surround has the same orientation as the center (Cavanaugh, Bair, & Movshon, 2002; Knierim & van Essen, 1992; Nelson & Frost, 1978). Large, uniform regions of orientation (such as those on the background) therefore experience greater suppression than small regions containing feature contrast (such as those on figures). To determine whether feature contrast could be used to assign figure and background with the N/U stimuli we simulated the stimuli used in this experiment. We created a matrix of pixels extending for 50° x 50° with a spacing of 0.1°. We created an N shape of the same dimensions as those used in the experiment. In different simulations the RF was centered either on the proto-ground, the proto-figure or, for comparison with earlier studies, a 4° square. Each pixel either had the same orientation as the center of the RF (coded as a 1) or the orthogonal orientation (0). We then used a series of circular apertures with diameter (d), where d ran from 0.4 to 40° in steps of 0.2° to measure feature-contrast at different spatial scales. We calculated feature-contrast (FC) as:

\[
FC(d) = \frac{\sum (1 - F_{RF})}{n}
\]

Where \( F_{RF} \) was the feature identity at each pixel that fell within the circular aperture of diameter, d and n was the total number of pixels inside the aperture. FC will have a value of zero when all pixels match the orientation at the center of the RF and a value of 1 when all pixels match the orthogonal
Figure ground modulation during proto-object segmentation

orientation. Values close to zero represent the situation typically seen on the background whereas values close to 1 would typically be found when the RF is situated on a small region of feature contrast embedded in a large uniform texture.

The results of the simulation show that for small apertures (diameter<11.1°) the proto-ground had a higher feature contrast than the proto-figure. So if feature-contrast was the only determinant of figure-ness then the proto-ground would be incorrectly assigned as a figure using apertures of this size. Beyond aperture diameters of 11.1° the feature contrast became larger on the proto-figure meaning that the visual system would have to integrate over regions of greater than 11.1° in diameter to assign figure-ground status correctly using feature-contrast with these N/U stimuli.

Analysis of eye position

We examined whether the crucial difference in V1 response between the proto-figure and proto-ground condition may have been due to differences in the mean eye position. For each electrode-array the trials that contribute to each condition vary due to the variation in the RF positions of the array. We therefore calculated for each array the mean eye position on trials where the proto-figure was centered on the RF and trials where the proto-ground was centered in the RF (Supplementary Figure 2). We compared the mean x and y eye positions on these trials during the fixation period (0-250 ms for the active task for Monkeys J and D) using two-sample t-tests. The results show that monkey D (4 arrays) had no significant differences in eye x or y position between the figure and center-ground conditions (all p > 0.2). In monkey J (2 arrays) the eye x position was significantly different for one array (p = 0.01), but not the other (p= 0.98). For the significant array the difference in mean x-position was small: 0.06° and, given that the neural difference between the
Chapter 5

figure and center was significant for both arrays, it is highly unlikely the neural differences observed were due to differences in the mean eye position.
Supplementary Figures

**Supplementary Figure 1** | Feature contrast (FC – see text for details) was calculated using apertures of different diameters centered on the center-ground (blue line) the figure (red line) or, for comparison to older studies, a 4° square (black line). With apertures of larger than 11.1° diameter the feature contrast became greater on the figure than the center ground.
Supplementary Figure 2 | The difference between proto-figures and proto-grounds cannot be explained by differences in mean eye position. The scatter-plots show the mean X and Y position of the eyes during the fixation period (0-250 ms) for trials in which the proto-figure (red dots) or proto-ground (cyan dots) were centered on the RF cluster. There were four different electrode-arrays for Monkey D, resulting in 4 clusters of RFs upon which the N/U figures were centered, these trials are shown in separate graphs here. Likewise, for Monkey J there were 2 arrays. The thick crosses show the mean x and y position +/- 1 standard deviation.
Supplementary Figure 3 | Generation of spatiotemporal maps of figure-ground modulation. 

**a.** Each array of micro-electrodes had a different average RF position (shown in cartoon form here as the colored squares labelled 1-4). The N/U stimuli were presented at different horizontal offsets relative to the average array RF center to create the different conditions. The left panel here shows the N/U centered on the average RF position of array #1, note that while the stimulus was well-centered for the targeted array, it was slightly shifted for the other arrays. The right panel shows an example of a stimulus centered on the average RF of array #2. For this hypothetical example, the average RF of array #1 was situated within 0.5° of the inner boundary of the N and this data would be excluded from the maps (note that while average array RFs are shown here for convenience, exclusion criteria were applied to individual electrode RFs). 

**b.** We calculated the horizontal distance between each N/U stimulus and the center of each RF. These distances were binned into 0.75° wide bins. The histogram shows the number of stimulus/RF pairs that contribute to the maps for different x-positions. The minimum number of stimulus/RF pairs in a bin was 22. 

**c.** Mean neural activity time-courses for each stimulus/RF were first smoothed with a 26 ms moving window and responses...
within a bin were averaged together. The resultant spatio-temporal MUA response is shown in the left panel. The map is dominated by the initial peak-response, making it difficult to see later differences between conditions. To correct for this issue we subtracted the mean response to the background (calculated as the average response from all RFs that were more than 8° away from the N/U center; shown in the mid-left panel) from each bin to create a map of figure-ground modulation (mid-right panel). We linearly interpolated the resulting map on a spatio-temporal grid (temporally: one grid-point every 2 ms, spatially: one grid-point every 0.25°). The final result is shown in the right panel; the colors indicate the difference from the background response in units of normalized MUA. This 2D map shows the same data as is shown in 3D in Fig. 3e.
Chapter 6

Discussion
As we are able to categorize a visual scene and make a decision about a stimulus within just one-tenth of a second, the capacity to visually perceive the external world might intuitively seem unremarkable. However, the underlying computational operations are extremely complex, and cases of visual agnosia and various attempts to develop computational models that mimic human vision for instance illustrate the astonishing capacity of the brain to represent the visual world. The capacity of the visual system to process all the information that reaches our retina is however also limited. We are not capable of processing the entire environment around us in a single take, and our brain therefore needs to be able to represent the visual world in a flexible manner to dynamically adjust to the features of the visual scene that are relevant for ongoing behavior. The computational operations to adjust to the ever changing behavioral demands are implemented in the brain by a series of processing stages along several cortical regions in a hierarchically organized system. By combining and transforming visual representations from lower areas at successive stages in very specific ways, this hierarchically organized visual system allows neurons in different areas to be specialized for a particular function. Neurons in V4 for instance contribute primarily during the detection of complex shapes, whereas other areas like V5/MT and areas in the parietal cortex are primarily involved in detecting moving stimuli and spatial relations between stimuli, respectively. This way these specialized regions can process different aspects of incoming visual information in parallel, and the hierarchical organized system can thereby cope with perceptual demands more efficiently. Although these specialized regions are particularly well suited to process specific aspects of the visual scene and are activated accordingly during the initial phase of visual information processing, which areas are involved and interact with each other during the later recurrent phase of visual information processing does not solely depend on
Discussion

the stimuli that are present in the visual scene. During this later phase of visual information processing the hierarchical organization of the visual system enables different brain areas to communicate with each other dependent on the context and specific behavioral demands, thereby allowing for flexible modulation of neural representations between the different areas along the cortical hierarchy. This communication between different areas during the later recurrent phase of processing is therefore thought to be essential for the implementation of various cognitive functions that depend on the flexible selection of information to adjust to specific circumstances. Although it is now widely recognized that different areas along the cortical hierarchy are involved in complex interactions during visual perception, the underlying computational operations and the dynamics along the cortical hierarchy during the implementation of various cognitive functions like visual awareness, working memory, attention and figure-ground segregation remain largely unknown. In this thesis we investigated how information is processed along the visual cortical hierarchy during visual perception by recording activity from neurons at different stages of the macaque visual system while the monkey performed complex visual tasks, all of which are thought to involve feedback processing across the visual cortical hierarchy.

First, in chapter 2 we addressed the question of how the brain gives rise to conscious experience. How awareness emerges in the brain is still heavily debated, but several studies\textsuperscript{6-8} strongly suggest that feedback from higher to lower areas in the cortical hierarchy is essential for perception. In this chapter we first showed that stimuli that reach awareness elicit a stronger initial feedforward response at all levels of the cortical hierarchy than stimuli that do not make it to awareness. We furthermore showed that this difference between the representations of perceived and non-perceived stimuli increased at higher levels of the cortical hierarchy and that the
intensity of the stimulus determines at which level of the cortical hierarchy the information gets lost, with stronger stimuli getting lost at higher hierarchical levels. Whether a stimulus will be consciously perceived or not therefore depends on the efficiency of feedforward propagation from lower to higher brain regions. Our experiments also showed that when a stimulus elicits sufficient activity in higher areas for the stimulus to be perceived so called ‘ignition’ takes place, in which conscious perception of a visual stimulus is associated with a later processing phase with enhanced sustained activity at all the stages of the cortical hierarchy. This sustained activity presumably reflects recurrent interactions between widespread brain regions that make the visual information globally available, and thereby enable the visual stimulus to be consciously perceived. One open question remains why sometimes a stimulus elicits sufficient activity to gain access to awareness, while at other times an identical stimulus does not elicit sufficient activity to gain access to awareness. This may depend, in part, on the state of the brain before the stimulus appears. To find out which factors determine whether a weak stimulus can gain access to conscious perception we together with the lab of Stefano Panzeri investigated whether pre-stimulus behavioral and neuronal markers of brain state predicted the perceptual outcome. Behavioral measures that were included consisted of the time before the monkey initiates a new trial by directing gaze to the fixation point after its appearance on the screen, the diameter of the pupil and the time-derivative of the pupil diameter. As direct measures for brain state we determined the pre-stimulus MUA and the power in the alpha, beta, and low and high gamma bands of the local field potential. These results show that these behavioral and neurophysiological measures of pre-stimulus cortical state have a weak relationship with perceptual outcome, generally in accordance with previous findings. A new finding is that a combination of markers enables better predictions with
Discussion

accuracies of 60-65%, and that combinations of these markers have separable influences on the animal's response bias and sensitivity. A bias to report target present was associated with a higher baseline firing rate across different brain regions, causing neurons to be closer to the threshold for ignition. In contrast, a higher sensitivity was associated with an improved propagation of neuronal activity to higher processing levels, thereby increasing the difference in activity levels between target-present and absent trials at the processing stage that determines the threshold for ignition. Additional measures of pre-stimulus brain state, not tested by us, might even further increase the predictive power. Although these results therefore suggest that pre-stimulus neural and behavioral markers can to a degree predict whether a stimulus is consciously perceived, part of the unpredictability is presumably intrinsic because it can be attributed to the intrinsic stochasticity of neurons at the different cortical levels.

In chapter 3 we investigated how sensory information is maintained in the brain when we have to remember something for a short period of time. What neural mechanism underlies this so called working memory remains unknown, but several observations suggest that NMDA channels might be crucially involved in maintaining the internally sustained activity that is observed in the PFC during working memory related processes, either through accommodating feedback processing\textsuperscript{12} or by optimizing the intrinsic dynamics of single neurons in these higher cortical regions\textsuperscript{13,14}. To test these theories we investigated the contribution of AMPA and NMDA receptors during spontaneous activity, the initial feedforward response and internally sustained activity during a task in which the monkey had to remember the spatial location of a stimulus for a short period of time. We showed that both AMPA and NMDA receptors contribute to persistent firing, and that NMDA receptors do not have a specific and critical role in persistent firing. Instead,
our experiments revealed a general multiplicative contribution of NMDA receptors to spiking activity which is similar for the initial feedforward response and during persistent activity. These results therefore suggest that the persistent activity that is observed in higher cortical areas during working memory related processes is maintained not solely by intrinsic dynamics of single neurons, but that internally sustained activity most likely involves recurrent processing during which reverberatory excitation (that is not specifically dependent on NMDA receptors) between neurons within a cortical area\textsuperscript{15,16} or reciprocal excitatory loops between (sub)cortical areas\textsuperscript{17-19} takes place.

In Chapters 4 and 5 we investigated how our visual system segregates objects from the background in a visual scene. Although several observations have shown that recurrent processes is involved in figure-ground modulation, it is still not known whether this modulation is due to feedback from higher visual areas\textsuperscript{20-23} or arises from purely local horizontal interactions between neurons within V1\textsuperscript{24}. In chapter 4 we assessed the contributions of figure enhancement and background suppression to figure-ground modulation by studying neuronal activity in areas V1 and V4 in monkeys performing a texture segregation task. We compared texture-defined figure-ground displays to homogeneous textures that lack figure-ground organization and found that both figure enhancement and ground suppression contribute to figure-ground modulation; the representation of figure elements was enhanced first in V4 and after a brief delay also in V1. After an additional delay the representation of background elements was suppressed. The laminar profiles of both figure enhancement and ground suppression were most pronounced in superficial and deep layers and was relatively weak in layer 4. Furthermore, the current-source density as a measure for the putative synaptic inputs that cause figure-ground modulation showed that increased
spiking activity was associated with stronger current sinks in feedback recipient layers 1, 2 and 5 for figure enhancement and weaker sinks for ground suppression. The laminar profile of figure-ground segregation therefore suggests that feedback connections from higher visual areas play an important role in figure-ground segregation and that this segregation relies on a number of different processes that unfold at characteristic time scales. Next, by using complex figures we in chapter 5 showed that figure-ground modulation cannot arise from purely local horizontal interactions between neurons within V1, and therefore most likely is due to feedback from higher visual areas. To surround the input to the receptive field of the neuron with edges that are matched for figure and background stimuli up to several visual degrees outside the receptive field of the neuron, we used textured backgrounds containing a complex object such as ‘n’ and ‘u’ forms. In line with previous results we found that neurons in V1 initially responded more strongly only if their RF was on the edge of the figure. This edge detection phase however was followed by a rapid modulation of the neural response across the entire figure, including regions in the center of the N/U which actually belong to the background. After approximately 100-150 ms from stimulus onset this central region became strongly suppressed and the figure label was restricted to locations occupied by the N/U. These results indicate that the visual system initially labels all regions surrounded by edges as figures for a brief period of time (50-100 ms), before an inhibitory form of feedback sharpens the spatial extent of the label so that it precisely matches the extent of the figure. Together with observations that the latency of the figure-ground modulation in V1 coincides with the later sustained response and that current-flow associated with figure-ground modulation occurs in the feedback layers of V1, these results indeed strongly support the idea that figure-ground segregation is not solved by purely local interactions between
Chapter 6

low-level neurons but instead is an evolving process which requires interactions between neurons at multiple levels of the cortical hierarchy. Although the described figure-ground modulation was already present in animals that had never seen the complex N/U objects, this process was strengthened by training and weakened when attention was directed to other objects. As learning and attention effects are generally thought to involve feedback processes, this furthermore suggests that feedback from higher areas influences low level representations on multiple timescales.

Together the studies that were covered in the chapters of this thesis strongly support the view that dynamic interactions between areas at different levels of the cortical hierarchy are essential for the implementation of cognitive functions that rely on flexible selection and maintenance of behaviorally relevant information. We showed that to select information and use it for cognitive behavior, the initial feedforward sweep needs to reach areas that are at the top of the visual hierarchy. Only if activity is sufficiently propagated to these higher visual areas a phase of “ignition” takes place. During this ignition phase, recurrent processing between lower and higher cortical areas takes place in which information is selected and its corresponding representations enhanced, thereby making the visual information globally available for cognition. How much the specialized areas along the cortical hierarchy contribute to information processing during perception depends on the level of correspondence between the functional role of each cortical area and its relevance to the problem that needs to be solved. If the behavioral task consists of the integration of a contour the behaviorally relevant modulation can be observed all the way down in V1, whereas neurons in area MT will modulate their activity if the direction of movement is important for solving a task. Similarly, if a task consists of localizing a red target, cells in a range of
Discussion

areas along the cortical hierarchy that are selective for red stimuli will enhance their activity\textsuperscript{29}. This task specific behavioral modulation of activity along the different areas and levels of the cortical hierarchy therefore strongly suggests that a given visual percept arises from the global set of cortical states and task-specific interactions between multiple areas along the visual cortical hierarchy, and not solely from the activity of a small number of cells at the top of the visual hierarchy as was traditionally thought. In this view, behavioral relevant modulation at lower sensory areas results from top-down information from higher cortical areas, thereby informing different levels at the cortical hierarchy which stimuli are important at a given moment in time. As the relevance of a stimulus can depend on many factors, these top-down influences are important during a wide range of cognitive functions, including spatial-, object- and feature oriented attention\textsuperscript{30-36}, reward\textsuperscript{37,38}, awareness\textsuperscript{8}, expectation\textsuperscript{39}, perceptual learning\textsuperscript{40}, working memory\textsuperscript{41} and figure-ground segregation\textsuperscript{42}.

Sources of top down modulation

The source through which top-down signals implement these different cognitive operations is still an area of active research. Traditionally the attentional control system for instance is thought to encompass a network of areas in parietal regions like LIP\textsuperscript{43} and frontal regions like FEF\textsuperscript{44} and the basal forebrain\textsuperscript{45}. Spatial- and object based working memory on the other hand has been localized in the dorsolateral\textsuperscript{46} and ventrolateral PFC\textsuperscript{47} respectively, while the reward system is thought to be implemented by a network spanning the mesolimbic dopamine pathway and prefrontal regions like the orbitofrontal cortex (OFC) and the insula\textsuperscript{48}. It is likely that in some cases the effects of different cognitive functions are related; a stimulus that is highly rewarding
for instance probably attracts more attention compared to a stimulus that is not rewarding, and, at least in experimental settings, stimuli that are attended are the ones that are rewarded\textsuperscript{49}. Furthermore, during cognitive functioning many of the cognitive control systems have to cooperate to select the appropriate information, and therefore cognition involves the participation of many different control- and neurotransmitter systems in a variety of brain regions. Using new optical and genetic techniques together with clever experimental paradigms, future studies could determine the source of the top-down selection signals in the brain during circumstances in which the effects of different cognitive control systems are related or interact, and determine how these interactions are implemented in the brain.

**Mechanisms of top down modulation**

Irrespective of the source of the top-down signal, all cognitive functions have in common that information needs to be flexibly selected and its corresponding representations in the brain enhanced.

*Changes in functional connectivity to enable dynamic and efficient integration of information as a function of cognitive state*

Top-down signals selectively gate connections along the cortical hierarchy that are appropriate for the task at hand. It has been shown for instance that depending on which stimulus components are relevant or irrelevant to the task being executed, feedback is able to change the stimulus selectivity of neurons in lower areas. McManus et al. (2011) for instance found that neurons in V1 that are selective for complex shapes change their stimulus selectivity according to the expected shape of the object\textsuperscript{39}. This way neurons change their function in a context-dependent manner under the instruction of feedback.
Discussion

from higher cortical areas, thereby changing the meaning of the information they carry. Because the higher cortical areas that interpret the information from lower areas are also the areas that sent the instruction to these lower level neurons to perform a particular calculation, the return signal is “interpreted” by the higher areas as the result of that calculation and is therefore not confused with other operations those neurons perform.

Besides changes in stimulus selectivity of individual neurons, Cohen & Newsome (2008) proposed another mechanism through which top-down signals change functional connectivity depending on the behavioral context\(^50\). They showed that pairs of neurons in MT change the degree of noise correlation depending on whether the animal was attempting to detect motion in the direction shared by the RF properties of the two neurons (thereby promoting cooperative interactions) or in an orthogonal direction (thereby promoting competitive interactions). The finding that noise correlations between pairs of neurons were higher when the animal was attempting to detect motion in the direction shared by the RF properties indicates that single cells are able to rapidly and flexibly participate in different ensembles of interacting neurons depending on the behavioral context, even in the presence of identical visual stimuli (see also ref\(^51\)). Although these studies showed that top-down biasing signals increase noise correlations among neurons with similar tuning, attention has been argued to reduce noise correlation between neurons\(^52,53\). Theoretical studies have argued that noise correlation can either increase or reduce the available information, depending on the algorithm by which neural responses are read out\(^54-56\). Increased noise correlations among neurons with similar tuning would reduce the reliability of the pooled responses because the correlated variability also enters in their average response\(^57,58\), whereas noise correlations might be beneficial for population read out when neurons are tuned to different features. By taking
the relative position of the neurons RF into account, Poort & Roelfsema (2009) argued that these opposing effects cancel each other out at the population level, so that the net effect of the noise correlations was negligible and noise correlations therefore have little influence on the coding of selective attention in area V1\textsuperscript{59}.

Correlation at a finer timescale, shorter than the integration time constant of the neuron, in which not the average firing rate but the relative timing of individual spikes is the critical factor, has also been suggested to increase the available information in the neural code. This way, through the formation of specific ensembles, synchrony has been proposed to bind features from the same object together\textsuperscript{60} and improve cross-area communication by optimizing the postsynaptic impact of spikes from one area upon the other\textsuperscript{61}. Although many studies reported that attention increases synchrony among local populations of neurons\textsuperscript{62} and neurons between two areas\textsuperscript{63}, this information has been shown to be irrelevant for binding\textsuperscript{64}. Whether synchrony between local populations or between neurons in different areas can be used by the brain to improve cross-area communication or for other computational purposes however remains unclear\textsuperscript{65,66}. How top-down effects exactly affect noise correlations and synchrony during different cognitive operations and how these effects influence the available information in the neural code to improve performance is therefore still not fully understood, and future studies investigating population coding should determine how interactions between neurons can have critical effects on behavior.

Besides affecting noise correlations and synchrony, attention has also been found to change functional connectivity through alterations of the size of the RF. By showing for instance that attention reduces the RF size of MT neurons, Womelsdorf et al. (2006) argue that highly modifiable RFs enable
the dynamic allocation of processing resources to attended locations, thereby supporting enhanced perception within the focus of attention by effectively increasing the spatial resolution\textsuperscript{67}. Furthermore, attention has also been shown to affect center-surround interactions, increasing collinear facilitation by lines outside the RF when attention is directed towards the RF\textsuperscript{26,68}. These contextual influences have been proposed to be mediated by long range horizontal connections within each cortical area, with top-down effects being responsible for the gating of these horizontal connections\textsuperscript{69}. Recently the lab of Yang Dan showed that surround suppression in V1 is indirectly caused by feedback connections from the PFC by the activation in V1 of a specific type of inhibitory neurons called somatostatin cells\textsuperscript{70}, thereby exposing a mechanism that selectively expresses functional connectivity of local horizontal connections through top-down influences from higher areas.

Together all these examples show that top-down signals implement great flexibility in adaptive tuning and functional connectivity within the cortical hierarchy, thereby enabling dynamic and efficient integration of information as a function of cognitive state.

\textit{Increases in firing rate to enhance behaviorally relevant representations}

In our studies we showed that neurons that represent behaviorally relevant information increase their firing rate during the later sustained response. Whether sensory information needs to access awareness (chapter 2), has to be maintained in working memory (chapter 3) or needs to be segregated into figure and background components (chapters 4 and 5), all these cognitive operations cause the top-down selection signal to increase the firing rate of neurons that represent information that is relevant for the task at hand. Recently, work in mice has begun to uncover potential mechanisms by which feedback would be able to modulate activity in lower areas. Beltramo et al.
(2013) for instance found that by selectively activating pyramidal cells in layer 5 of mouse visual cortex, they could strongly enhance the activity of neurons in the rest of the cortical column\textsuperscript{71}. The opposite effect was found by the lab of Scanzianani, who showed that activating pyramidal cells in layer 6 of mouse visual cortex suppresses neuronal activity in the rest of the column\textsuperscript{72} via fast spiking inhibitory neurons in the deep layers that project to all other layers in the column\textsuperscript{73}. Whether these mechanisms are recruited during specific cognitive operations or whether they serve general gain control mechanisms that are involved in a wide range of cognitive operations remains to be investigated.

**Future Research**

In this thesis we investigated how information is processed along the visual cortical hierarchy during visual perception. Our studies and results from other experiments has taught us a lot about the dynamic interactions between areas at different levels of the cortical hierarchy that are essential for the implementation of cognitive functions, but at the same time there is still a lot to be learned.

The anatomical connections within the visual cortical hierarchy for instance are very complex. First, the visual cortical hierarchy violates a strict serial hierarchy at even the earliest stages of visual processing. V1 for instance projects directly not just to V2, but also to V3\textsuperscript{74}, V4\textsuperscript{75,76} and MT\textsuperscript{77-79}. Second, some of the feedback connections within the visual cortical hierarchy are unidirectional, and third, other feedback connections bypass intermediate areas that allows for direct communication between high and low stages of the cortical hierarchy\textsuperscript{80}. Fourth, reentrant interactions between cortical areas within the visual cortical hierarchy are, besides direct cortico-cortical
Discussion

connections, mediated by pathways passing through the pulvinar nucleus in the thalamus\textsuperscript{81}. How exactly the different areas are connected to each other within the visual cortical hierarchy is yet to be fully mapped. Furthermore, how different connection patterns relate to different stages of information processing during visual information processing is not at all fully understood, and therefore continues to be an active area of research.

Also, although the view that progressively more complex representations are formed by the combination of inputs from earlier stages has been generally accepted ever since Hubel and Wiesel’s pioneering experiments, we surprisingly still have very limited knowledge of the visual features that neurons encode beyond V1. The fundamentals of shape-selectivity in V4 for instance, despite some early ideas and observations, remain largely unknown. Similarly, although we for instance know that cells in face areas are selectively activated by faces\textsuperscript{82} and facial parts\textsuperscript{83,84}, we do not understand how an entire population of cells represents a face in any of these areas. The experimental challenge in probing these higher levels of the visual cortical hierarchy therefore is to create a set of stimuli that will allow us to determine the selectivity of neurons as animals perform object recognition tasks, thereby fully elucidating how visual representations are combined and transformed at successive stages to map in detail the specialized function of different areas along the cortical hierarchy.

Another experimental challenge will be to disentangle feedforward and feedback influences during visual processing. Neurons in many visual areas are coactive during the complex interactions between cortical areas during the perception of a visual stimulus, and it has therefore been proven very difficult to separate the influences of lower areas onto higher areas from the effects that go in the opposite direction. Although recording neural activity
simultaneously in the different layers of the cortex in the early stages of the cortical hierarchy might provide a way to distinguish feedforward from feedback influences because these processing streams are still segregated at the level of the different cortical layers\textsuperscript{85-87}, projections beyond V1 to V2 do not strictly follow those rules because it has been observed that there are also feedforward connections that terminate across a broad range of lamina\textsuperscript{88,89}. 

New optical and genetic techniques that enable the mapping and selective (in)activation of specific brain areas, connections and cell-types will bring great opportunities for disentangling feedforward and feedback influences and determine causation. They furthermore will be very important for determining the source of top-down selection during different cognitive operations and for elucidating the mechanisms through which top-down effects modulate neuronal activity and interactions between neurons to select and enhance behaviorally relevant representations in the brain.
References


42. Poort, J. et al. The role of attention in figure-ground segregation in areas V1 and V4 of the visual cortex. Neuron 75, 143-156 (2012).


Discussion


Discussion


Summary
Summary

Understanding how our brain generates our perception of the visual world is one of the major challenges that remain to be addressed in neuroscience.

By combining visual representations in a feedforward direction at a series of processing stages along several cortical regions in a hierarchically organized system, our visual system allows visual information to transform from simple line-like features in the first visual area (V1) to more complex features like faces in higher visual areas.

In addition to this feedforward sweep, there is also a later recurrent phase of information processing. During this later phase of information processing the hierarchical organization of the visual system enables different brain areas to communicate with each other dependent on specific behavioral demands. Although this later recurrent phase of processing is thought to be essential for the implementation of various cognitive functions that depend on the flexible selection of information, how information is processed along the visual cortical hierarchy and the dynamics between cortical areas during the implementation of these cognitive functions remain largely unknown.

In this thesis, we addressed fundamental questions related to the organizational and computational principles in the hierarchically organized visual system by recording activity from neurons at different stages of the macaque visual system while the monkey performed complex visual tasks.

We found that whether a stimulus will be consciously perceived or not depends on the efficiency of feedforward propagation from lower to higher brain regions, and that to select information and use it for cognitive behavior the initial feedforward sweep needs to reach areas that are at the top of the visual hierarchy. We showed that a combination of behavioral and neurophysiological measures of pre-stimulus cortical state predicts
Summary

perceptual outcome with accuracies of 60-65%, and that combinations of these markers have separable influences on the animal's response bias and sensitivity. Our experiments also showed that when a stimulus elicits sufficient activity in higher areas for the stimulus to be perceived, this conscious perception of a visual stimulus is associated with a later processing phase with enhanced sustained activity at all the stages of the cortical hierarchy. This sustained activity presumably reflects recurrent interactions between widespread brain regions that make the visual information globally available, and thereby enable the visual stimulus to be consciously perceived.

We showed that both AMPA and NMDA receptors contribute to persistent firing that is observed in higher cortical areas during working memory related processes, and that NMDA receptors do not have a specific and critical role in persistent firing but contribute to spiking activity in a general multiplicative way. These results suggest that persistent activity during working memory related processes is not maintained solely by intrinsic dynamics of single neurons, but that internally sustained activity most likely involves recurrent processing during which reverberatory excitation between neurons within a cortical area or reciprocal excitatory loops between (sub)cortical areas takes place.

We showed that figure-ground modulation cannot arise from purely local horizontal interactions between neurons within V1, and therefore most likely is due to feedback from higher visual areas. The laminar profile of figure-ground segregation also suggests that feedback connections from higher visual areas play an important role in figure-ground segregation, and that this segregation relies on a number of different processes that unfold at characteristic time scales. We found that both figure enhancement and ground suppression contribute to figure-ground modulation; the representation of
figure elements was enhanced first in V4 and after a brief delay also in V1. After an additional delay the representation of background elements was suppressed.

Together, our experiments strongly suggest that the task specific behavioral modulation of activity along the different areas of the cortical hierarchy during the later phase of information processing reflects recurrent interactions between widespread brain regions. This way, a given visual percept arises from the global set of cortical states and task-specific interactions between multiple areas along the visual cortical hierarchy, thereby for instance enabling a visual stimulus to be consciously perceived, maintained in working memory or assigned to figure or background regions.
Acknowledgement
Acknowledgement

During the years I worked on this thesis I had a lot of support and help from people, inside the lab but also outside of science. In this section, I hope to acknowledge everyone who has been important to me during this period.

In de eerste plaats wil ik Pieter enorm bedanken voor de begeleiding die ik heb gekregen bij het doen van mijn onderzoek. De ruimte die je me liet om dingen op te lossen wanneer de experimenten niet optimaal verliepen, samen met de wetenschap dat ik altijd bij je terecht kon als ik ergens tegen aanliep en je dingen die niet helemaal kloppen gelijk opmerkt, werkte voor mij heel geruststellend. Ik heb altijd heel veel bewondering gehad voor je discipline en expertise, en ook voor het begrip en de manier waarop je door de jaren heen met mijn beslissingen en twijfels bent omgegaan wil ik je graag nog bedanken.

Kor, zonder jouw adviezen en perfecte dura scrapes had ik mijn experimenten niet kunnen doen. Bedankt voor je expertise en je geduld Kor (met de aapjes en met mij), ik vond het altijd erg fijn en gezellig om met je te werken. Ook Anneke wil ik heel erg bedanken voor de fijne en goede ondersteuning met de aapjes.

Jan, dankzij je geschreeuw voelde ik me gelijk op m’n gemak tijdens mijn eerste stage op het NIN. Onze gesprekken en je wijze woorden over de CPN, PVDA, je filosofie cursussen en vooral AJAX! zal ik niet vergeten!

Ruud, jouw geduld met mij en de microdrive waar ik talloze keren iets aan wilde veranderen is ongeevenaard. Heel erg bedankt voor je geduld en al
het werk dat je hebt gedaan, je vakmanswerk is essentieel geweest voor het uitvoeren van een van mijn projecten.

Chris v/d T, heel erg bedankt voor je continue bereidheid om me te helpen. Ik kan me nog goed herinneren dat je me tijdens m’n stage na een MATLAB cursus vroeg wat ik tijdens die cursus nu eigenlijk had geleerd omdat ik nog steeds niet kon programmeren. Ik ben je eerlijkheid erg gaan waarderen, en aan onze samenwerking tijdens het electrochemistry project, gesprekken tijdens het oplossen van MATLAB- en setup problemen en de clashes tijdens onze voetbalwedstrijden bewaar ik fijne herinneringen.

Matt, I really admire the way you combine your love for science with all the other many things you like and are good at. Thanks for your willingness to help everyone in the lab and your ability to make the NIN a better and funnier place!

Timo, je was m’n steun en toeverlaat in NY. Ik denk nog vaak met weemoed terug aan onze gesprekken en lange avonden samen. Ondanks dat ik het zelf niet meer kon opbrengen heb ik veel van je doorzettingsvermogen geleerd, ik vond het super om je passie en bevlogenheid van dichtbij meegemaakt te kunnen hebben. Samen met o.a. Louise, Meg, Danique en Winrich heb je er voor gezorgd dat ik heel veel leuke herinneringen heb aan die tijd, daarvoor heel erg bedankt allemaal!

Bruno and Chris Klink, together with Timo you made our office the best place to start my PhD. Your knowledge and experience in science helped me countless times during my PhD, and also on a wide range of topics
outside science I learned a lot from you guys. Thanks for all our joyful conversations and time together!

Danique, we zijn zo goed als tegelijk aan onze PhD begonnen. Zonder jou was mijn periode bij het NIN niet alleen een stuk minder leuk geweest, maar zou ik waarschijnlijk ook talloze deadlines en andere organisatorische verplichtingen hebben gemist. Bedankt dat je (zelfs nog na onze NIN periode) een beetje op me hebt gepast!

Roxana and Jasper, thank you for your guidance during my internships and subsequent time together.

Anne, Dev, Pia, Sreedeep, Ulf, Areg, Jessy, Enny, Catherine, Jeannette, Rob, Xing, Aurel, Jonathan and other colleagues at the NIN, I really enjoyed spending time with you, thanks for making the NIN such a pleasant place.

Tini, Wilma, Heidi en Ernita heel erg bedankt voor al jullie hulp bij organisatorische en praktische zaken.

Pap, Mam & Jac. Jullie steun, ongeacht de consequenties van mijn nogal radicale beslissingen, maar ook jullie geduld en adviezen om me te helpen uitzoeken waar ik naar op zoek ben heeft me enorm geholpen en goed gedaan.
Acknowledgement

Ook de rest van de familie hartstikke bedankt voor jullie interesse en steun. En oma, ondanks dat je zelf altijd een slag om de arm hield en me al vertelde dat je het misschien niet zou redden vind ik het heel jammer en verdrietig dat je niet bij mijn verdediging kan zijn.

Veerle, ik ben super blij dat ik zo lang over m'n proefschrift heb gedaan zodat ik na New York weer bij het NIN aan de slag kon en we daardoor nu zo gelukkig zijn samen. Hou van je.

Als laatste, en ondanks dat jullie stuk voor stuk een slechte invloed hebben gehad op het werk voor dit proefschrift; Timma, Rikka, Jeffa, Jossa, Micha, Hilla, Dani, Fipa, Eefje, Niek, Uri, Marloes, Diamond, Yonina, Mickey, Dipo, Chris, Rodney, Vedran, Anna, Isabel en ongetwijfeld ook andere vrienden die ik nu helaas vergeet, bedankt voor jullie vriendschap, jullie zijn altijd een welkome en fijne afleiding voor me geweest!