

Computational Mechanism of Filling-In in the Visual System

By

RAJANI RAMAN

(PHYS05201004007)

SAHA INSTITUTE OF NUCLEAR PHYSICS

*A thesis submitted to the
Board of Studies in Physical Sciences
In partial fulfillment of requirements*

For the Degree of

DOCTOR OF PHILOSOPHY

of

HOMI BHABHA NATIONAL INSTITUTE



June 2017

"There are only two mistakes one can make along the road to truth; not going all the way, and not starting."

Buddha

Homi Bhabha National Institute

Recommendations of the Viva Voce Committee

As members of the Viva Voce Committee, we certify that we have read the dissertation prepared by **Rajani Raman** entitled "*Computational Mechanism of Filling-In in the Visual System*" and recommend that it may be accepted as fulfilling the thesis requirement for the award of Degree of Doctor of Philosophy.



Date : 2/6/2017

Chairman : Prof. Satyajit Saha



Date : 2/6/2017.

Guide & Convener : Prof. Sandip Sarkar



Date : 2/6/2017

Examiner : Prof. S P Arun



Date : 2/6/2017

Member : Prof. Supratik Mukhopadhyay



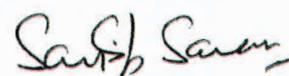
Date : 2/6/2017

Member : Prof. Dhananjay Bhattacharyya

Final approval and acceptance of this thesis is contingent upon the candidate's submission of the final copies of the thesis to HBNI.

I/we hereby certify that I have read this thesis prepared under my/our direction and recommend that it may be accepted as fulfilling the thesis requirement.

Date : 2/6/2017
Place : Kolkata


Guide

STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at Homi Bhabha National Institute (HBNI) and is deposited in the Library to be made available to borrowers under rules of the HBNI.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgment of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the Competent Authority of HBNI when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

RAJANI RAMAN

DECLARATION

I, hereby declare that the investigation presented in the thesis has been carried out by me. The work is original and has not been submitted earlier as a whole or in part for a degree / diploma at this or any other Institution / University.

RAJANI RAMAN

Acknowledgements

First of all, I would like to thank my Ph.D. advisor, Prof. Sandip Sarkar, for supporting and giving me the freedom to pursue my research, for insightful discussions and constant motivation. His guidance has helped me throughout my Ph.D. tenure and writing of this thesis. I thank Apoorva Bhandari for his quick advice which turned out to be the basis of this thesis. I would like to thank Prof. Rajesh P. N. Rao for his useful suggestions and the discussions regarding this work.

The best outcome of these years of my Ph.D. has been to come across wonderful, humble, and loving friends. They influenced me a lot in a whole range of areas- Music, Movie, Art, Philosophy, Life perspective, History, Literature, Languages, Technology, Travel, Photography (you name it !) and continue to do so. I particularly thank Aminul, with whom my frame of mind matches the most, for consistently re-assuring me that the journey is more important than the destination. I am thankful to Hitesh, from whom I learned how to be a modest, impartial human being, and use one's computer smartly. I am so fortunate to have loving friends like- Atanu, Spa, Niladri da, Sudip. You guys made my life complete. I also thank, KD, Amit, Prithwish, Jayati, Dev for being such great friends. In addition, I am indebted to Dipankar almost half of my post-MSc's fellowship for letting me copy his assignments! I owe a lot to these friends for loving me, educating me, and supporting me. The moment of our togetherness, outings, and the tours will always breathe in my memory.

I would like to thank my seniors Ajanta di, Debismita di, and Purba di for their constant encouragement. In fact, I am very fortunate to being around with awesome members of Room 337 (a power-house for all kinds of discussion) - Abhik, Sourav, Bankim, Arpita, Prashant, and Sridhar. I wish them a great life ahead.

I am indebted to Shiv Shakti (Master sahib, who always has a positive attitude and the ability to smile despite whatever might be the situation) for making me believe that I can be a researcher and give something back to the society. I am grateful to my friends from my early days- Guddo Bhaiya, Prashant, Om Prakash and Firoz, for inspiring me and ushering love and support. I thank all my friends whom I came across at different phases of my life (too many to list here but you know who you are!) and infused a positive vibe by their mere presence.

I would like to take this opportunity to especially thank my parents, my sister, and my brother-in-law. My hard-working parents have sacrificed their lives for my cherished aspiration. My sister has been the backbone to me through all these years; she has taken all my responsibilities on her so that I could focus on my endeavor. I owe her a lot for her unconditional love, as I would not have made this far without her. Last but not least, I thank Mommy (Suman), Lichi (Richa), Gudii (Gudia), and Kisi (Kinshu) for their immense love.

Rajani Raman
Kolkata, India

Dedicated to my sister, Sangeeta

List of Publications arising from the thesis

Journal

1. **“Predictive Coding: A Possible Explanation of Filling-In at the Blind Spot”**
Raman et. al,
[PLoS ONE 11\(3\): e0151194](#), (2016).
2. **“Significance of Natural Scene Statistics in Understanding the Anisotropies of Perceptual Filling-in at the Blind Spot”**
Raman et. al,
[arXiv:1607.02865 \[q-bio.NC\]](#), (2016). (Accepted in **Scientific Reports**)

Conference

1. **“A Possible Explanation of Oriented Bar Filling-in at the Blind-Spot in the light of Hierarchical Prediction Mechanism”**
Raman et. al,
[J. Phys. Conf. Series 759, 012027](#) (2016).

Rajani Raman

Contents

Synopsis	xix
List of Figures	xxviii
1 Introduction	1
1.1 Motivation	1
1.2 Organization of this thesis	3
2 Background	5
2.1 A Biological Primer	5
2.1.1 Filling-in at the Blind Spot	5
2.1.1.1 The Blind Spot	5
2.1.1.2 Filling-in	7
2.1.2 Early Visual System	9
2.1.2.1 The Visual Pathway	9
2.1.2.2 Cortical Layers	10
2.1.2.3 Retinotopy	10
2.1.2.4 Receptive fields	11
2.2 Theoretical and Computational Primer	13
2.2.1 Theoretical Problem of vision	13
2.2.2 Bottom-up Approach of Vision	13
2.2.3 Bayesian inference: Modern approach	15
2.2.3.1 Coding Scheme	17
2.2.3.2 Natural image statistics and sparse coding	18
2.2.4 Predictive Coding	20
2.2.4.1 General Architecture of HPC	21
2.2.4.2 Network dynamics and learning rule of HPC	22
3 Mechanism of Filling-In at the Blind Spot	25
3.1 Introduction	25
3.2 Simulation	28
3.2.1 Network	28
3.2.2 Training	29
3.2.3 Blind spot implementation	30

3.3	Results	31
3.3.1	Filling-in of shifting bar	32
3.4	Discussion	36
4	Properties of the Filling-In	41
4.1	Introduction	41
4.2	Simulation	43
4.3	Results	44
4.3.1	Tolerance of filling-in	45
4.3.2	Anisotropy	46
4.3.2.1	Anisotropy in orientation selectively	48
4.3.2.2	Anisotropy in filling-in	49
4.3.2.3	Anisotropy in tolerance of filling-in	50
4.3.2.4	Comparison with the psychophysical results	54
4.3.2.5	Relation between natural image statistics and filling-in at the blind spot	58
4.4	Discussions	63
5	General Discussion	67
5.1	Summary	67
5.2	Discussion	69
5.3	Conclusion	74
A	Coding Length of HPC	77
B	Architecture of the PE Module	79
C	Development of Receptive Field in BS area	81
D	Filling-in of the natural scene	85
	Bibliography	89

Synopsis

Vision is an extraordinary phenomenon. Dealing with under-sampled retinal input is one of the remarkable capabilities of the visual system among its other miraculous aptitudes. Filling-in at the blind spot is one of the examples of how brain interpolates the informational void due to the deficit of visual input from the retina. Because of the absence of photoreceptors at optic disc, the retina is unable to send the corresponding signal to the brain and thereby, hides some portion of the visual field. The concealed visual field is known as the blind spot. However, we never notice any odd patch in our visual field, even in monocular vision, but rather we see the complete scene; filled up in accordance with the surrounding visual attributes [1,2]. This completion is known as perceptual filling-in or simply filling-in. In addition to the blind spot, filling-in also occurs in other visual input deficit conditions (e.g. filling-in at the artificial and natural retinal scotoma) including visual illusions (e.g. Neon colour spreading, Craik-O'Brien-Cornsweet illusion, Kanizsa shapes).

For years, filling-in completion has demanded the attention of researchers from many areas: Neurophysiology, Psychology, Computational neuroscience, and philosophy. This resulted in various hypotheses which range from passive to active completion. Recently, many psychophysical, physiological and imaging studies have revealed a

deeper insight of this phenomenon. These studies brought out at least two very important points [3]. Firstly, the filling-in is an active phenomenon; there exists a neural correlation to the filling-in at the blind spot. And secondly, the early parts of the visual system are mainly associated with this phenomenon. Recently, physiological studies have reported that the neurons, in the deep layer of primary visual cortex at the location retinotopically related to the blind spot, exhibit non-linear response during filling-in of a bar across the blind spot [4, 5]. Psychophysical studies have further elucidated that, though filling-in occurs readily for the aligned pair of bars placed on opposite sides of the blind spot, it exhibits some tolerance with the small degree of variation in attributes (e.g. misalignment, orientation difference) [6]. Furthermore, filling-in exhibits anisotropic behavior; where the degree of filling-in depends on the bar orientation (certain orientation is favorable than the other) [6, 7].

Despite this early progress in the psychophysical and physiological domain, the actual neural and computational mechanism of filling-in has remained unexplained [5]. How the visual system manage to fill-in the informational void? What kind of computational principles and neural mechanisms are involved? What is the significance of observed non-linear responses in filling-in? In this thesis, we addressed these questions in a computational framework of predictive coding in conjugation with natural images statistics [8], which is recently argued as a general computation principle [9] and has explained a range of properties of visual cortex [10].

The problem of vision has been argued as an inference or an estimation problem; where an organism tries to estimate the hidden physical cause (object attributes such as shape, texture, and luminance etc.) behind the generated image that organism receives as an input. In the hierarchical predictive coding framework (HPC), it is assumed that the image generation in the outer world involves hierarchical, multilevel, spatial and temporal interactions between the physical causes. The goal of the visual system is, thus, to estimate (or internally represent) these multilevel hidden physical causes efficiently; which is accomplished using recurrent prediction-correction

mechanism along its hierarchy using predictor-estimator module (PE module).

In this framework [8], on the arrival of an input, a PE module at each visual processing level generate the prediction (or estimate) on the basis of the learned statistical regularities of natural scenes. Each higher area (say V2) then sends the generated prediction to its immediate lower level (say V1) by feedback connections and in return receives the error signal, by feed-forward connections, which is used to correct the current estimate. An equilibrium state is achieved after the completion of a few prediction-correction cycles; where the estimate matches the input signal. This optimum-estimate is regarded as a representation of the input at that level. The achieved optimum-estimate at different levels of the network is depicted as a perception of the input image.

This computational process, and hence the architecture of the network, basically originates from probabilistic estimation theories. In Bayesian framework, the process of prediction-correction is a manifestation of maximum a posterior probability (MAP), where maximizing the posterior probability $P(\mathbf{r}, U | \mathbf{I})$ with respect to the activity vector, \mathbf{r} , and represents the estimate of the input, \mathbf{I} , and weighting matrix U provides the dynamics and learning rules respectively of a predictive-estimator module. In other words, it is a process of minimizing the error between generated image and the input image \mathbf{I} . The generated image is a linear combination of learned weighting vectors (i^{th} column of U). The coefficient (i^{th} element of \mathbf{r}) is the activity of i^{th} PE neuron. The reconstructed image, corresponding to optimal-estimate is designated as “perceptual images” in this study.

In this work, I simulated a three-level linear hierarchical predictive network. The middle-level network, which is equivalent to V1, consists of nine PE modules. These modules receive input from front level, which is equivalent to the output of LGN, and send the output to the solitary module at the last level. The last level module is equivalent to V2. Therefore, the PE module at V2 receives input from all nine V1

PE modules and sends back the feedback signal to all of them. This architecture is based on the fact that the visual area higher in hierarchy operates on a higher spatial scale.

Natural images from the different environment were used for training. Variance normalized batches of image patches extracted from randomly selected locations from the randomly selected images were given as input to the network. For each batch of image patches, the network was allowed to achieve optimal estimate (of \mathbf{r}), under the constraint of sparse coding [11], and the average of these estimates was used to update the efficacy matrix U of neurons, initially assigned random values. The neurons V1 was trained first and subsequently the neurons in V2.

To mimic the blind spot, the feed-forward connection in a certain area was removed from the model network, which was ‘pre-trained’ with usual feed-forward connections. The ‘pre-training’, the training before the creation of the blind spot, captures the fact that the active neurons in deep layer (5/6) corresponding to filling-in has been reported to be of the binocular type. These neurons were found to respond to the inputs from both eye and hence, possess binocular receptive field [3, 4]. We designate the network with the blind spot as BS network and the one without the blind spot as a non-BS network.

The learned synaptic weights of model neurons at V1 resemble the Gabor-like structure similar to the receptive field of simple cells at V1, distributed in the different orientation and spatial frequency, reported earlier in several studies [8, 12, 13]. The weighting profile of neurons at the V2 resembles relatively more complex visual features: long bar, curve, etc.

Both BS and non-BS network were exposed to horizontal bar stimuli of different length. One end of the bar was fixed at a position outside of the blind spot, whereas, the position of other end was varied (by one pixel at each instant) across the blind spot. The response of PE neurons in the central module in the model network

(designated as BS module) at V1 and the sole module at V2 was recorded for the different end position of the bar. These responses were used to generate “perceptual images” to explicitly bring out the correspondence between these responses and the perceptual filling-in completion.

We found that the response of neurons in the BS module remained constant and relatively low as long as the bar end remained inside the blind spot and this resulted in the perception of a bar of a constant length outside the blind spot. On the other hand, when the bar crosses the blind spot, the responses are elevated non-linearly, and the filling-in completion occurred. These results are consistent with the findings of physiological studies [4].

Moreover, the non-linearity involved in filling-in was investigated explicitly. We found that the response to aligned pair of bars, which stimulate the both sides of the blind spot simultaneously, was larger than the sum of responses to the bar presented on either side of blind spot individually at a time. This indicates that the abrupt change in the magnitude of the response during filling-in completion cannot be explained by the stimulation of the receptive field extending out from the opposite side of the blind spot.

To investigate the tolerance of filling-in, we performed investigations with two different attributes: misalignment and orientation difference. Two bar segments were presented on the opposite sides of the blind spot. One of these bar segments was fixed while the other was either shifted (horizontally misaligned bars) or rotated depending on the chosen attribute.

We recorded the response of PE neurons, in the BS module, at V1 and generated the corresponding “perceptual images”. The bar appears completed when both segments are perfectly aligned and the filling-in remains largely unaffected for some degree of misalignments and orientations difference. The filling-in fades away quickly with a further increase in misalignment and orientation difference. This result indicates that

the filling-in completion is highly favorable for perfect alignment and is tolerant to a certain extent to the increasing deviation. Similar results were also reported in earlier psychophysical studies [6, 14].

To study the anisotropy in filling-in, we first investigated whether our model network could possess anisotropy via training with natural images. To do this, we obtained the orientation tuning distribution of the neurons in V1. We used bar stimulus of different orientation and frequency to determine the orientation tuning of particular neurons by registering their optimal response. We found that the distribution is higher for the horizontal orientation, followed by vertical and then non-cardinal. This result is very much in-line with the results related to the anisotropy of orientation distribution in natural images [15, 16] and orientation tuning distribution of neurons in primary visual cortex [17–19].

To investigate the anisotropy in filling-in, the trained network was exposed to stimuli consisting of a pair of bar segments in horizontal and vertical orientations (across the blind spot). One side of both bar segments was fixed inside the blind spot, whereas other ends were expanding in opposite direction. The responses of PE neurons were recorded. The filling-in-value – the average pixel value inside the blind spot in the reconstructed “perceptual image”– starts increasing beyond a certain critical minimum length of bar segments. This result exhibits the “minimum-length requirement” properties of filling-in. Moreover, the comparative observation of results for Horizontal and vertical arrangements indicates that the critical length needed for filling-in is lesser for the horizontal bars. In addition, for the same length, the filling-in performance is better for the horizontal case. This anisotropic property is in complete agreement with studies [7]. To investigate the anisotropy in tolerance of filling-in, we repeated the experiments (as described in the first part), with misaligned and oriented bar pair in the horizontal and vertical configuration. We found that though the filling-in value is larger for the aligned bar cases in the horizontal configuration, it drop down more quickly in comparison to the vertical configuration with increasing

misalignment and orientation difference. These results indicate that the visual system favors the horizontal configuration in order to fill-in but exhibits less tolerance to the deviation. On the other hand, the vertical configuration is inferior for filling-in but it is relatively more tolerant to the deviation [6].

These results can be understood under the computational principles of predictive coding. For an input stimulus around the blind spot, higher areas (V2) generates unified estimate (including the estimate corresponding to blind spot region) of the input stimuli on the basis of the learned statistical regularities of natural images. This estimate remains uncorrected due to the absence of error carrying feed-forward connection in BS region at V1 and therefore, local optimum- estimate is achieved essentially by top-down prediction. Influenced by learned statistical regularities, higher areas predict a long continuous bar across the blind spot and this result in the perception of completion. The observed properties of filling-in result from the degree of similarity between statistics of stimuli around the blind spot and the natural image statistics.

More frequent features in the natural scenes tend to be more likely for completion across the blind spot. In the perspective of anisotropy, the over-representation of horizontally oriented feature (lines, bar etc.) in natural scene cause the over-representation of it in the learned receptive field in the primary visual cortex. This leads to the superiority of horizontal bar for the filling-in. But what about its inferiority when it comes to tolerance of filling-in? The higher population tuning toward horizontal orientation could result into more specificity about the estimate they generate and therefore less tolerant for deviations. This could be analogous to the orientation tuning profile of a single neuron in the visual cortex, where more tuned neurons are less tolerant to the deviations.

In conclusion, in this work we have attempted to present a fresh perspective of the computational mechanism of the filling-in process at the blind-spot in the framework

of hierarchical predictive coding, which provides a functional explanation for a range of neural responses in the cortex. We demonstrate that filling-in is a manifestation of a hierarchical predictive coding principle and, the nature of filling-in is predominantly guided by the learned statistical regularities of the natural scene. We also show that the anisotropies of filling-in at the blind-spot and the associated tolerances are the outcome of the intrinsic anisotropies of natural scenes. Our studies suggest that natural scene statistics plays a significant role in determining the filling-in performance at the blind-spot and shaping the associated anisotropies. Moreover, these studies also advocate that Hierarchical Predictive Coding in conjugation with natural scene statistics can provide a framework for encoding computational mechanisms of perceptual filling-in phenomenon (at the blind-spot) that can serve as a link between natural scene statistics, cortical organization and the perceptual experience.

This thesis is divided into two major parts. The first part includes studies related to the computational mechanism underlying the filling-in, which includes the explanation of non-linear responses and its correspondence to filling-in. Moreover, it is also shown that the tolerance of filling-in could also be understood in the framework of hierarchical predictive coding of a natural image. In the second part, it is shown that the anisotropy in filling-in and tolerance of filling-in arises from the anisotropy in natural image statistics.

This thesis is composed of five chapters-

- The first chapter contains a brief introduction, motivation and the organization of the thesis.
- The second chapter contains an overview of the functional mechanism of early visual systems and the hierarchical predictive coding (HPC) formalism in the backdrop of efficient coding hypothesis and natural scene statistics.
- In Chapter three we present our investigations, using simulation, about the computational mechanisms of filling-in. We show that filling-in could be a

manifestation of the general computational mechanism of hierarchical predictive coding, in which the properties of filling-in could be guided by learned statistics of the natural scene.

- In the fourth chapter, we investigated the origin of anisotropy in filling-in in the light of natural image statistics. We demonstrated a general link between the anisotropy in natural image statistics, visual cortex and the percept related to orientation tuning.
- Chapter five summarizes the contributions of this thesis including limitation as well as future directions.

List of Figures

2.1	The retina	6
2.2	Blind-Spot demonstration	7
2.3	Visual pathway	9
2.4	Retinotopy	11
2.5	Receptive fields	12
2.6	Learned RF with sparse prior	19
2.7	General HPC architecture	22
3.1	Schematic illustration of bar completion experiment	27
3.2	Three level HPC model network	29
3.3	Natural images	30
3.4	Model blind Spot	31
3.5	Learned synaptic weights	32
3.6	Shifting-bar investigation	33
3.7	Responses profiles	34
3.8	Response elevation in BS region at level 1	34
3.9	Response profile at level 2	35
3.10	Nonlinearity in the response profile	37
4.1	Anisotropy in orientation selectivity	44
4.2	Stimuli	46
4.3	Tolerance in misalignment	47
4.4	Tolerance in disorientation	48
4.5	Anisotropy in the natural scene	49
4.6	Filling-in anisotropy	51
4.7	Anisotropy in tolerance of filing-in of misaligned bars	52
4.8	Anisotropy in tolerance of filing-in of disoriented bars.	53
4.9	Comparison with psychophysical results.	57
4.10	Validation investigation.	62
B.1	PE module architecture	80
C.1	Receptive field development	83
D.1	Filling-in of natural image	86

*Two roads diverged in a wood, and I,
I took the one less traveled by ...*

Robert Frost in "The Road Not
Taken"

1

Introduction

1.1 Motivation

Vision is an extraordinary phenomenon. Coming through lousy lens of eye the light falls on irregular mosaic of back to front photoreceptors by passing through layer of blood vessels over the retina. The information is then get compressed in order to transfer from 126 million photoreceptors to one million optics nerve. Dealing with these under-sampled retinal input is one of the remarkable capability of the visual system among its other miraculous aptitudes. Visual cortex receives the train of spikes corresponding to highly under-sampled fuzzy retinal images. However, we

manage to acquire complete and detailed perception of the world. Filling-in at the blind spot is an example of similar phenomena, where the visual system interpolates the informational void originated due to the absence of photoreceptor at the blind spot. Despite the absence of information corresponding to the blind spot, we never perceive any odd patch in our visual field. Main reason behind this is, normally, the retinal input from the fellow eye compensate for the informational void originated due to the blind spot. However, even in monocular vision we do not notice any void but rather we see the complete scene. Our visual system fills that absence with the visual attributes in the surrounding. How the visual system manages to do such interpolation? What computational principle and neural bases are involved?

For years, filling-in completion has demanded the attention of people from many areas: neurophysiologists, psychologists, computational neuroscientists, and philosophers. These attention has led toward various hypothesis which ranges from passive to active completion. Recently, many psychophysical, physiological and neuroimaging study has been carried out. These studies reveal at least two very important points. Firstly, the filling-in is an active phenomenon; there exists a neural correlation to the filling-in at the blind spot. And secondly, the early part of the visual system are mainly associated with this phenomenon. The second one, obtained through psychophysical studies (and validated by physiological studies), reflect from the fact that the simple structures like bars, lines etc completes readily across the blind spot whereas, there are limitations for the completion of relatively complex structures. Recently, neurophysiological studies have reported that the neurons, in the deep layer of primary visual cortex at the location retinotopically related to the blind spot, exhibit non-linear response corresponding to filling-in of a bar across the blind spot. In the psychophysical domain, the recent studies had reported other properties of filling-in like, tolerance, anisotropy, and anisotropy of tolerance. Despite this progress, the neural and computational mechanism of filling-in has remained unknown. In this thesis, we address these issues in a general computational model of predictive coding

and the natural image statistics.

A line of theoretical research has grown, recently, under the common umbrella of Bayesian inference framework of perception. In this framework, it is argued that the statistics of the natural scene (and hence the statistics of sensory input) imposes the organizing principles behind the observed neural properties of the visual system. In line with such hypothesis, hierarchical predictive coding (HPC) model assumes that the visual system learns the statistical regularity of natural scene. And on the basis of this, a group of neurons at each higher visual processing levels attempts to estimate (or predict) the inputs to the lower-level, while the neurons in the lower processing levels signal the error (to the prediction) to the higher processing level to adapt so as to minimize the discrepancy. A multi-level optimum estimate encoded by those neurons is therefore argued as a representation of cause of the sensory signal.

In the last two decade, various properties of the visual system have been explained under the computational model of HPC. HPC explains how the interaction between top-down prediction and bottom-up error signal give rise to the extra-classical receptive field (RF) at the primary visual cortex, spatiotemporal RF at LGN, MT and many other properties of visual cortex.

We hypothesized that the similar prediction-correction mechanism under the backdrop of natural image statistics could account for filling-in and the related properties.

1.2 Organization of this thesis

Chapter 2 comprises of mainly two section: 1. the biological primer and 2. The computational primer. The first section contains the overview of the physiology of visual system (mainly the early visual system) where we emphasize the origin of the problem: blind spot. A brief overview of the existing functional mechanism of the early visual system is also included. The second section contains the general

theoretical framework which forms the foundation for this thesis. Firstly, the Bayesian inference framework of perception is introduced, which is a common basic principle for many theoretical models, and its relation to natural image statistics. Secondly, the framework of predictive coding is explained with emphasize on hierarchical predictive coding model. This part describes the HPCs general computational principle and its origin from probabilistic estimation theories.

Chapter 3 investigates, using simulation, the computational mechanism of filling-in. This chapter explains how HPC model network, after trained with natural images, accounts for the non-linear response observed in primary visual cortex during filling-in of a bar. The representation of these responses indeed shows the filling-in completion. Results of the investigations suggest that the filling-in could be a manifestation of the general computational mechanism of hierarchical predictive coding, in which the properties of filling-in could be guided by learned statistics of the natural scene.

In the 4th chapter, investigations on the origin of tolerance and anisotropy in filling-in in the light of natural image statistics is presented. The visual system favors the filling-in of a horizontal bar over the vertical bar. Whereas, the vertical bar gets the upper hand in the case of tolerance of filling-in. In this context, we analyze the orientation-tuning density of learned receptive field in our model network and investigate its possible resultant effect as the nature of anisotropy in filling-in. Findings of this chapter, also demonstrate a general link between the anisotropy in natural image statistics, visual cortex and the perception related to orientation tuning.

In chapter 5, a summary of this thesis work is presented along with discussion and the possible limitations and scope for future investigations.

There is pleasure in recognizing old things from a new viewpoint.

Richard Feynman

2

Background

2.1 A Biological Primer

2.1.1 Filling-in at the Blind Spot

2.1.1.1 The Blind Spot

Retina is the ‘frontal desk clerk’ of the organization called visual system. Its job is to receive the foreign delegate, the 2-D intensity pattern: an image, extract the

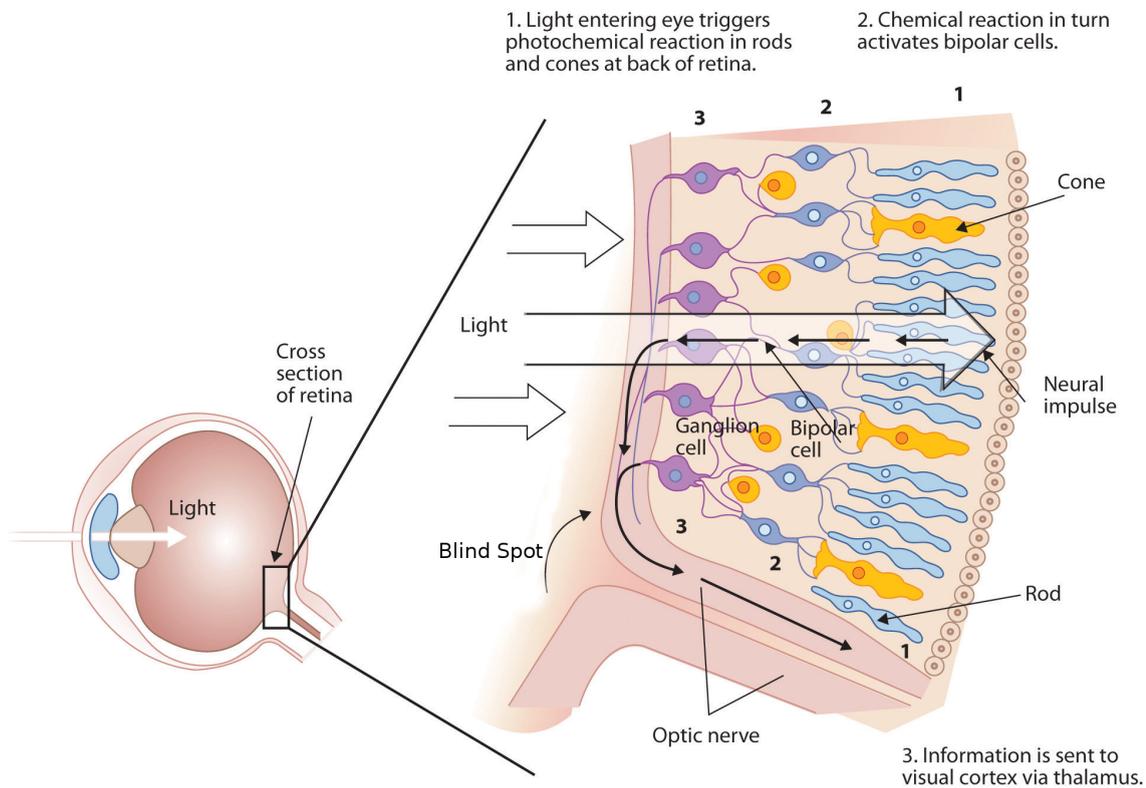


FIGURE 2.1: *The retina and the blind spot. (Figure modified from 'From Flat World Knowledge, Introduction to Psychology, v1.0', CC-BY-NC-SA)*

information, process a bit (well, in fact, a lot!) and then convey it to the visual system in the language visual system understands: the train of spikes.

Fig 2.1 shows the structure of retina which resides in the rear part of the eye. The photoreceptors code the image in the form of chemical signal and send it to the subsequent level cells. The information eventually leaves the retina in the form of the spike (action potential) as the output of ganglion cell via the optic nerve. This picture of retinal processing is highly simplified. In fact, retina performs much-sophisticated processing before sending it to the brain, which is out of the scope of this thesis. For a wonderful review of retinal processing see [20, 21].

Fig. 2.1 shows that the ganglion cell situated at the top layer while the photo-receptors lies on the rear side of the retina (and hence, does not receive get the direct light).

FIGURE 2.2: *Blind-Spot demonstration*

The axons of ganglion cells pass, in a bundle called optic nerve, from the surface of the retina to the brain. Such arrangement leaves behind a photo-receptor free zone in the retina known as the blind spot.

2.1.1.2 Filling-in

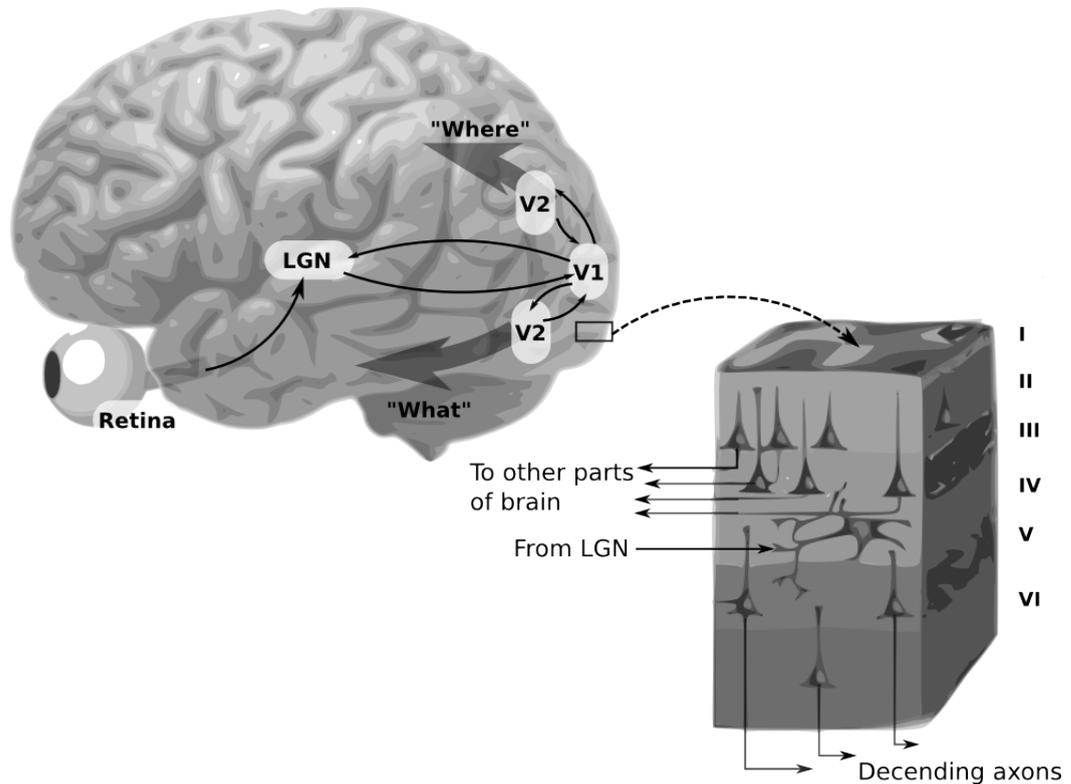
Our visual system does not get any information corresponding to the blind spot, nevertheless, we never notice any odd patch in our visual field. In fact, we perceive the same visual attributes there at the blind spot, as present in the surrounding of the blind spot [1]. This phenomenon is called the filling-in. One of the reasons that we do not perceive the odd patch is that the input from the other eye compensates for the loss. But even in monocular vision, the filling-in occurs.

We can demonstrate this phenomenon by closing your right eye, fixate on the upper cross sign and hold the Fig. 2.2 about 1 feet from your face. As we can observe, the dark circle is replaced by the surrounding white plane background surface. What happens here? Above arrangements projected the dark circle exactly inside the blind spot of retina — causing the disappearance of the circle— and visual system somehow managed to fill that informational void with the information in the surrounding. Similarly, If you fixate on the lower cross the broken bar will appear completed. How brain perform filling-in at the blind spot? This quest forms the basis of this thesis.

In general, filling-in take place in a variety of conditions. The conditions have been broadly categorized into three type. First is, the filling-in corresponding to the deficit of visual input. In this situation, filling-in take place when a particular area of the visual field is deprived of the visual input [1, 22–24]. The filling-in at the blind spot falls into that category. Filling-in at the scotoma– a region in the visual field that get concealed due to confined damage to some part of the visual system –is another example of such situation. The filling-in, in these conditions, occurs rapidly. Another situation arises in normal vision (without deficit of visual input) in the case of steady fixation and the stabilized retinal image. In these situations, a prolonged fixation on, or a stabilized retinal image of, an object causes the object to become invisible. The region corresponding to the object is eventually gets filled up with the visual attributes in the surrounding. The example of such phenomenon are: Toxler effect [25, 26], filling-in at an artificial scotoma [22, 27, 28], etc. The third kind of situation arises in well-known illusions: Neon color spreading, Craik-O’Brien-Cornsweet illusion [29, 30], Kanizsa shapes [31], etc. This is again the case of the normal vision and prolonged fixation.

Although the discovery of filling-in process goes back in nineteenth century, it is recent that the coming together of the ideas from the different research stream – psychophysics, neurophysiology, neuroimaging and, computational neuroscience- revitalized the potential interest in this phenomenon. Though the studies from the various approaches have provided several important development, the understanding of filling-in is far from complete. What is the mechanism (neural and computational) of filling-in of different filling-in phenomenon? How these mechanisms are related. These are the general question remained to be answered. In this thesis, however, we mainly concentrate on the filling-in mechanism at the blind spot and we would discuss the possible mechanism sharing to the other filling-in phenomena.

We will come back to this in the next chapter. Before that, we should acquire some overview of the early visual pathway and it functional accounts.

FIGURE 2.3: *Visual pathway*

2.1.2 Early Visual System

2.1.2.1 The Visual Pathway

Hierarchy exists throughout the visual processing (see Fig 2.3). The output of retina reaches to the cortex via LGN, a sub-cortical area act as a relay station. The first stage of the cortex is comprised of V1 and V2 areas, known as early cortical areas. Then, two separate processing pathway split from the early visual area, ventral and dorsal pathway [32]. The ventral path, comprised of V4 and IT, process the form and color information, and also known as ‘What’ pathway. The dorsal path, comprised of MT, process the motion and spatial information, and also known as ‘Where’ pathway. (A good textbook account of these and other well-known facts, which we are going to discuss in this section, about the visual pathway, can be found at [33,34])

Though above descriptions mainly indicate the feed-forward connectivity, there exist

reciprocal feedback connections among the different processing levels [35]. For example, V2 sends back the feedback signal to the V1 and, V1 to LGN. The function role of this reciprocal architecture is still not well understood. This thesis relies on one of the propositions that are made to explain the role of feedback circuitry, which is introduced in the next section.

Nearly 40 to 50 percent of our brain is devoted to the vision [36]. Given this grandness and complexity, our understanding of the visual system is far from complete. Leaving out the discussion related to higher visual areas, we here mainly overview the early visual areas (LGN, V1, and V2) which provide the sufficient background needed to explore the subject in this thesis.

2.1.2.2 Cortical Layers

As is usual in all the cortex, V1 possesses 6 distinct horizontal layers (see Fig. 2.3) [37]. Layer 4 receives input from LGN by the feed-forward connection, while layer 6 (and 5) sends the feedback connection to the LGN (and other subcortical targets). Layer 2/3 sends feed-forward output to the other cortical areas while superficial layers receive the feedback connection from them.

2.1.2.3 Retinotopy

Both hemispheres contain the V1 area. The information of left visual field goes to the V1 of right hemisphere while the information of right visual field goes to the left hemisphere. Except for this discontinuity, the visual field is mapped in continuously in the early visual cortex (Higher areas do not follow such scheme though). In other words, neighboring points in the visual field is processed by the neighboring neurons in the early visual cortex. This map is called retinotopic map [38] (see Fig. 2.4.

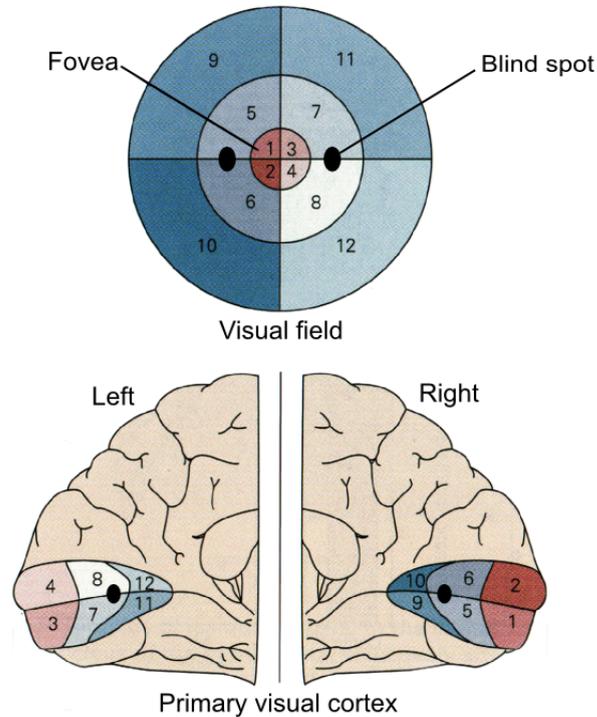


FIGURE 2.4: *Retinotopy in primary visual cortex. (Figure modified from Wikimedia Commons CC-BY-NC-SA)*

Early visual cortex possesses a distorted retinotopic map. For example, in V1 area of the human brain, 50% of visual area process only 2% of the central visual field (see Fig.) [39] and this cause the distortion corresponding to the magnification of central vision. Other than this, geometrical distortion emerges as an effect of translation of radial visual field to the non-radial V1 area, where concentric circles and radial lines translate into vertical and horizontal lines in V1. The Fig. 2.4 also shows the retinotopic representation of blind spot in the V1 area.

2.1.2.4 Receptive fields

The response of a typical neuron, in the early visual areas, depends on the intensity structure of a tiny area of the visual field. Normally this area is referred as the classical receptive field. But more generally, the receptive field of a neuron is characterized by the intensity structure that makes that neuron responds maximally.

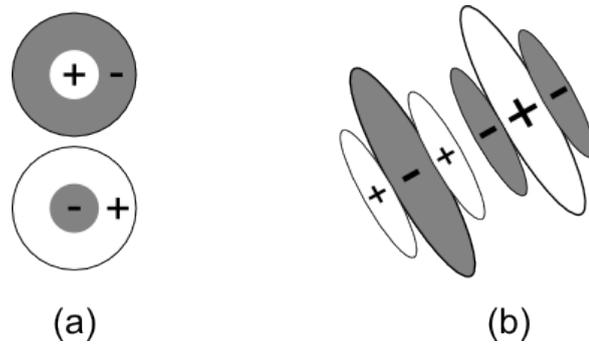


FIGURE 2.5: *Receptive fields (RF). (a) ganglion cell RF. Upper center surround RF represents the on-center RF (indicated by '+' sign), while lower one represents the off-center RF (indicated by '-'). (b) Simple cell off-center and on-center RF.*

The receptive field of neurons in the particular areas has been considered as the important factor in understanding the function of that area or, more specifically, of those neurons. For example, ganglion cell–neurons in the outer layer of the retina—as well as cells in the LGN, possess circular center-surrounded receptive field [40] (see Fig. 2.5a). The light beam in the center of receptive field excites the response of some ganglion cells whereas, stimulation in surrounding leads to the inhibition of response. These neurons are known as on-center cells. There are other kinds of cells known as off-center showing response opposite to the ON-center.

Primary visual cortex accommodates, mainly, two type of neurons; simple cells and complex cells, based on the structure of their receptive field [41, 42]. Simple cells typically possess oriented elongated receptive field with separate on and off sub-regions (see Fig. 2.5b) Simple cells, respond strongly to the bar and edge stimulus of a particular orientation and at a particular location [43]. Moreover, these cells are strongly selective for the spatial phase of the stimulus.

Complex cells, on the other hand, do not possess separate on-off regions, hence, they show response invariant to spatial phase of stimulus. These cells, however, respond to the bar and edges of a particular orientation and at a particular location [43].

2.2 Theoretical and Computational Primer

2.2.1 Theoretical Problem of vision

How does vision work? What is the organizational principle of the physiological and computational structure of the visual system? To understand this, we first have to understand what are the computational problems visual system solves in order to provide us a coherent, reliable visual perception of the outer world.

To start with, Our retina receives a 2D image as an effect, caused by an ambitious 3D world. This image is produced by a complex interaction between various parameters, for example, illumination, surface geometry, reflectance, at multiple spatial and temporal scales. This image is in turn sampled by— not so perfect and noisy —photoreceptor mosaic (in addition to the fact that the light has to cross several layers of retinal cells in order to fall on the photoreceptors in the retina. The retina then translates this under-sampled, noisy, ambiguous, 2D image into train of spikes and transmit it to the brain, through a very narrow channel (in compare to the photoreceptor population) of the optic fiber.

Now, the brain has a task, seeming almost impossible, to decode correctly interpret sensory data (in form of train of spike) to get a detailed, coherent, reliable perception of the *cause* of those sensory data: the world. Moreover, the visual system has to perform this difficult task with its limited anatomical and metabolic capital.

We know that brain manage to accomplish that very efficiently. So efficiency that we take the vision as granted phenomenon.

2.2.2 Bottom-up Approach of Vision

In the initial computational approach, pioneered by Marr [44], the problem of vision has been seen as the problem of image analysis. In this approach, the role of the early

visual system is to filter the input signal to extract the feature, like edge, boundary etc. from the image. Which is ultimately is grouped and represented in higher areas and compare to the stored templates of the objects, in order to recognize what information input signal is carrying about the world.

This framework centered around the processing in feed-forward manner, where neurons in lower area (say V1) extracts relatively simpler features and sends it output to the higher area (say V2). The extraction is represented as the neuronal response r as a result of filtering of input signal (image) $\mathbf{I}(x, y)$ with receptive field profile $\mathbf{W}(x, y)$, which is basically weighted sum of the input pixel $\mathbf{I}(x, y)$ -

$$r = \sum_{x,y} w(x, y)\mathbf{I}(x, y) \quad (2.1)$$

Where, $w(x, y)$, elements of the vector \mathbf{W} , is weighting element corresponding to the input $I(x, y)$ and it provides the mathematical means of a receptive field. This equation can be re-written as-

$$r = \mathbf{W}(x, y)\mathbf{I}(x, y) \quad (2.2)$$

Now, the response of a ganglion cell can be predicted by filtering the local input image with its receptive field that is modeled by the difference of the Gaussian (DoG) function [45] (it also modeled by Laplacian-of-Gaussian), sometimes known as Mexican hat.

Filtering image with DoG function, which is basically a spatial band-pass filter, brings boundary of intensity variation [44]. And had been hypothesized as a local edge detector. Similarly, the simple cell is modeled by the 2-d Gabor filters and its role has been hypothesized as an extraction of local contour information [46–49].

This approach, however, despite its early success, does not go very far in explaining a range of property of visual cortex [50]. Moreover, it does not provide enough light

to the problem of vision as a phenomenon in the backdrop of the complexity we discussed in the previous section.

The main reason for this inability is its exclusion of importance of prior information visual system seems to have acquired about the world. This prior information should play an important role in the visual processing implemented by top-down feedback connection in the visual system. This goes with the fact that there exists feedback connection in roughly equal number to the feed-forward connection. But the Marr's approach rely upon pure bottom-up computational approach, in which information streams from lower to higher processing levels.

Second reason is, in this approach, the visual system remains independent of the type of input (natural, or artificial) and it works like a general computational machine. But the fact is our visual system is not a general processing machine. It has evolved for performing a certain type of computations necessary for us to survive. Therefore, Marr's approach lacks adaptive computation performing ecologically valid task.

One can expect promising outcome from the alternative frameworks which incorporate all these points.

2.2.3 Bayesian inference: Modern approach

An insight of problem of vision is that it is an inverse problem: one where the *causes* of sensory data need to be *inferred* from sensory data [8, 51–57]. Last few decade has seen the strongly re-emergence of a promising frameworks where vision has seen as a ‘Bayesian inference’ [58–60]. This notion has its root that goes back to the Helmholtz [61], who proposed that the visual perception is a result of the knowledge driven inference our brain make about the hidden physical *cause* (object attributes such as shape, texture, and luminance etc) of the sensory signal it receives.

In this framework, it is hypothesized that our visual system *learns* the internal model of the outer world [12, 57, 62–66]. On the basis of this learned prior information, visual system *estimate*, optimally, the hidden cause from underlying ambiguous effect in the form received sensory signals [55, 59, 67–69].

So, the visual system has two fundamental tasks: (i) making an estimate and (ii) learning parameters involved in image generation. The precise estimate of cause is impossible to compute. Because one to one mapping of cause and effect in our complex world is difficult. One cause can give rise to different sets of effector different cause can give rise to similar effect. Therefore, in this probabilistic Bayesian framework, the fundamental tasks reduce to the finding the hidden physical cause \mathbf{r} and parameters U that maximize the posterior probability, $P(\mathbf{r}, U|\mathbf{I})$, which is the probability distribution of causes \mathbf{r} and parameters U and given a input signal \mathbf{I} .

The posterior probability $P(\mathbf{r}|\mathbf{I})$ can be obtained using the Bayes theorem (for simplicity assuming the \mathbf{r} and U are statistically independent)-

$$P(\mathbf{r}, U|\mathbf{I}) = \frac{P(\mathbf{I}|\mathbf{r}, U)P(\mathbf{r})P(U)}{P(\mathbf{I})} \quad (2.3)$$

Here, probability density $P(\mathbf{I}|\mathbf{r}, U)$ is the distribution of generative model and it sits in the heart of ‘Bayesian inference’. It learns the causal-matrix (internal model of cause and effect). On the basis of this information, it generates the hypothesis of the hidden causes (of sensory input) by generating the sensory pattern and comparing it to the input sensory signals. This approach quite opposite to the image analysis, which we came through in the previous section, and sometimes known as analysis by synthesis approach. The factor $P(\mathbf{r})$ provides the bias for causes, which contribute to the estimation of the cause by weighing up the generative model for the more frequent occurrence of causes than the others. $P(\mathbf{I})$ is the normalizing constant.

In summary, the fundamental task is, therefore, to find the optimal values -

$$(\hat{\mathbf{r}}, \hat{U}) = \max_{\mathbf{r}, U} (P(\mathbf{r}, U | \mathbf{I})) \quad (2.4)$$

Finding optimum U , for keeping \mathbf{r} constant, is a process of learning. While finding optimum \mathbf{r} on the basis of learned U is the process of estimation or *representation* of physical cause.

2.2.3.1 Coding Scheme

On arrival of an input image \mathbf{I} , a linear generative model tries to generate an image \mathbf{I}' , by linear combination of learned parameter (or basis vector) vector U_i with coefficient r_i -

$$\mathbf{I}' = \sum_i r_i U_i \quad (2.5)$$

In neural network convention, vector U_i represents the receptive field of neuron with response r_i . This can be re-written as-

$$\mathbf{I}' = U\mathbf{r} \quad (2.6)$$

Where, U is a matrix, called efficacy matrix, encompassed U_i as a i^{th} row and, \mathbf{r} is a vector, called response vector, encompassed r_i as i^{th} element.

In a deterministic generative model the process of finding the optimal value of \mathbf{r} and U , thus, would relate to values that minimize the square error between the input image and generated image.

$$E = |\mathbf{I} - U\mathbf{r}|^2 = (\mathbf{I} - U\mathbf{r})^T (\mathbf{I} - U\mathbf{r}) \quad (2.7)$$

Where, T stands for the transpose of matrix $(\mathbf{I} - U\mathbf{r})$.

But as our visual process is a stochastic one, a probabilistic generative model is more appropriate to consider. We can express this by assuming error as Gaussian noise of zero mean and variance σ -

$$P(\mathbf{I}|\mathbf{r}) = \exp\left(-\frac{(\mathbf{I} - U\mathbf{r})^T(\mathbf{I} - U\mathbf{r})}{2\sigma^2}\right) \quad (2.8)$$

The goal of Equ.2.4 can also be realized by finding the optimum values for the minimum of $-\log$ of the posterior probability-

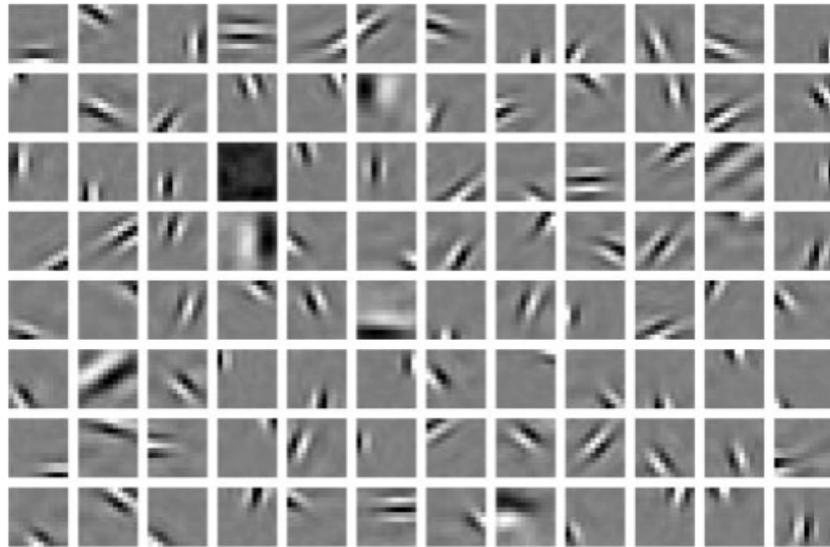
$$E = -\ln P(\mathbf{r}, U|\mathbf{I}) = \frac{1}{2\sigma^2} \underbrace{(\mathbf{I} - U\mathbf{r})^T(\mathbf{I} - U\mathbf{r})}_{\text{Square error term}} + g(\mathbf{r}) + h(U) \quad (2.9)$$

Where, $g(\mathbf{r})$ and $h(U)$ are the negative \log of prior probabilities $P(\mathbf{r})$ and U respectively. In information theoretic framework, E is known as *coding length*.

The first term of Equ.2.9 is a square error (weighted by the inverse of the variance). So, the fundamental task of vision can be seen as an inferential or prediction error minimization in the backdrop of the constraint of statistics of input (provided by prior probabilities $P(\mathbf{r})$ and $P(U)$).

2.2.3.2 Natural image statistics and sparse coding

A good prediction (inference/estimate), hence, a good vision, relies upon how good the internal model of the world is. The internal model of the outer world has been hypothesized to be a reflection of the statistical structure of our natural environment [51, 70] (for a good overview see [71]). This assumption is made on the basis of the fact that our visual system has been evolved to cope with the natural environment. Our natural world is full of regularity (redundancy) and hence predictable.

FIGURE 2.6: *Learned RF with sparse prior*

Being adaptive to the more frequent structure and regularity would provide the computational advantages in the form of inference [72] as well as transmitting information forward *efficiently* by reducing redundancy [11, 51, 73, 74].

Statistical regularity in natural scenes, therefore, could be expected to be reflected in the properties of visual system. Olshahausen, recently, in seminal studies [12, 75, 76] named *sparse coding*, showed that the training a neural network with natural image in the constraint of of sparse, kurtotic prior probability distribution (using the equation 2.9 with discarded last term) like, $P(r_i) = \exp(-\alpha \log(1 + r_i^2))$, leads to the development of gabor like receptive field (RF) of simple cell in the network (see Fig. 2.6)

Sparse coding principle, simply, constraint the coding length such that only small set of neurons from the available larger pool would be able to generate \mathbf{I}' (and hence code \mathbf{I}) as close as possible to an incoming signal \mathbf{I} . This constraints is the implementation of one of the hypothesis which comes under the umbrella of *efficient coding*, which, in the backdrop of limited anatomical and metabolic capital, argues that the goal of early sensory systems is to recode the sensory input in a way to reduce redundancy in natural scene without the loss of relevant information.

Many other studies [72, 76] have explored the statistical models of natural images and its relation to the properties of the visual system. For example, the structure of the receptive field of the ganglion/LGN cells [73, 74, 77], complex cells [72] and the cells at the V2 [78], has been shown to be a direct reflection of natural image statistics.

2.2.4 Predictive Coding

The biologically plausible theory for Bayesian inference is provided by Predictive coding [8, 59, 68]. In this approach, the visual system achieves the fundamental goal of encoding possible causes of sensory inputs (and learning the internal model) by concurrent prediction - correction mechanism, which, as we have discussed, is implemented by the generative model by minimizing the prediction error. The visual system constantly attempts to generate a top-down prediction of the input driving sensory signal on the basis of the learned regularity of natural scene, and signals forward just the residual bottom-up error signal to correct the initial prediction. This idea is based on the anatomical architecture of the visual system which is hierarchically organized and reciprocally connected [35].

More specifically, in a hierarchical predictive coding (HPC) model: the model of cortical processing, it is assumed that the image generation in the outer world involves hierarchical, multilevel, spatial and temporal interactions between the physical causes. The goal of the visual system is, thus, to inference (estimate, or internally represent) these multilevel hidden physical causes efficiently; which is accomplished by the visual system using recurrent prediction-correction mechanism along its hierarchy (See Fig 2.7a)

In past decade, HPC has gained growing support as the general coding principle of the cortex. Recently, several neuronal tuning properties in different visual areas such as the lateral geniculate nucleus (LGN), primary visual cortex (V1) and middle temporal level (MT) have been explained using this framework [79–82]. For example,

in his standard HPC model, Rao [8] suggested that the extra-classical properties of neurons in V1 could be understood in terms of the predictive-feedback signal from the secondary visual cortex (V2) which is made in a larger context and in the backdrop of learned statistical regularity of the natural Scene.

Though HPC has supported perception, action and attention in implemented by cortex in general (For an excellent review see [9, 10, 83], in this thesis we concentrate on the visual cortex and the visual perception.

2.2.4.1 General Architecture of HPC

In this framework, on the arrival of an input, predictor estimator modules (PE module), at each visual processing level, generate the prediction (or estimate) on the basis of the learned statistical regularities of natural scenes. Each higher area (say V2) then sends the generated prediction to its immediate lower level (say V1) by feedback connections and in return receives the error signal, by feed-forward connections, which is used to correct the current estimate. An equilibrium state is achieved after the completion of few prediction-correction cycle; where the estimate matches the input signal. This optimum-estimate is regarded as a representation of the input at that level. The achieved optimum-estimate at different levels of the network is depicted as a perception of the input image.

In general, a single PE module (See Fig 2.7b) consists of: (i) Predictive estimator neurons (PE neurons) which represent the estimate of current input signal \mathbf{I} with response vector \mathbf{r} (state vector), (ii) neurons, carrying prediction signal $U\mathbf{r}$ (for the input \mathbf{I}) to lower level by feed-back connections, whose synapse encode encoding efficacy matrix U , (iii) neurons, carrying feed-forward error signal $(\mathbf{I} - U\mathbf{r})$ from lower level to higher level, whose synapses encoded rows of efficacy matrix U^T , and (iv) error detecting neurons which carry the residual error signal $(\mathbf{r} - \mathbf{r}^{td})$ to the higher level corresponding to the prediction \mathbf{r}^{td} from the higher level.

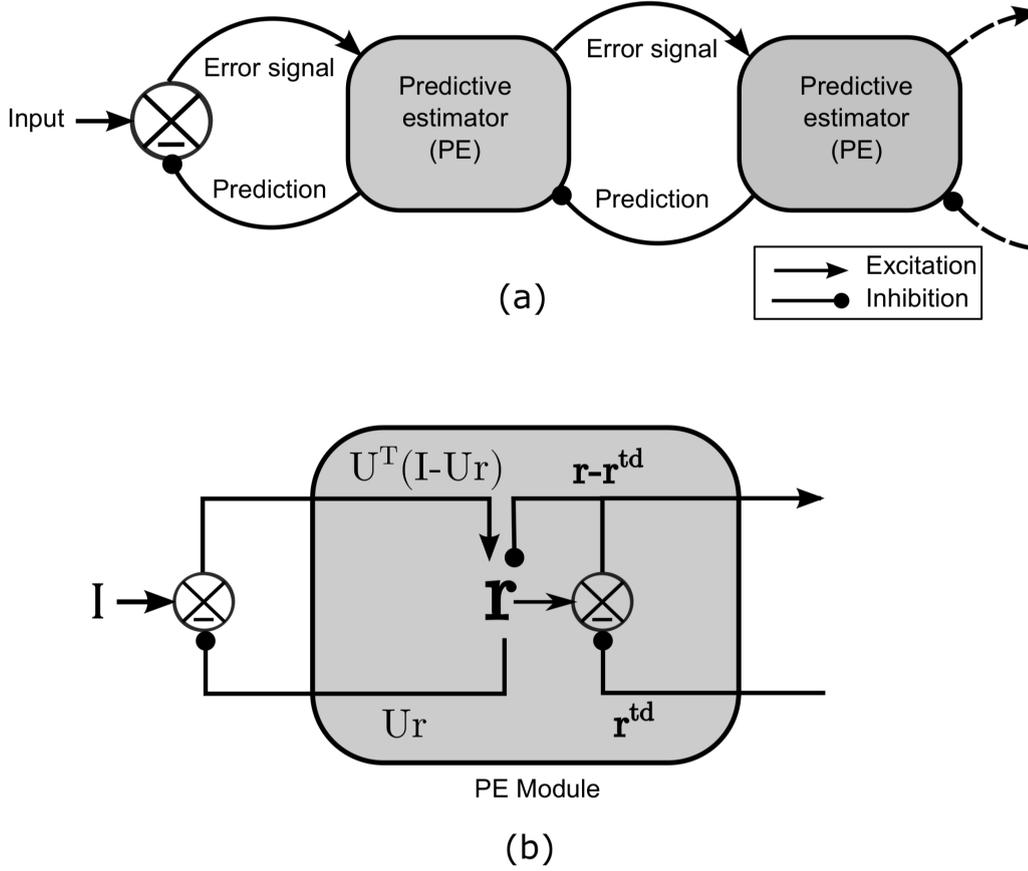


FIGURE 2.7: *General HPC architecture (adopted from Rao. [8]). a) On arrival of input, predictive estimator module at each higher visual processing level makes the estimate and sends prediction signal to its next lower level by feedback connection and receives the corresponding prediction error by a feed-forward connection. The error signal is used by the predictive estimator module to correct the estimate for better prediction. b) General predictive estimator (PE) module constitutes of (i) neurons to represent the estimate of the input I by their response vector \mathbf{r} by minimizing the bottom-up $(\mathbf{I} - U\mathbf{r})$ and top-down $(\mathbf{r} - \mathbf{r}^{td})$ error, (ii) feed-forward error carrying neurons has the efficacy matrix U , which encode the basis vectors their synaptic weights (or receptive fields), (iii) prediction $U\mathbf{r}$ carrying neurons and (iv) top down error detecting neurons.*

2.2.4.2 Network dynamics and learning rule of HPC

The dynamics, the learning rules and hence the above-mentioned architecture of a general PE module stem from the total coding length E_T (extended form of equation 2.9, for details see Appendix A) of hierarchical network-

$$E_T = \frac{1}{\sigma^2} (\mathbf{I} - U\mathbf{r})^T (\mathbf{I} - U\mathbf{r}) + \frac{1}{\sigma_{td}^2} (\mathbf{r}^{td} - \mathbf{r})^T (\mathbf{r}^{td} - \mathbf{r}) + g(\mathbf{r}) + h(U) \quad (2.10)$$

Here the hierarchal probabilistic generative model distributions $P(\mathbf{I}|\mathbf{r}, U)$ and $P(\mathbf{r}^{td}|\mathbf{r})$ are assumed to be as Gaussians of zero mean and variances σ^2 and σ_{td}^2 respectively.

Minimizing the coding length E , with respect to \mathbf{r} (using the gradient descent method) provides the dynamics of PE module as,

$$\frac{d\mathbf{r}}{dt} = -\frac{k_1}{2} \frac{\partial E_T}{\partial \mathbf{r}} = \frac{k_1}{\sigma^2} U^T (\mathbf{I} - U\mathbf{r}) + \frac{k_1}{\sigma_{td}^2} (\mathbf{r}^{td} - \mathbf{r}) - \frac{k_1}{2} g'(\mathbf{r}) \quad (2.11)$$

here, k_1 is a rate parameter that governs the rate of descent towards a minimum of E , and U^T is the transpose to weight matrix U . The steady state of this dynamical equation provides an optimum-estimate, which is regarded as the representation of the input.

As we discussed, coding length E , roughly, can be seen as the mean square error at the input and the output level of a PE module, subjected to constraints of prior probabilities. Minimization of the coding length is equivalent to optimization of estimate by recurrently matching of estimate to the corresponding “sensory driven” input from lower area as well as “context driven” prediction signal from the higher area. The prediction signal $U\mathbf{r}$ is the linear combination of basis vectors U_i 's. The U_i is the i^{th} column of the matrix U , and represents the receptive field for i^{th} neuron. The weighted coefficient in this combination, r_i , represents the response of i^{th} neuron having receptive field U_i . The visual representation of the prediction $U\mathbf{r}$ corresponding to optimum-estimate \mathbf{r} is, in this study, termed as “perceptual image.”

Furthermore, the minimization of coding length E , with respect to U using gradient descent method provides the learning rule for basis matrix U as,

$$\frac{dU}{dt} = -\frac{k_2}{2} \frac{\partial E}{\partial U} = \frac{k_2}{\sigma^2} (\mathbf{I} - U\mathbf{r})\mathbf{r}^T - \frac{k_2}{2} h'(U) \quad (2.12)$$

here k_2 is learning rate, which operates on the slower time scale than the rate parameters k_1 , and \mathbf{r}^T is the transpose of state vector \mathbf{r} . This learning rule can be seen as of Hebbian-type.

In the studies that follows in the next chapters, the prior probability, $P(\mathbf{r})$, on state vector \mathbf{r} , is chosen according to sparse coding. The kurtotic prior distribution ($P(r_i) = \exp(-\alpha \log(1 + r_i^2))$) constrains the dynamics for the sparse representation of the input. This distribution gives us:

$$g'(r_i) = 2\alpha r_i / (1 + r_i^2) \quad (2.13)$$

which is used in equation (2.11). The prior probability distribution, $P(U)$ has been chosen here to be Gaussian type, which finally gives us:

$$h'(U) = 2\lambda U \quad (2.14)$$

which is used in the equation (2.12). Here α and λ are variance related parameters.

There are things known and there are things unknown, and in between are the doors of perception.

Aldous Huxley

3

Mechanism of Filling-In at the Blind Spot

3.1 Introduction

Filling-in at the blind spot is one of the examples of how brain interpolates the informational void due to the deficit of visual input from the retina. Because of the absence of photoreceptors at optic disc, the retina is unable to send the corresponding signal to the brain and thereby, hides some portion of the visual field. The concealed visual field is known as the blind spot. However, we never notice any odd patch in our visual field, even in monocular vision, but rather we see the complete scene;

filled up in accordance with the surrounding visual attributes (i.e., color, brightness, texture or motion) [1]. This phenomenon is known as filling-in at the blind spot.

There have been three main hypotheses speculating the mechanism that underlie perceptual filling-in at the blind spot: active completion, passive re-mapping completion and, completion by ignorance. In the active completion hypothesis, it has been proposed that some active neural process in the retinotopic representation of the blind spot (BS area) in the visual cortex should be responsible for filling-in [1, 84, 85]. In other words, there should be a correlated neuronal filling-in corresponding to the perceptual filling-in in the BS area of the cortex. Whereas, passive re-mapping hypothesis argues that the filling-in at the blind spot could be a result of passive cortical re-mapping, where the informational ‘hole’ in the BS area is ‘sewn up’ so that the neighboring points in the cortex receive input from the opposite side of the blind spot [86, 87]. Completion by ignorance hypothesis, on the other hand, proposes that the filling-in does not involve any kind of neural process; the brain simply ignores the informational void [88–90].

The converging evidence from the recent psychophysical, neurophysiological, neuroimaging studies supports the idea that the filling-in is an active process: some neural computation is involved in the process of filling-in. Moreover, these studies suggest that such process mainly occurs in the early visual cortical areas. For example, studies [3, 84] on monkeys show that perceptually correlated neural activities are evoked in the deep layer of primary visual cortex, in the region that retinotopically corresponds to the blind spot (BS) region, when filling-in completion occurs. In another experiment, Matsumoto et. al. [4] showed that some neuron in BS region in deep layer of primary visual cortex (BS neurons), which possess larger receptive fields that extend beyond the blind spot, exhibits non-linear elevated response when a long moving bar cross over the blind spot and perceptual completion occurs. Fig.3.1 shows the schematic summary of Matsumoto experiment; while the drifting end of the bar was inside the blind spot, the perception of the bar was of a short isolated one and

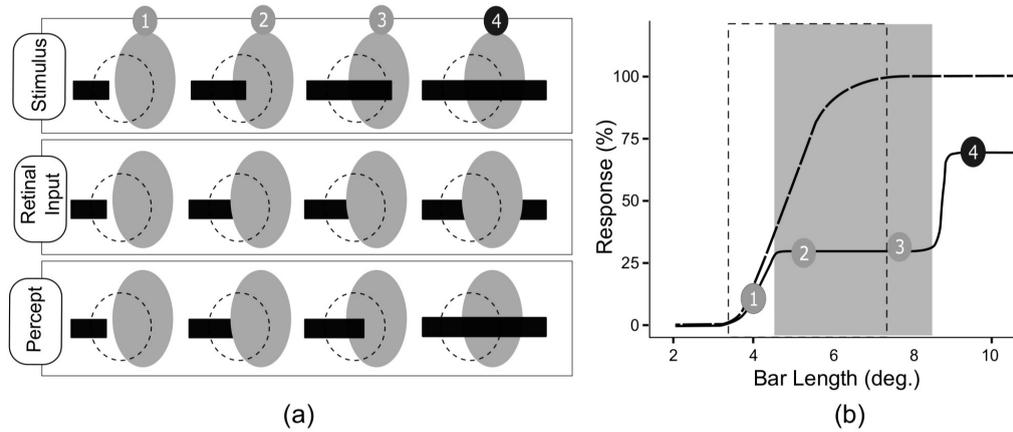


FIGURE 3.1: Schematic illustration of bar completion experiment (adapted from Matsumoto and Komatsu [5]). (a) The gray oval area represents the blind spot, whereas the dashed circle represents the receptive field of a neuron. The actual stimulus, the corresponding retinal input, and perception at position 1, 2, 3 and 4 are shown. One end of the bar stimuli was kept fixed outside the BS region, and the other end was free to drift across the blind spot. (b) The response of a typical neuron in BS region at the deep layer of primary visual cortex is presented. The gray rectangle indicates the blind spot and the dotted rectangular area represent the receptive field of the typical neuron. The solid line is the response obtained through the eye connected to the blind spot (BS eye) under review, and the dotted line is the response of the same neuron obtained through the fellow eye.

the corresponding neural responses were low and constant. However, the moment bar end crossed the blind spot, the neural response elevated rapidly and completion of the bar was perceived. These elevated response exhibit nonlinearity; the response to the long bar that stimulates simultaneously the both sides of the blind spot was larger than the sum of responses to the stimuli presented on either side of blind spot separately.

These studies explain some fundamental question about filling-in: whether some neural computation is required for the filling-in to occur, and the involvement of early visual cortex in the computation. However, there are some more fundamental questions that remained unanswered. What computational and neural mechanism is involved in the filling-in process? Whether the filling-in is involved in a special mechanism or some general computation mechanism can account for the same? How the observed nonlinearity corresponding to filling-in could be understood in that computational framework.

In this chapter, we will try to throw some the light on these questions under the framework of the Hierarchical Predictive coding of natural images. To do this, we have conducted simulation studies involving horizontal bars on three leveled (LGN-V1-V2) HPC model network having a blind spot which was emulated by removing the feed-forward (LGN-V1) connection. In this investigation we have employed shifting bar stimuli as described in Fig3.1, to study the properties of our model network and recorded the model predictive estimator neurons (PE neurons) at BS region in the V1. We found that these neurons exhibit, similar to the one observed in experiments, non-linear response and represent the filling-in completion when bar crosses the blind spot. These results suggest that the filling-in process could naturally arise out of the computational principle of hierarchical predictive coding of natural images.

3.2 Simulation

3.2.1 Network

A three-level linear hierarchical predictive network (See Fig 3.2) is simulated to imitate the early visual cortex. Level 0, Level 1 and Level 2 is equivalent to LGN, V1, and V2 area respectively. Level 1, consist of 9 modules, receives input from level 0 and sends the output to the solitary module at level 2. The PE module at level 2 receives input from all the nine level 1 PE modules and sends back the feedback signal to all of them. This architecture is based on the fact that the visual area higher in hierarchy operates on a higher spatial scale.

Each of nine PE modules at level 1 consists of 64 PE neurons, 144 prediction carrying neurons, 64 afferent error carrying neurons and 64 error detecting neurons for conveying the residual error to level 2. The layer 2 module consists of 169 PE neurons, 576 prediction carrying neurons and 169 error carrying neurons.

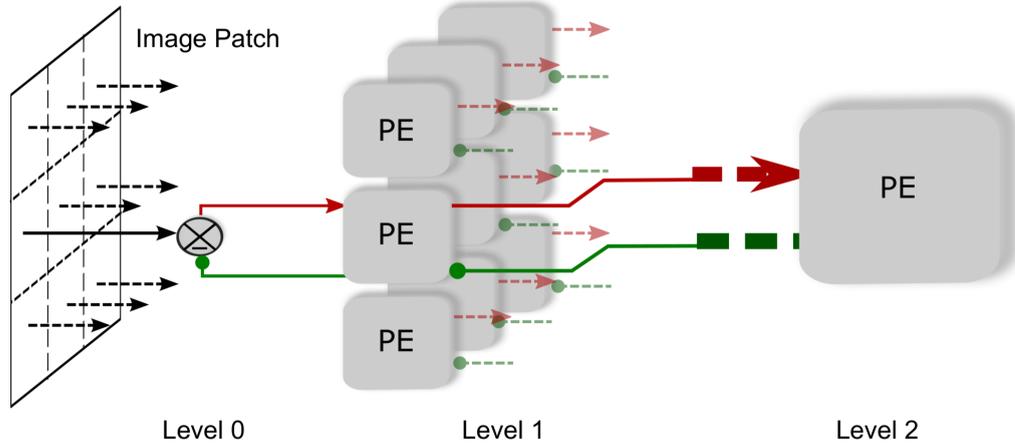


FIGURE 3.2: Three level HPC model network. Each of nine level 1 PE modules sends prediction to level 0 by feedback connection and receives error signal corresponding to their local image patches by a feed-forward connection. On the other hand, the PE module at level 2 sends the prediction signal to all level 1 modules and in reply, receives the error signal collectively from all these modules. Level 2, therefore, encodes larger visual patch and hence possess the larger receptive field.

3.2.2 Training

Six images of different natural environment were used for the training (see Fig3.3a). These images were preprocessed first to compensate the early visual processing (Ganglion cell, LGN) [77]. This pre-processing involved DC removal and the filtering of images with circular symmetric whitening/lowpass filter with spatial frequency profile $W(f) = f \exp(-(f/f_0)^4)$ (see [12, 75]). Here, Cutoff frequency f_0 was taken to be 200 cycles/image. Thereafter, Variance normalized 1000 batches of 100 image patches of size 30×30 pixel, which were extracted from randomly selected locations from the randomly selected pre-processed images, were given as input to the network. A single 30×30 -pixel image consisted of nine tiled 12×12 -pixel image patches, which were overlapped by 3 pixels (see Fig3.3b) and which were fed to the corresponding level 1 PE modules. For each batch of image patches, the network was allowed to achieve steady states (according to the equ. (2.11)) and the average of these states was used to update the efficacy of neurons (according to the equ. 2.12), initially assigned random values. During training, to prevent the efficacy vectors U_i (columns of U or rows of U^T) from growing unbound, the gain (L2 norm), $l_i = \sqrt{U_i \cdot U_i}$, were adapted, as

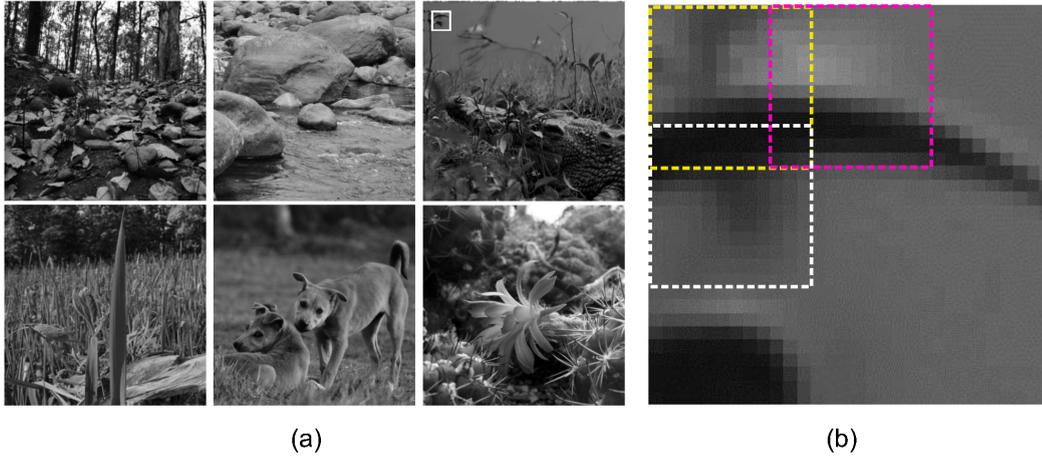


FIGURE 3.3: *Natural images.* a) These images, taken from of different natural environments, are used for simulation. b) A typical sample of 30×30 pixel image patches extracted from the natural image (top rightmost) from the position shown by the white rectangle. Each of these patches is broken down to 9 sub-patches of 12×12 pixel each with 3 overlapping pixels. Three such sub-patches are shown here by three dotted rectangles in yellow, magenta and white. Each of these sub-patches forms the local input to the 9, level 0 modules in the HPC model network.

$l_i^{new} = l_i^{old} (\langle r_i^2 \rangle / \sigma_{goal}^2)^\gamma$, so that the variance of r_i remain at appropriate level (for the details see [75]). Here σ_{goal}^2 is desired variance of r_i and α is gain adaption rate. The level 1 was trained first and then the level 2. Parameter values used in this study are: $k_1 = 1$, $k_2 = 3$, $\sigma^2 = 3$, $\sigma_{td}^2 = 10$, $\alpha = 0.05$ at level 1 and 0.1 at level 2, $\lambda = 0.0025$, $\sigma_{goal}^2 = 0.05$, $\gamma = 0.02$.

3.2.3 Blind spot implementation

To mimic the blind spot the feed-forward connection in a certain area was removed from the model network, which was pre-trained with usual feed-forward connections. The removal was implemented by setting the efficacy of early feed-forward (level 0 - level 1) neurons, that carry the error signal corresponding to the middle region (of size 8×8) of input patches (of size 30×30), to zero (see Fig.3.4). This “pre-training”, the training before the creation of the blind spot, captures the fact that the active neurons in deep layer (5/6) corresponding to filling-in has been reported to be of the binocular type. These neurons were found to respond to the inputs from both eye

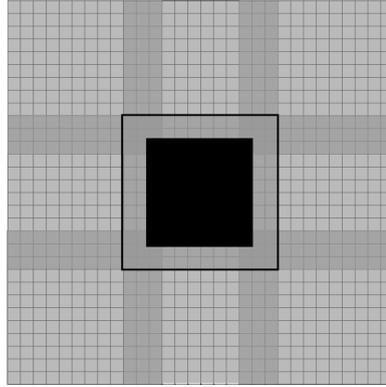


FIGURE 3.4: *Model blind spot. A 30×30 connection map is shown. Central dark square (8×8 pixel) indicates the connection-less area corresponding to the blind spot. The outside square outside of this indicates area coded by central module (BS module) of the network. The shaded patches indicate the overlapped areas (of 3 pixel width) shared by neighboring modules.*

and hence, possess binocular receptive field. Additionally, these neurons also exhibit greater sensitivity to the inputs from the other eye (non-BS eye) [3,4]. It is, therefore, natural to assume that, in normal binocular vision the feed-forward input from the non-BS eye will cause the receptive fields (of these deep layer neurons) to develop.

3.3 Results

The HPC network was allowed to learn the synaptic weight of model neurons, by exposing it to natural image patches under the constraints of the sparseness of model neuron responses (see method). The learned synaptic weights of neurons carrying feed-forward signal of one of the modules at level 1 and level 2 are shown in Fig 3.5. The weighting profiles at level 1 (Fig 3.5a) resemble the Gabor-like receptive field at V1, which is similar to the results reported earlier in several studies [8, 12, 80]. The weighting profile at the level 2 (Fig 3.5b) resembles the more abstract visual features: long bar, curve, etc. The blind spot was emulated in the network by removing feed-forward connection (see method), whereas, the training was performed on a network by keeping this connection intact. We designate the network with the blind spot as BS network and the one without the blind spot as a non-BS network.

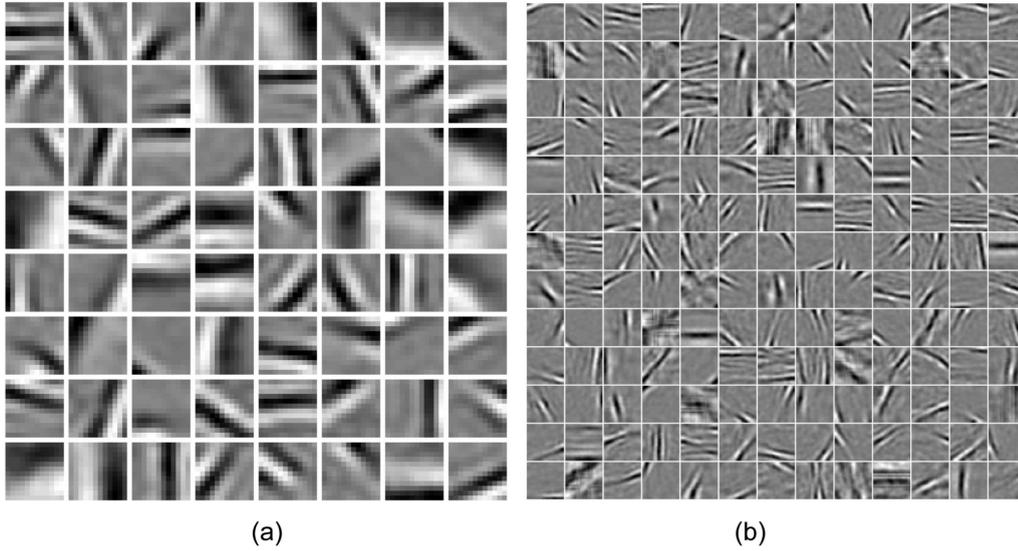


FIGURE 3.5: *Learned synaptic weights. (a) Learned receptive field of 64 feed-forward neurons of the size 12×12 pixels of BS module at the level 1 and, (b) Learned receptive field 169 feed-forward neurons of the size 30×30 at the level 2.*

3.3.1 Filling-in of shifting bar

We simulated the bar-shifting experiment on the trained HPC model network. Both BS and non-BS Network were exposed to images of a horizontal bar of different length. One end of the bar was fixed at a position outside of the blind spot, whereas, the position of other end was varied (by one pixel at each instant) across the blind spot. Images of the bar for six different end positions are shown in Fig 3.6a. The response vector, \mathbf{r} , of PE neurons in the central module in the model network (let say BS module) at level 1 and the sole module at level 2 was recorded for the different end position of the bar.

Fig 3.7 shows the bar plots of the response of 64 neurons in BS module at level 1 for six different bar position in both model networks(BS and non-BS). The comparison shows that almost the same set of a small number of neurons responded in both networks. The receptive field of the highly responsive neurons in this set possesses a horizontal bar-like structure. We plotted the response of some of these highly responsive neurons against the bar position (varying by one pixel) (see Fig 3.8) which show that these neurons exhibited elevated response when the varying end of bar

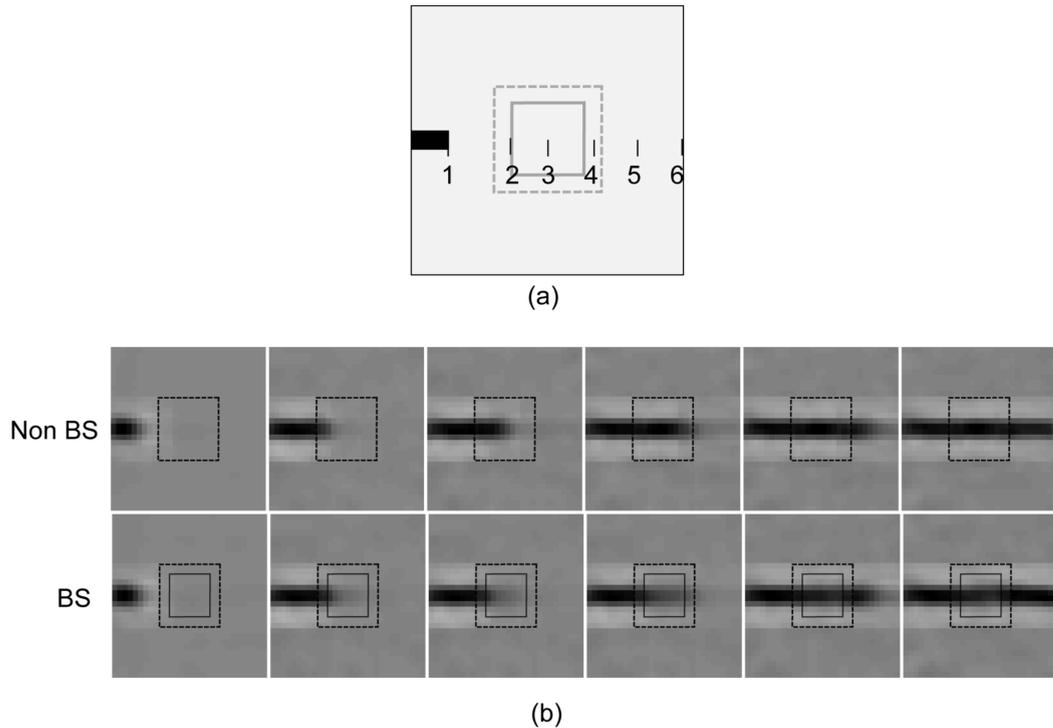


FIGURE 3.6: *Shifting-bar investigation.* a) A typical 30×30 pixel stimulus is shown here. The darkened object in the stimulus is a bar, whose endpoint is represented by the number 1. Five more stimuli were constructed by shifting the bar end to positions 2 to 6. The larger rectangle of size 12×12 pixels (shown by the dotted line at the center) indicates the extension of BS module and the smaller one of size 8×8 (shown by the solid line) indicate the extension of the blind spot. b) Generated 30×30 “perceptual images” corresponding to response profile of PE neurons at level 1 of the HPC network for non-BS (top row) and BS (bottom row) cases are shown.

crosses the blind spot. These elevated responses, in BS network, become reasonably close to the maximum response exhibit by these neurons in the non-BS neuron. The closeness of responses indicates the representation of objects in the BS network is similar to the one in the non-BS network. This is reflected in the corresponding “perceptual images” (see Fig 3.6b) reconstructed using the generative process. The response profile of level 2 neurons is shown in Fig 3.9. The most active PE neurons at level 2 exhibit similar response as level 1 neurons and possess horizontal bar like the receptive field, which is quite expected.

It is evident from these results that, in the case of BS Network, as long as the bar end remained inside the blind spot the response of neurons, in the BS module, remained constant and relatively low (Figs 3.7 and 3.8) which results in the perception of a

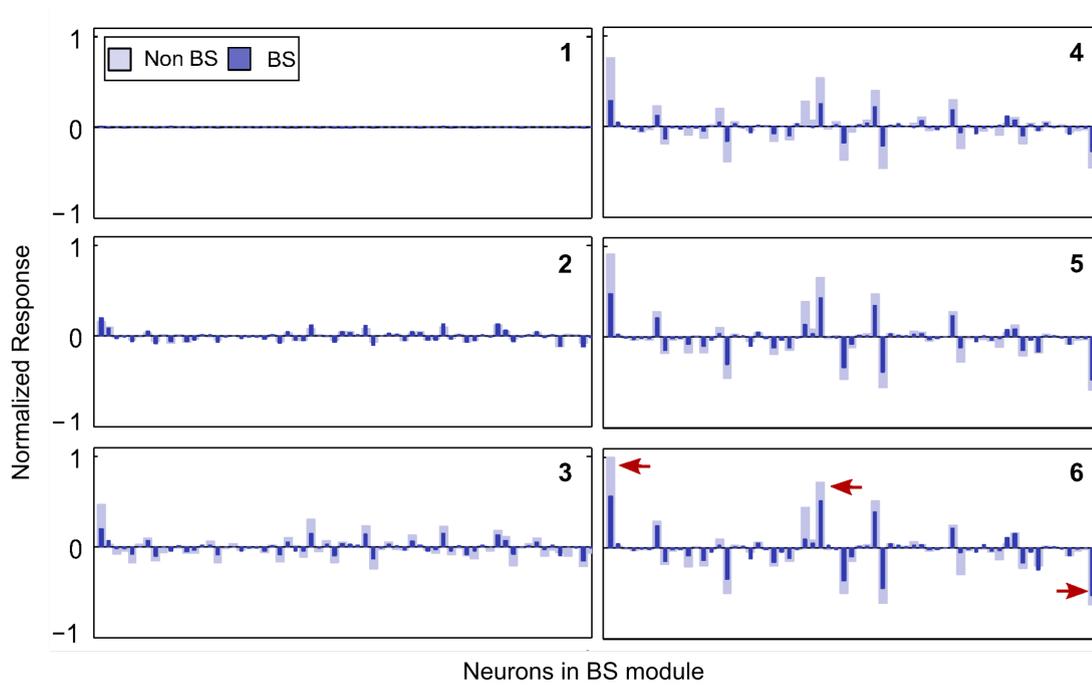


FIGURE 3.7: Responses profiles. Normalized responses of 64 PE neurons at BS module, corresponding to the six stimuli discussed in Fig 3.6a are presented. The dark blue bar represents the response of PE neurons for the BS network, whereas, the light blue bar represents the responses for the non-BS network. Three most highly active neurons (in bottom leftmost bar plot) are marked by red arrows.

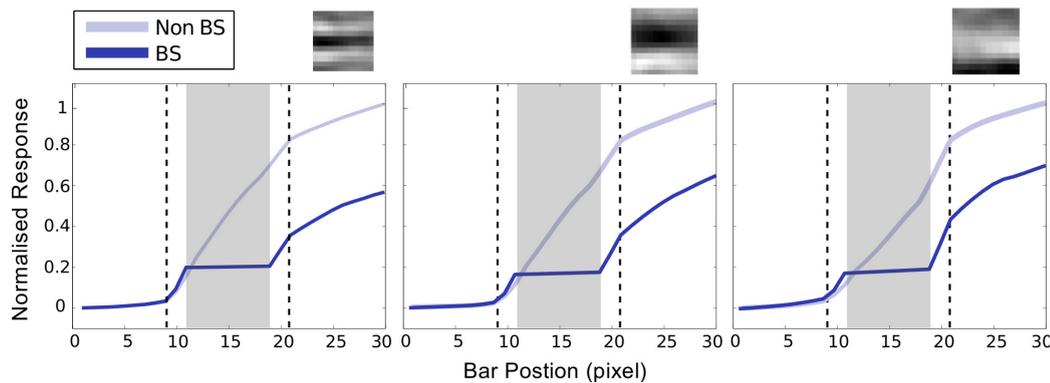


FIGURE 3.8: Response elevation in BS region at level 1. Plots of the absolute value of normalized response are shown against the bar position for three highly active neurons (indicated by red arrows in the sixth bar blot of Fig 3.7) In these plots, dotted rectangular area indicates the extension of BS module whereas, the solid gray rectangular area indicates the extension of the blind spot. The receptive fields of these three neurons are shown at the top of the respective plots, which show that these neurons participated in encoding information of a horizontal bar. To compare the relative activity of the neurons we have plotted the absolute value of the responses instead of signed values of responses

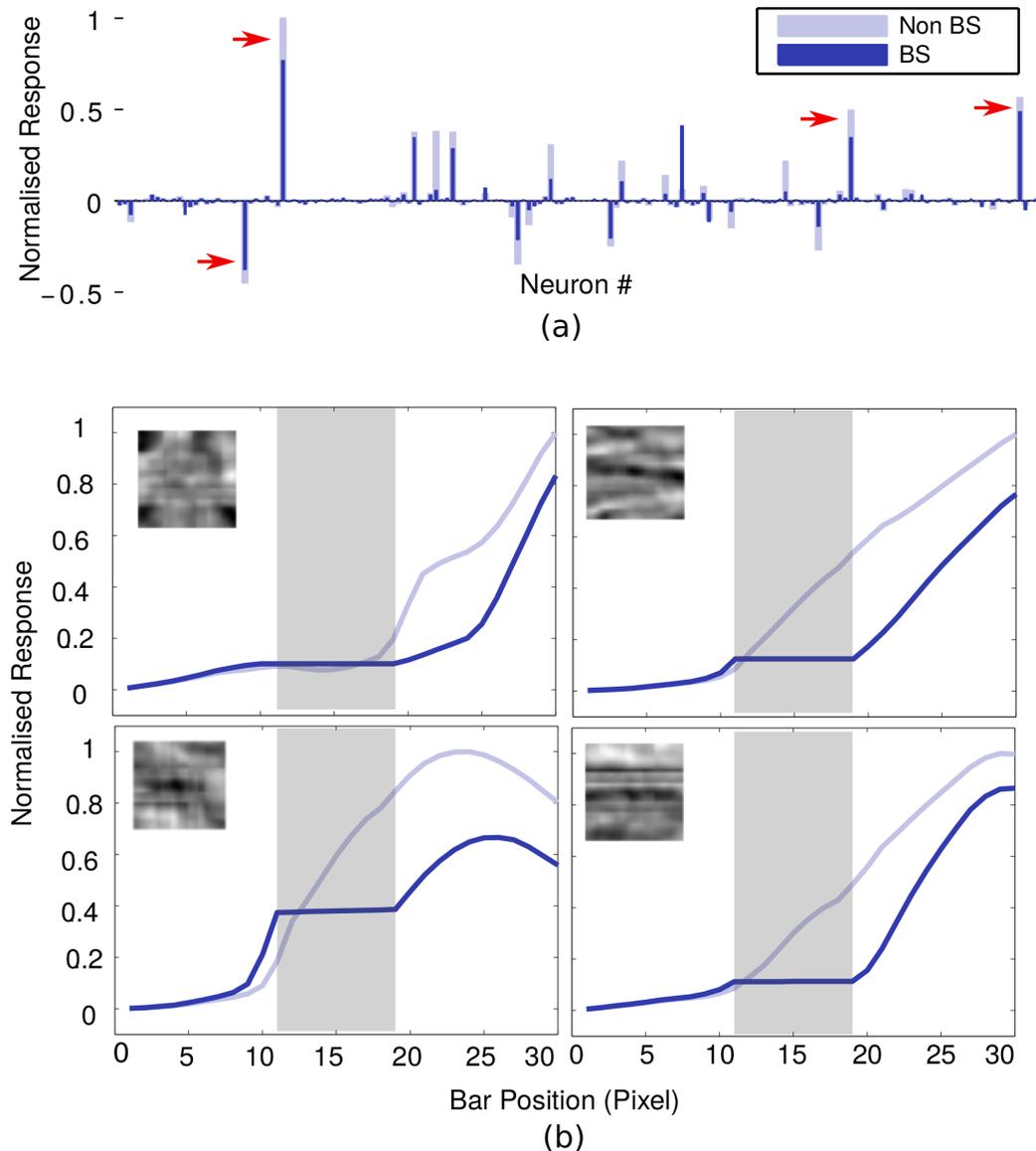


FIGURE 3.9: *Response profile at level 2. (a) The normalized response of 169 neurons at level 2 corresponding to shifting bar stimuli for the end postilion 6. (b) Plots of the normalized absolute value of response of most active neurons at level 2 (marked as red arrow in (a)). The receptive field of these neurons is shown in the inset of their corresponding plots.*

bar of constant length on one side of the blind spot. On the other hand, when the bar crosses the blind spot, the responses are elevated significantly, and the filling-in completion occurred. These could also be understood by observing the relative deviation of the response of each neuron (typically the highly responsive neurons) in both networks (Fig 3.7) when bar end crosses the blind spot. These results are consistent with the findings of neurophysiological studies on macaque monkeys [4].

It is evident from Fig 3.8 that when filling-in occur, the response profile changes abruptly in a non-linear fashion. In order to verify the explicit correlation between the non-linear response and the filling-in completion, the response of the most responsive neurons was examined by exposing the BS network with different stimulus combinations (a, b, c, ab and a+b) as described in Fig 3.10. The responses of these neurons to these four stimuli were compared. Stimuli are shown in the inset at the bottom of the figure and the corresponding responses are presented as bar plots above each stimulus. As shown in Fig 3.10, the response to ab stimulus is significantly larger than the sum of the responses to a and b (shown as a+b) even though stimuli a and b separately stimulated similar areas of the RF exposed ab stimulus. This indicates that the abrupt change in the magnitude of the response during filling-in completion can not be explained by the summing over the responses of the receptive field extending out from the opposite side of the blind spot. In other words, the abrupt response increase (non-linearity) is correlated with the perceptual completion and is not predictable from a simple summation rule.

3.4 Discussion

To understand the mechanism of filling-in, we should recall that in HPC, feed-forward connection propagates up the residual error, corresponding to the current prediction made by higher area, for the betterment of the next prediction. The optimum-estimate, where prediction closely matches the “driving sensory” input as well as “contextual signal” from the higher area, which produces a minimum prediction error, is then depicted as a perception of the input. However, the blind spot is characterized by the absence of such feed-forward connection. Therefore, the estimate made by higher areas prevails in the absence of error signal and this provides the ground for the filling-in completion to occur at the blind spot.

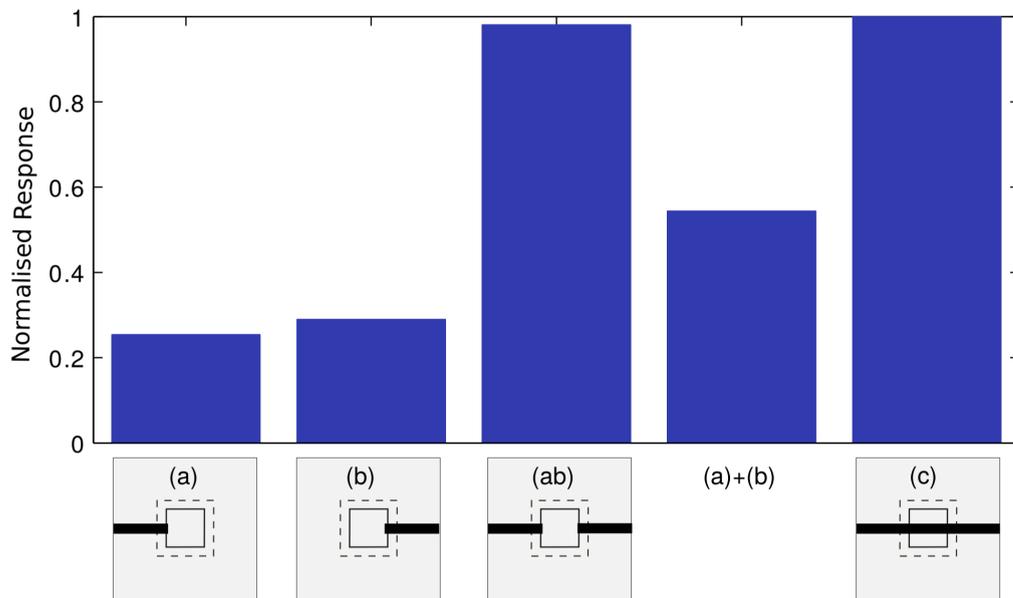


FIGURE 3.10: *Nonlinearity in the response profile.* Average normalized responses of neurons (in BS module) to various stimulus conditions are presented. Estimated responses to stimuli a , b , c and ab , as well as the sum of the responses to a and b , are shown, where ab is a combination of stimuli a and b . Each stimulus is schematically shown below each bar plot, where each bar plot shows the mean of normalized responses of 8 most responsive neurons in the BS module. Conventions are same as shown in Fig 3.8.

Neurons at V2 learn the regular feature like long bar, curve etc. from the natural scene and operate on a larger area and context. Therefore, the initial estimate, which is based on these learned regular features, prevails and becomes an optimum-estimate for the inputs which only match with those regular features in the surrounding of the blind spot. This process initiates the filling-in at the V2. In the absence of feed-forward connection in BS region, the corresponding local optimum-estimate at level 1 will, therefore, evolve by matching the “context-driven” feedback signal from level 2. This process at level 1 locally captures all the course of the completion process at level 2. Thus, the properties of the filling-in are primarily determined by the matching of statistics of input stimuli around the blind spot and natural statistics learned by the network. Better the degree of matching, higher chances of completion. For example, in the bar shifting experiment while the moving end of bar remained inside the blind spot, the incoming sensory input, which is of a short bar residing on

one side of the blind spot, deviates reasonably from learned statistical regularity, in which the bars are usually longer (extended across the blind spot) [8]. That turns out as a non-completion of the bar and PE neurons at BS module, whose response represent bar, exhibit low response. On the other hand, when bar crosses the blind spot, the likelihood of matching of the input signal and the learned regular features (long and continuous bar) suddenly increases and that leads to the abrupt elevation of the response of PE neurons which encodes the bar in BS module. This process resembles the AND-gate functionality and reflects the nonlinearity in response profiles shown above (Fig. 3.8 and 3.10).

Regarding the “pre-training” mentioned in the method section, one can argue that in the absence of feed-forward input, even the feedback signal from the higher area can possibly cause the associative receptive fields to develop, which could provide the basis for internal representation in the BS area. But this is not the case with neurons representing the blind spot in which, as we have already discussed, there seems to exist the feed-forward connection from the other eye (non-BS) in a normal binocular vision to cause receptive fields of the deep layer neurons to develop. In this case, one can suggest that these neurons might get relatively reduce input strength in the BS region since these are getting input from only one eye rather than from both eyes and that can lead to different weighting profile. Without going into details of the integration of input, which can take the value of relative strength from half to one depending on assumption of integration (linear or nonlinear), we can discuss that even the reduction of input, in a reasonable amount, may not give rise to any qualitative change in the learned receptive fields of the neurons (see Appendix C) because the nature of the receptive fields is mainly governed by the statistical feature of the input. We, therefore, argue that this situation may not alter the generality of our approach.

In conclusion, recent studies on filling-in at the blind spot reveal that neural activities in BS area, in the deep layers of V1, are associated with filling-in completion. For

example, in the shifting bar completion experiment, nonlinear neuronal activities are reported. Here, we have explained these activities, through modeling, in terms of hierarchical predictive coding principles of natural images and moreover, we have also demonstrated that these activities represent the filling-in operation. These results suggest that the filling-in could be a manifestation of a hierarchical predictive coding principle and, the nature of filling-in could be predominantly guided by the learned statistical regularities of the natural scene.

*Man is not himself only... He is all that he sees; all that flows
to him from a thousand sources... He is the land, the lift of its
mountain lines, the reach of its valleys. . . .*

Mary Austin

4

Properties of the Filling-In

4.1 Introduction

In the previous chapter, we demonstrated how the filling-in phenomenon could arise from the general computational mechanism of HPC. We speculated that the properties of filling-in could be explained on the basis of the properties of the natural image statistics.

In this chapter, we will investigate mainly two reported properties of filling-in: tolerance and anisotropy. When two aligned bars are presented on opposite sides of the blind spot such that the gap fully falls inside the blind spot (essentially equivalent

to a bar extended the both side of the blind spot, as we discussed in the previous chapter), the bars are usually perceived as a continuous one. Moreover, this filling-in continues (however, in reduced strength) to occur up to a certain degree of differences, in alignment or orientation, between the bar pair. This property is known as tolerance of filling-in.

Psychophysical investigations also revealed that filling-in and the tolerance of filling-in are not isotropic but sensitive to orientation configuration (horizontal, vertical) of the stimulus (bar-pair). Araragi et. al. [7] have demonstrated that a certain minimum critical length of bar-pair (extended beyond the blind spot) is required in order of filling-in to occur. For the horizontal configuration, this critical length is shorter in comparison to the vertical configuration. He also showed that for the identical length, horizontal configuration tends to have better filling-in over vertical configuration. This phenomenon is designated as anisotropy in filling-in. In his other studies, Araragi et. al. [6, 14] showed that the anisotropy can also be observed in the tolerance of filling-in. However, contrary to the horizontal dominance in anisotropy in filling-in, vertical bar exhibited higher tolerance over the horizontal configuration in filling-in: vertical bar-pair appears continuous for a comparatively larger difference in alignment or orientations. This phenomenon is designated as anisotropy in filling-in-tolerance.

Anisotropy has also been reported in other visual phenomena related to orientation perception. Studies with grating stimuli show that visual sensitivity and acuity of human (and other species) is typically better at cardinal (horizontal and vertical) orientations than at oblique orientation (those ± 45 degree from cardinal) [91–93]. This has been termed the ‘oblique effect’ [94]. On the other hand, studies involving natural broadband stimuli show the opposite where oblique orientations have upper hand over cardinal ones [15, 16, 19, 95, 96]. This phenomenon is known as ‘horizontal effect’. These studies brought out the differences in bias between horizontal and vertical orientation and demonstrated that our visual system favors horizontal configuration

over vertical. It has been suggested [15,96–98] that the statistics of natural scenes is primarily responsible for the observed anisotropy in the orientation perception. Image analysis reveals that the orientation content in natural scenes is biased more towards horizontal than vertical, and the least bias is observed towards the oblique. This asymmetry raises a logical question whether the orientation-selective neurons in the cortex are influenced by the prevalence of horizontal orientation in the environment during development. Indeed, it has been demonstrated experimentally [17, 96, 99–101] that adult ferret and cat V1 contains an over-representation of neurons coding horizontal orientations.

How the properties of natural scene statistics influence our perceptual judgment? To be more specific, whether the over-representation of features in the natural scene has any role in the emergence of the observed tolerance in filling-in at the blind spot? We investigate this question using a three level HPC model network.

Firstly, we investigate the tolerance of filling-in for the two stimulus configuration: misaligned and disoriented bar-pair. Then, secondly, we examine the anisotropy in filling-in and anisotropy in filling-in-tolerance in the vertical and horizontal stimuli configurations. We observed a correspondence between the anisotropies related to the orientation of feature found in the natural scene and apparently different anisotropy observed in various psychophysical experiments. Moreover, this chapter explicitly brings out the possible relation between natural scene statistics, cortical organization and the perceptual experience at the blind spot.

4.2 Simulation

A similar three-level model network and training procedures as described in the last chapter were used for simulation. However, for robust and statistical rigorous investigation, we increased the number of neurons at each module at every level. The

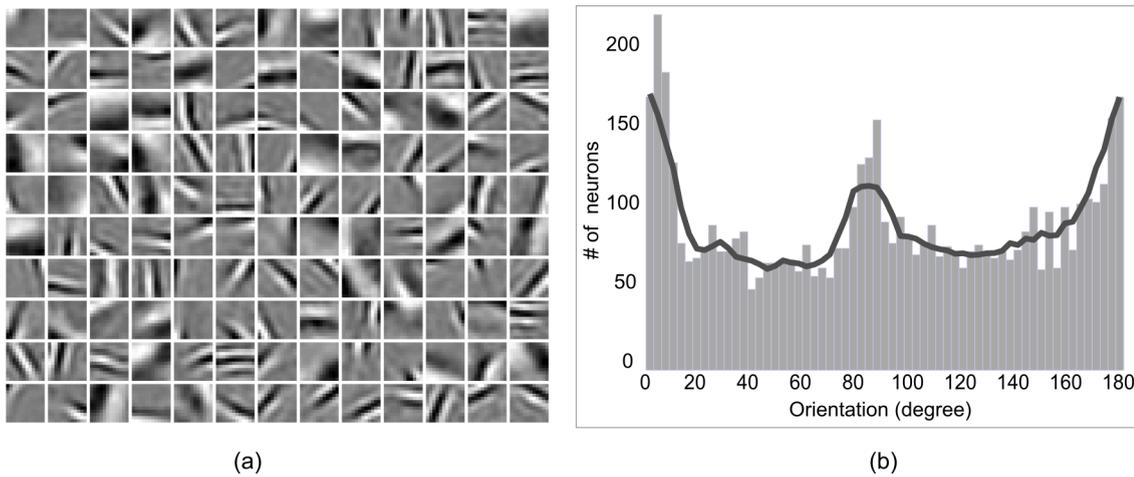


FIGURE 4.1: *Anisotropy in orientation selectivity (a) learned weighting profiles of 130 neurons at one of the 9 modules at level 1 for the single training set. (b) Orientation distribution at level 1 for all the neurons (130×40). The envelope (continuous line) is obtained from the running average of 7 bins of the histogram.*

modules at level 1 consist of 130 feed-forward, 130 PE neurons, and 144 feedback neurons. The level 2 module contains 256 feed-forward neurons, 256 PE neurons, and 1170 feedback neurons. All the parameters of the network in this chapter are same as taken in the previous chapter. To make the investigation more rigorous we also performed 40 different training cycles to learn 40 different sets of efficacy matrix (receptive field).

4.3 Results

The training resulted in 40 different set of Gabor-like weighting profiles at level 1 (Fig. 4.1a), as shown in the last chapter, which was distributed in the different orientation and spatial frequency and resembled the simple cell receptive field at V1. Level 2 weighting profiles resembled more abstract features (corner, curves, long bar etc.)

4.3.1 Tolerance of filling-in

To investigate the tolerance in filling-in for the misaligned bar, the model network was exposed to a pair of horizontal bar segments placed on both sides of the blind spot (see Fig.4.2a). One bar was kept fixed at one side of the blind spot while the position of the other one was shifted vertically in small steps to vary the misalignments. The response of PE neurons in BS module was recorded with changing misalignment. This process was repeated 40 times with 40 different cycles of training. Investigations with the different cycle of the training can be considered analogous to the psychophysical investigation performed on the different participant (human), which leads to more statistical rigors in the results. All the subsequent investigations reported in this chapter follows the same number of repetitions. From these responses, equivalent “perceptual images” were reconstructed, which are shown in Fig.4.3a. Perceptual images show that filling-in continues to occur (though in reduced strength) for a small degree of misalignment but cease to appear after a certain degree of misalignment.

To quantify the filling-in, pixel values in the middle (central 2×2 pixel wide region in the blind spot, indicated small red square in Fig.4.3a) of the perceptual image were averaged. We define this average as the ‘filling-in-value’, where the greater negative value indicates better filling in. Fig.4.3b shows the plot of this values, from all the perceptual images corresponding to 40 training, with misalignment of bar-pair. As we can observe, the filling-in value is greater in the case of perfect alignment and gradually deteriorates with increasing misalignment.

To investigate the tolerance in disorientation, the network was stimulated with a stimulus as shown in Fig.4.2b. The stimulus consists of a fixed bar and a rotating bar. The fixed bar is placed horizontally. The other bar, the test bar, was rotated in steps of 10 degrees from the aligned position (0-degree difference in orientation) to the perpendicular position (90-degree difference in orientation). We recorded the response of PE neurons corresponding to these stimuli configurations and perceptual images

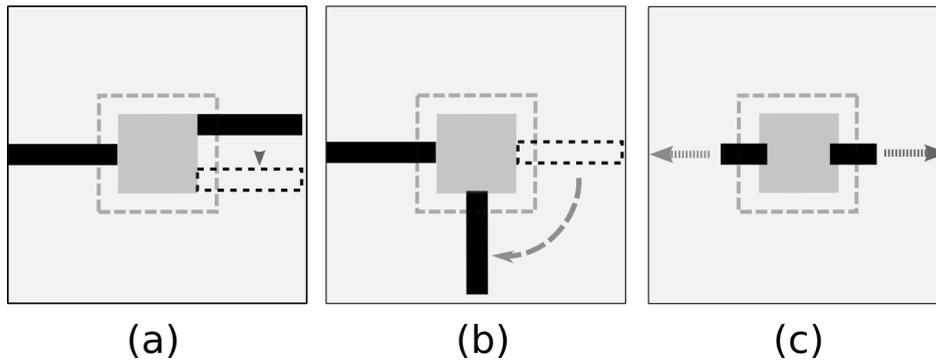


FIGURE 4.2: *Stimuli.* (a) *Misaligned bar stimulus:* Two bar stimuli are shown at the opposite end of the blind spot, which is indicated with the gray square (8×8 pixels) in the center. The dotted square (12×12 pixels) denotes the area exposed to the central module (called BS module) of one of the nine level 1 modules (see Fig.3.2). The bar at the left side of the blind spot remains fixed while the right side bar moves in the vertical direction in steps of one pixel every time (b) *Rotating bar stimulus:* In this case, the left side bar remains fixed but the bar at right side rotates in steps of 10 degrees. (c) *Expanding bar stimulus:* One end of both bars was fixed inside the blind spot, whereas other ends were expanding together in sync in steps of one pixel in opposite directions. Extension of bars has been measured from the border of the blind spot.

were generated accordingly. The perceptual images and the corresponding plot of filling-in-value for those stimuli are shown in Fig.4.4a and Fig.4.4b, respectively. These results show that the filling-in performance was better for the aligned bars but it deteriorated with increasing difference in orientations.

This result indicates that the filling-in completion depends on alignment and the filling-in is best for perfect alignment. Moreover, this filling-in continues (however, in reduced strength) to occur up to a certain degree of differences, in alignment or orientation, between the bar pair. These results are consistent with earlier psychophysical findings [6, 14].

4.3.2 Anisotropy

In this section, we test the hypothesis that the prevalence of certain features in natural scenes is capable of providing a mechanistic explanation of anisotropy related to the perceptual filling-in reported by human observers. Our objective is summarized in

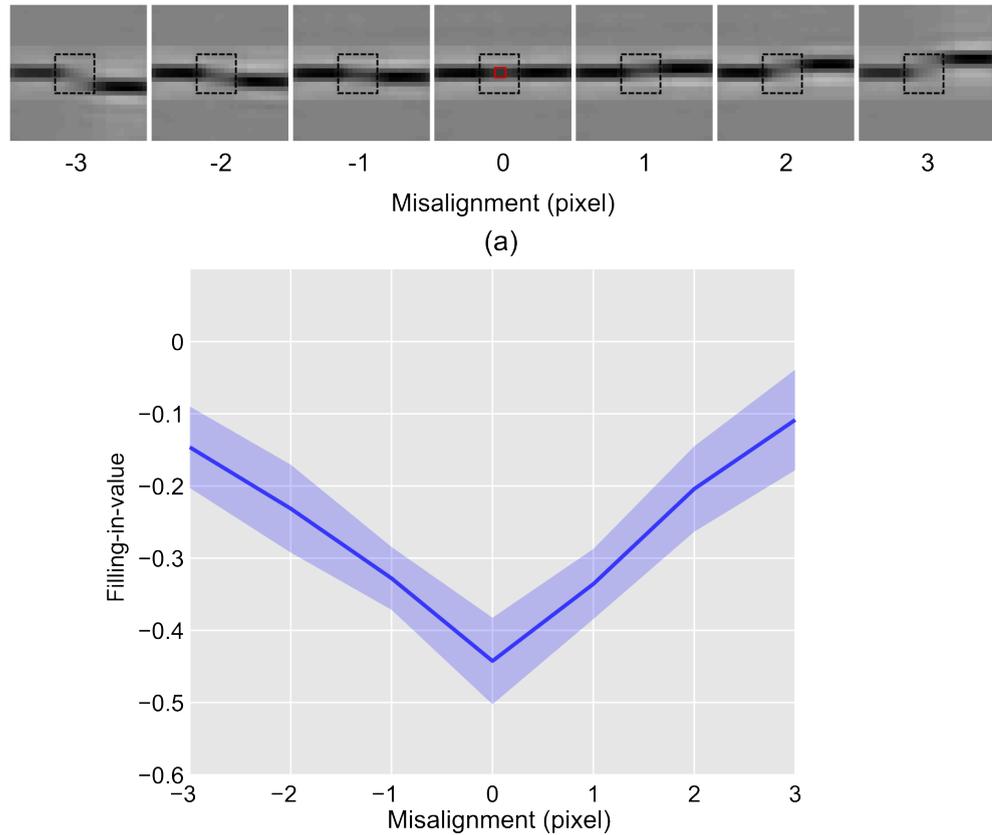


FIGURE 4.3: *Tolerance in misalignment* a) A typical stimulus is shown, where two non-aligned bar segments are presented on opposite sides of the blind spot. For our study, one bar was fixed (left one) and the other one was shifted vertically by one pixel per instant for the seven positions emulating seven stimuli. b) The generated “perceptual images” for those stimuli (as discussed in (a)) corresponding to the recorded response profile of PE neurons at level 1.

Fig.4.5, where we have schematically depicted the proposition that there is a link between the anisotropy present in the natural scene and the anisotropy reported in perceptual filling-in investigations. This supports the general speculation[6,8,10–12] that orientation anisotropy in natural scene plays a significant role in determining the anisotropy in the cortex as well as the anisotropy in perceptual orientation preference. As a premise, we first explored the capability of the HPC model network to learn the anisotropic distribution of features present in the natural image via training, which will validate the previously known results. Then we went on to investigate whether the learned statistics (learned internal model) could explain the anisotropy in filling-in and the anisotropy in tolerance of filling-in reported in other psychophysical studies.

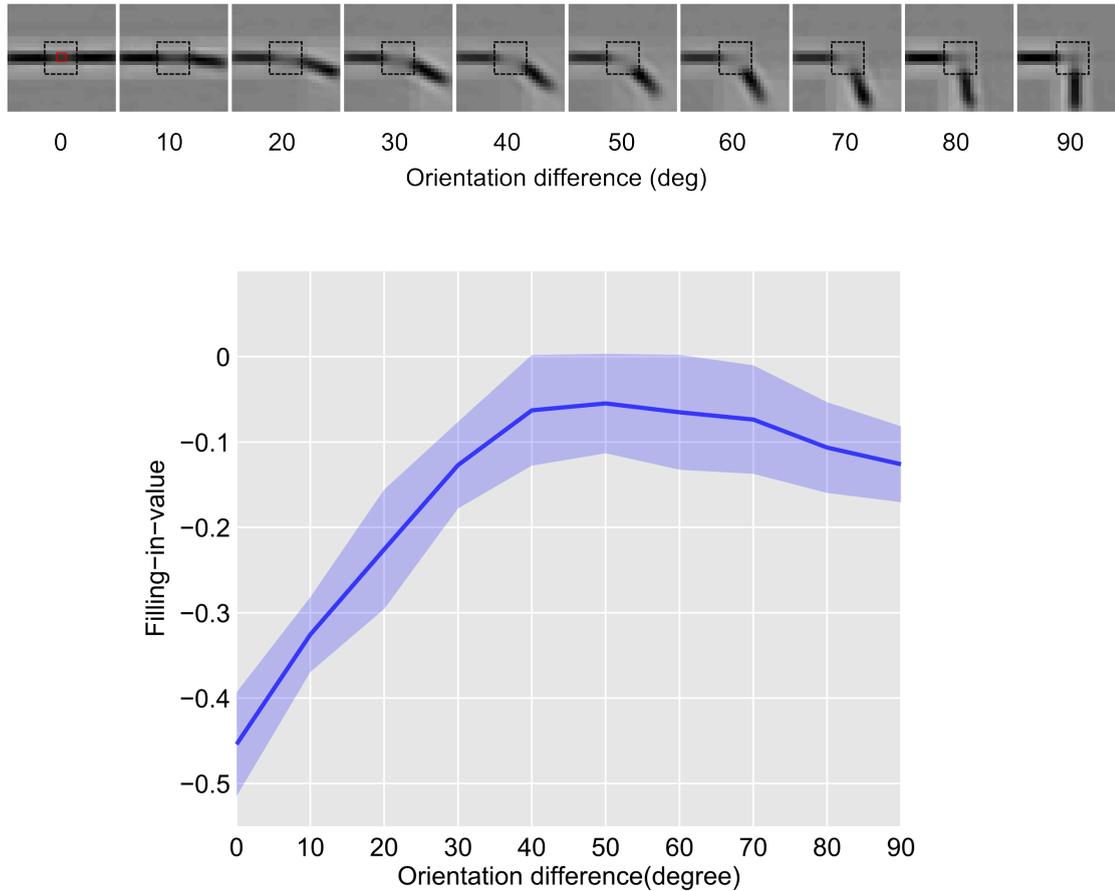


FIGURE 4.4: *Tolerance in disorientation* a) Stimulus used for the study is shown. One part of the stimulus, the horizontal black bar, was fixed. The other part was rotated clockwise from its initial horizontal position (dotted bar) to the vertical position in steps of 10 degrees. The central rectangle is BS and the outer rectangle (dotted) is the receptive field. Rectangular regions are shown for illustration only, b) Perceptual images, generated from the response of PE neurons at level 1, corresponding to all ten orientations of stimulus are shown, c) Average response at the center (2×2) of the BS region in (b) is plotted as a function of orientation angle.

4.3.2.1 Anisotropy in orientation selectivity

To investigate the presence of any anisotropy, we measured the orientation tuning distribution of the trained neurons in V1. To do this, we utilized bar stimulus of different orientation and frequency and determined the orientation tuning of a particular neuron by registering their optimal response. Fig. 4.1b shows the distribution of orientation tuning of neurons in V1. It is evident from the distribution that greater number of neurons are oriented towards the horizontal orientation, followed by vertical and the non-cardinal orientation. This anisotropic distribution is very much in-line with

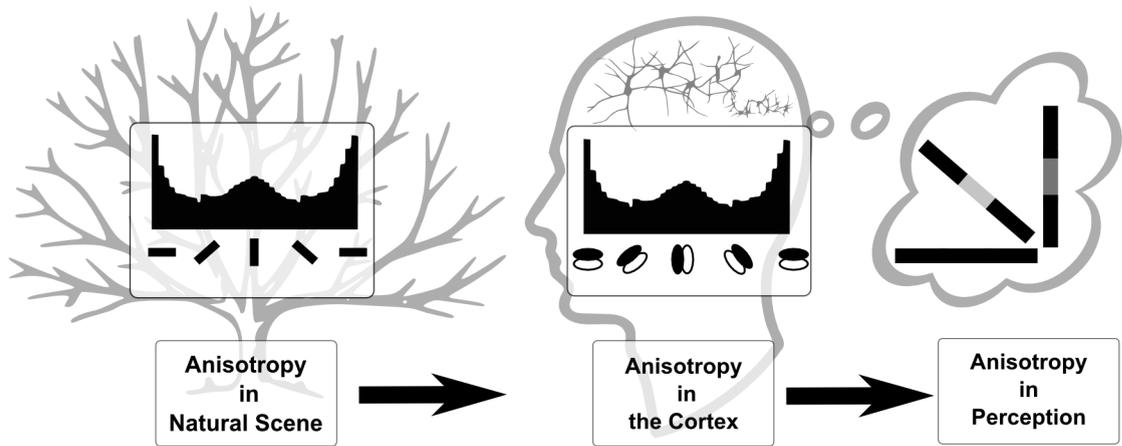


FIGURE 4.5: *Anisotropy in the natural scene, the cortex, and perception: Our aim is schematically presented in this diagram. We want to establish the link between the anisotropy in the contours in the natural scene, orientation preference of neurons in the cortex and orientation bias in human perception*

the reported anisotropy of orientation distribution in natural scenes [15, 16, 97] and orientation tuning distribution of neurons in primary visual cortex [17, 96, 99–101].

4.3.2.2 Anisotropy in filling-in

To investigate the anisotropy in filling-in, the learned network was exposed to a pair of expanding bar segments, as shown in Fig. 4.2c, oriented in the horizontal direction. One end of both bars was fixed and other ends were free to expand together in sync as described in the Fig. 4.2c. The network was also stimulated with similar stimuli but oriented in the vertical direction (not shown). The responses of PE neurons were recorded as a function of bar extension (length) for both orientation configurations.

Fig. 4.6b shows the plot of ‘filling-in-value’ as a function of the bar extension for both configurations. Inspection of Fig. 4.6b shows that the filling-in starts improving when the lengths of the bar segments exceed a certain minimum. This can be visualized from the perceptual images (Fig. 4.6a) where beyond a certain minimum length, the bars appear continuous. This result is similar to the ‘minimum-length requirement’ [7]

of filling-in. Moreover, the comparative investigation of results for horizontal and vertical arrangements indicates that the minimum critical length needed for the onset of filling-in is lesser for the horizontal orientation. In addition, for the same length, the filling-in performance is better (more negative ‘filling-in-value’) for the horizontal case. This anisotropic property is completely in agreement with psychophysical studies [7].

To validate our results, a two-way ANOVA was conducted that examined the significance of effect of degree of bar extension and the configuration (horizontal /vertical) on the filling-in-values. We found that the effect of extension [$F(10,858) = 933.93$, $p=0$], configuration [$F(1,858) = 585$, $p=0$], and, the interaction between them [$F(10,858) = 24.09$, $p=0$] was significant.

4.3.2.3 Anisotropy in tolerance of filling-in

We repeated the tolerance investigations (described in the section 4.3.1) for the vertical configuration. In misalignment case, the other bar (above the blind spot) was shifted horizontally by one pixel per instant. While in the disoriented case, the other bar was rotated by 10 degrees per instant up to 90 degrees.

The response of PE neurons in BS module was recorded with changing misalignment and disorientation and the perceptually equivalent images were generated from these responses, which are shown in Fig. 4.7a (bottom row) for the misalignment case and Fig. 4.8a (bottom row) for the disorientation case. Inspection of the plot of ‘filling-in-value’ in the Fig. 4.7b and Fig. 4.8b indicate that the general nature of filling-in: filling-in is best in the case of perfect alignment but deteriorates with increasing misalignment or disorientations, remains similar for the vertical configuration. However, the two configurations exhibit the different degree of filling-in for the different misalignment/disorientation.

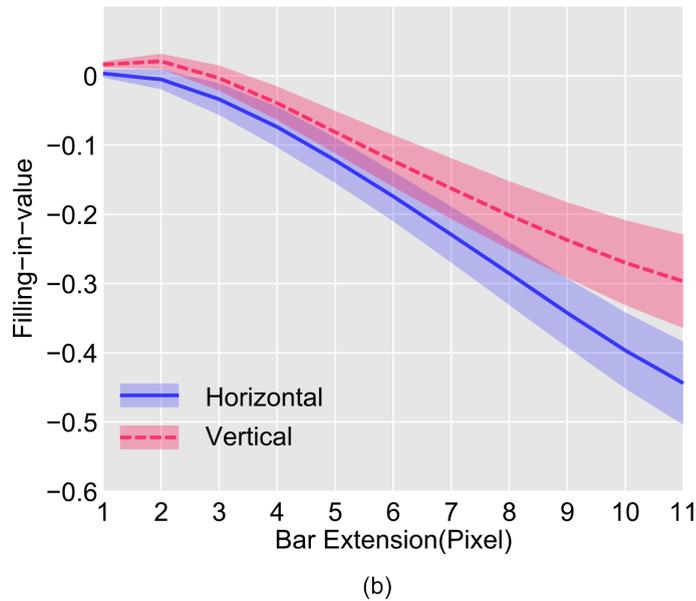
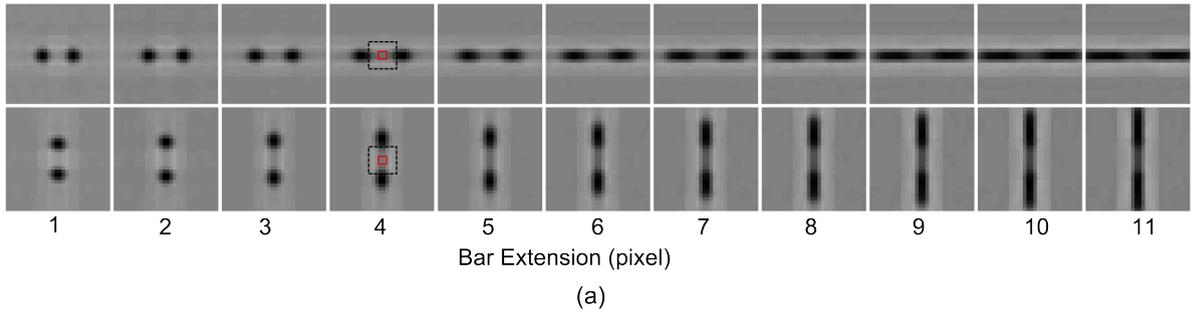


FIGURE 4.6: *Filling-in anisotropy. (a) Perceptually equivalent images are shown, which are generated from the response of PE neurons while the network was stimulated with stimuli depicted in Fig. 2a. (b) The plot of ‘filling-in-value’ in BS area of the images in (a) as a function of bar extension measured from the edge of the blind spot. The lines represent the average and the shaded portion indicates the standard deviation for the 40 training set.*

In misalignment case, comparison of ‘filling-in-value’, plotted in Fig. 4.7b, show that it is more negative (better filling-in) for the horizontal orientation compared to that of the vertical orientation. This is a signature of anisotropy of filling-in (discussed in the previous subsection). Moreover, it is also evident that the average slope of the curves is different and it is higher for the horizontal case. This indicates that the rate of change of the ‘filling-in-value’, for the horizontal orientation, is more sensitive to the change in misalignment. In other words, filling-in, in the case of vertical orientation, is more tolerant to misalignment compared to that of the horizontal orientation. This could be considered as a signature of anisotropy of tolerance of filling-in.

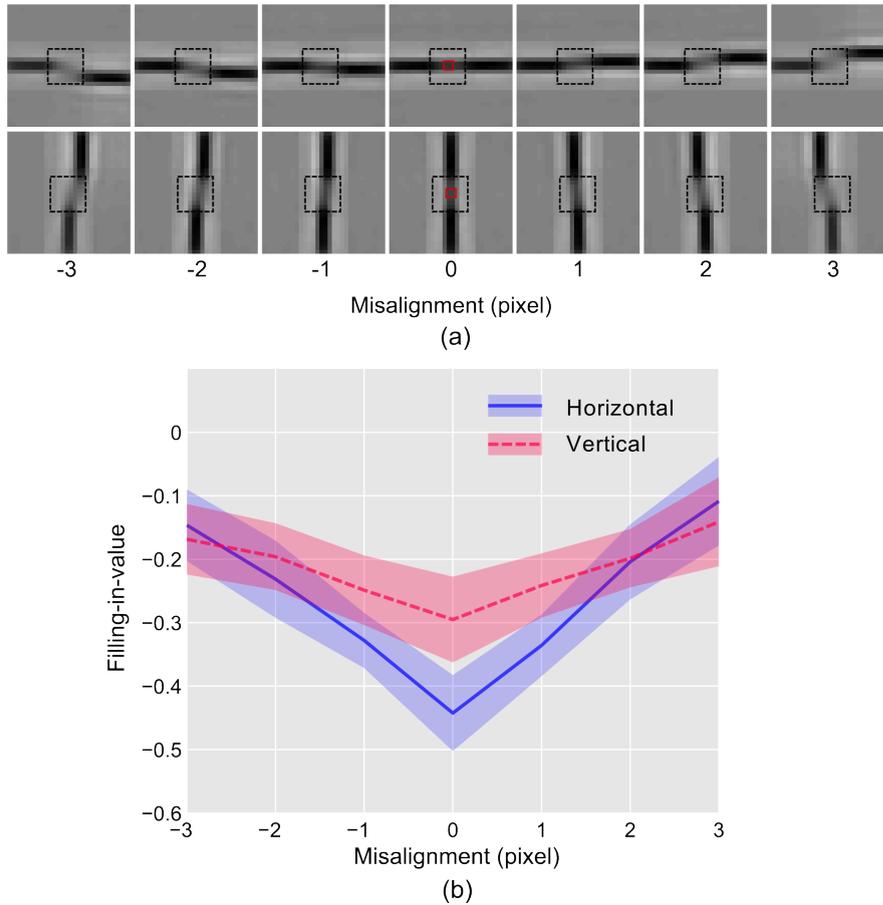


FIGURE 4.7: *Anisotropy in tolerance of filling-in of misaligned bars. (a) Perceptually equivalent images are shown, which are generated from the response of PE neurons while the network was stimulated with stimuli depicted in Fig. 4.2b. (b) The plot of ‘filling-in-value’ in BS area of the images in (a) as a function of misalignment between the bars. Convention for lines and the shades are as described in Fig. 4.3b.*

In disorientation case (Fig. 4.8b), the comparison shows that for the vertical configuration the ‘filling-in’ value remains less negative (indicating relatively inferior filling-in), throughout the entire range of difference from 0 degrees to 60 degrees (thereafter the difference becomes indistinguishable). However, its rate of deterioration remains lower in comparison to the horizontal configuration. These results indicate that the horizontal configuration favors filling-in but exhibit greater sensitivity to the changes in orientation difference (less tolerant); on the other hand, the vertical configuration is little less favorable for filling-in but less sensitive to the changes in orientation difference (more tolerant). These qualitative signatures of anisotropic sensitivity support the results of psychophysical studies [6].

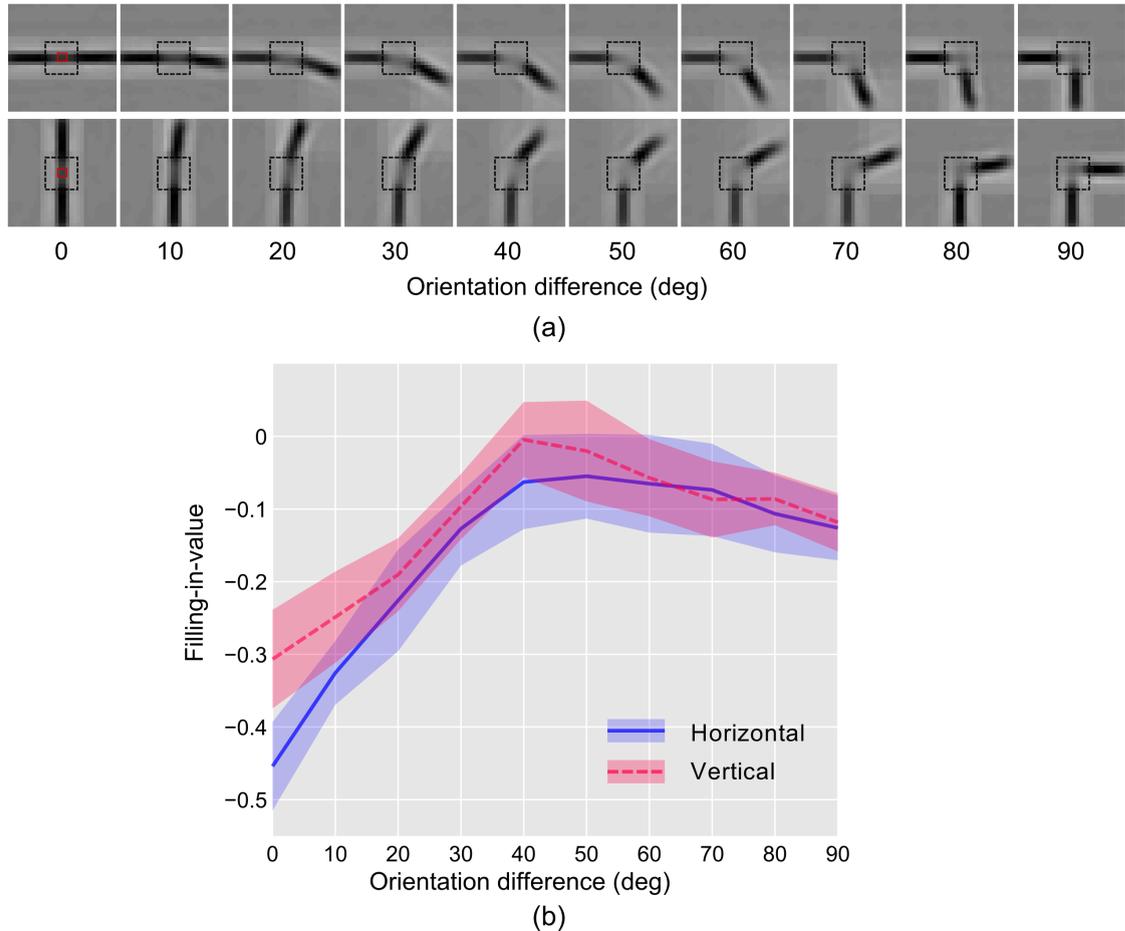


FIGURE 4.8: *Anisotropy in tolerance of filling-in of disoriented bars. (a) Perceptually equivalent images are shown, which are generated from the response of PE neurons while the network was stimulated with stimuli depicted in Fig. 4.2c. (b) The plot of ‘filling-in-value’ in BS area of the images in (a) as a function of orientation difference between the bars. Convention for lines and the shades are as described in Fig. 4.3b.*

A two-way ANOVA was conducted that examined the significance of effect of degree of bar misalignment/disorientation and the configuration on the filling-in-values. We found that the effect of misalignment [$F(6,546) = 175.91, p < 0.001$], configuration [$F(1,546) = 81.96, p < 0.001$], and, the interaction between them [$F(6,546) = 26.53, p < 0.001$] was significant. The effect of disorientation [$F(9,780) = 334.4, p < 0.001$], configuration [$F(1,780) = 104.66, p < 0.001$], and, the interaction between them [$F(9,780) = 13.12, p < 0.001$] was also very significant.

4.3.2.4 Comparison with the psychophysical results

For the purpose of direct comparison with psychophysical results [6, 7, 14], we have redrawn our results (Fig. 4.6b, Fig. 4.7b, and Fig. 4.8b) in Fig 4.9 taking into account the concept of visual angle (VA) and a threshold. In our study, the extent of the blind spot is 8×8 pixels. On the contrary, if we approximate the blind spot to be a square region, the average size of the spot is 5×5 degree [3, 4]. Therefore, we have used a scaling factor of .625 for converting pixels to degrees. Additionally, we have introduced an artificial threshold (at 50 %), which is used to obtain quantitative estimates. Bar diagrams compatible for comparison with psychophysical experiments are plotted on the right of each of the plots.

In psychophysical investigations, the anisotropy in filling-in (discussed in Fig. 4.6) was measured using staircase method [7]. On the contrary, in our study, we have measured the filling-in-value which directly corresponds to the activity of the PE neurons of our model network. Change in the activity of these neurons, for exposure to different stimuli, encodes properties of the filling-in process. As shown in Fig. 4.6b, in our model network, neurons exhibited higher activity (more negative filling-in-value) when exposed to horizontal bar stimulus compared to the activity induced by vertical bar stimulus for a given bar length. Therefore, bar length that induces similar levels of neural activity will be different for different configuration (horizontal or vertical), and this may provide an estimate of anisotropy. We have estimated these lengths ('minimum length requirement') by considering a threshold at 50 % filling-in-value corresponding to the vertical case (in red) as shown in Fig. 4.9a. The estimated bar lengths are plotted as bars in Fig. 4.9b that are similar to the results reported in experiments [7]. We did not consider a threshold corresponding to the 50 % filling-in value in the horizontal case because that would have estimated longer bars, whereas our focus is to find the minimum lengths of bars.

For the presentation of results related to the tolerance in filling-in, we have conceived a general notion of the tolerance as a rate of change of filling-in-value, with increasing difference in attributes. Faster change (higher rate) will indicate lesser tolerance, and this is advantageous because one can predict the tolerance by inspecting the slope of the curve representing the changing filling-in-value, which is directly available from the simulation study. Several psychophysical studies [6, 14], on the other hand, have defined tolerance in filling-in as the maximum difference in attribute above which filling-in is not perceivable with certainty. This definition is more compatible with the outcome of psychophysical experiments. Therefore, a different presentation of our results is necessary for direct comparison with psychophysical findings.

For comparison, we have normalized the results shown earlier in Fig. 4.7b and Fig. 4.8b. The normalization is performed by dividing the filling-in values represented by a given curve (for a specific configuration) by the magnitude of the maximum filling-in value for that specific case, and this is repeated for each of the plots separately. To compare with the psychometric plot of psychophysical results we have, in the resultant plots, considered -1 to represent 100% probability and 0 to represent 0% probability. Additionally, following the general convention where the y-axis ranges from minimum probability (at the bottom) to the maximum (at the top), we have inverted our plots which are shown in Fig. 4.9c and 4.9e. These normalized plots range from 50% probability at the bottom to the 100% probability at the top and tolerances are estimated from these plots. We have introduced an artificial threshold at the 50% probability, where it is assumed that beyond this threshold filling-in cannot be perceived with certainty.

In line with the definition of tolerance (maximum difference for which filling-in cannot be perceived with certainty) compatible with psychophysical experiments, tolerances are represented as vertical bars in Fig. 4.9d and Fig. 4.9f. Estimated lengths for misalignment shown in Fig. 4.9d corroborate the experimental findings presented in the psychophysical study [6] for the vertical as well as for horizontal configurations. The

estimated orientation difference for horizontal and vertical configurations are shown in Fig. 4.9f. In this plot, the tolerances are 19 degrees and 24 degrees for horizontal and vertical configuration respectively while the psychophysical results [6] provides these value as 40 degrees and 55 degrees. Though the magnitude of tolerances obtained in our investigation differs from the tolerances reported in psychophysical experiments, it is interesting to observe that the ratio (1.26) of vertical to horizontal tolerance is very close to the value (1.37) obtained experimentally. It shows that qualitatively our results agree well with the experimental findings. These results presented in Fig 4.9d and Fig 4.9f clearly show the vertical dominance in the case of tolerance of filling-in for misalignment and as well as for disorientation. Filling-in phenomenon for misaligned bars was reinvestigated recently [14] in the context of linear and curvilinear completions at the blind spot. In this study, perceptual completion was defined as the case where participants perceived continuous bar irrespective of the apparent shape being straight or curved. This definition is identical to the misalignment case we have considered in this article as well as the one reported earlier [6]. Though the tolerance of misalignment reported in later study [14] is higher approximately by a factor of 2 than the results reported earlier [6], the vertical dominance in filling-in is preserved and the ratio of vertical to horizontal misalignment remained very close to that reported in psychophysical study [6]. Therefore, our results related to misalignment investigation to some extent corroborates the results reported in psychophysical study [14].

The common asymptotic shape, as observed in psychometric plots, near 100 % probability [14], is not apparent in our plots. The most plausible reason is the lower resolution we have achieved in our simulation, where we have considered a 8×8 pixels wide blind spot that provided 4 data points corresponding to four misalignments. For the same limited resolution, the results of collinear experiments [14] could not be discussed with the results of our investigation. The collinear filling-in has been shown for the very narrow misalignment which is not possible to investigate in the

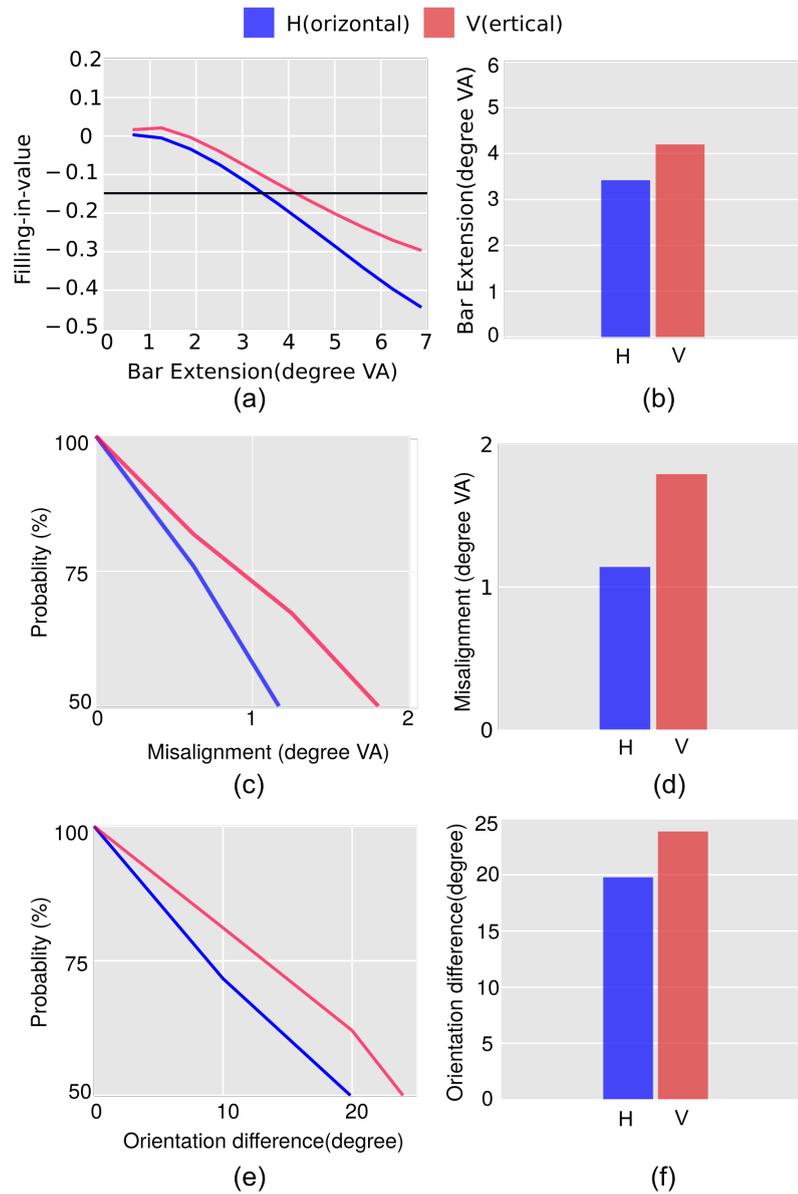


FIGURE 4.9: results of Fig 4b, 5b, and 6b are redrawn in (a), (c), and (e) respectively for the comparison. The visual angle is represented as VA in the plots. (a) The horizontal line represents the threshold corresponding to the 50 % of the maximum filling-in-value for vertical configuration, and the estimated bar lengths corresponding to this threshold are plotted as bars in (b). (c) Normalized plots, as explained in the text, for the positive misalignments are presented, which continued from the 100% to 50% probability (artificial threshold). The amount of misalignment at this threshold for the horizontal and vertical cases are shown as bars in (d). (e) Similar normalized plots for orientation difference are shown here and estimated orientation difference at 50% threshold is shown as bar plots in (f)

current context. However, a model network with a better resolution could be able to shed some light on these phenomena.

4.3.2.5 Relation between natural image statistics and filling-in at the blind spot

How anisotropy, then, arises from the response of the model network? We have shown (Fig. 4.1b) that, in agreement with natural scene statistics, the distribution of the orientation preference of the learned receptive fields at V1 reflects the over-representation of neurons tuned towards horizontal orientation. This demonstrates that the model network could encode the anisotropies of natural scene statistics through learning.

As discussed in the last chapter, the likelihood of filling-in of features (bars with different attributes) is guided by its likelihood of occurrence in the natural scene. Features that are more frequent tend to be more likely candidates for filling-in. In this perspective, we argue that the over-representation encoded by the learned receptive fields at V1 dominates the prediction at the blind spot that leads to filling-in of discontinuity. This happens because in the absence of the feed-forward connections (in the network representing blind spot region) top-down predictions biased by the learned internal model dominates. Thus, the prevalence of horizontally oriented features (lines, bar etc.) in the learned internal model results in the superiority of horizontal features in filling-in. This is reflected as more negative ‘filling-in-value’ in all three horizontal cases (blue line) in Fig. 4.6, Fig. 4.7, and Fig. 4.8.

How vertical superiority arises in tolerance of filling-in? The nature of variation in filling-in-value, shown in Fig. 4.7b (or Fig. 4.9a) and Fig. 4.8b (or 4.9b), can be explained by taking into account the orientation tuning distribution of neurons shown in Fig. 4.1b. Inspection of Fig. 4.1b reveals that neurons tuned toward horizontal orientation have a higher population and sharper distribution. In comparison, neurons

tuned toward vertical orientation have a lower population and relatively broader distribution. The sharper distribution (and higher population) of neurons tuned toward horizontal orientation results in a more specific estimate for filling-in that would be less tolerance despite the fact that better filling-in will be observed for that orientation. On the other hand, broader distribution (and lower population) of neurons tuned toward vertical orientation results in higher tolerance and the lesser response results from the comparatively lower population. Therefore, in the case of horizontally oriented stimuli, the filling-in performance deteriorates at a faster rate with increasing difference in stimulus attributes compared to that of vertically oriented one.

These arguments can be readily put forward for explaining the anisotropy in tolerance of filling-in for disoriented bar stimuli (Fig. 4.8). For a given configuration (horizontal or vertical), the rotating segment of the stimuli makes varying angles with the fixed segment. Because of this, the filled-in section that resides inside the blind-spot will have to be aligned at varying angles either toward vertical or horizontal depending on the configuration. For every angle (0 to 90 degree), neurons having the similar orientation preference matching that of the filled-in section (in the blind spot) that connects the pair of bars will be activated for filling-in. For horizontal configuration, neurons having horizontal orientation preference as well as neurons having close to horizontal orientation preference are activated (depending on the stimuli in Fig. 4.8b). Because of the sharper distribution of neurons with orientation preference toward horizontal, a smaller orientation difference (with the horizontal) of the rotating bar will activate a certain population of neurons with similar orientation sensitivity but this population will be comparatively much smaller compared to the population that have been de-activated due to the increase in orientation difference. This will result in a larger decrease in response of the neurons, which is reflected as a faster decrease (lesser tolerance) in responses with increasing stimulus deviation from the horizontal orientation. Similar arguments can be given to explain the slower decrease (greater

tolerance) in responses of neurons (because of broader distribution) in the case of vertical configuration.

In the case of misaligned bar investigation (Fig. 4.7), one bar is kept fixed and the other is shifted (either vertically or horizontally) to simulate the varying amount of misalignment. Because of this, the filled-in section of the pair of bars (inside the blind-spot) will have to be aligned at varying angles either toward vertical or horizontal depending on the configuration. For every misalignment, neurons having orientation preference similar to that of the filled-in section become activated for filling-in. Therefore, as discussed before, the filling-in-value will be determined by the population of neurons tuned to a specific orientation and the nature of variation (with increasing misalignment) will be determined by the width of the distribution of neurons. This is reflected as better filling-in (more –ve filling-in-value) and faster deterioration in filling-in with increasing difference in attributes in case of the horizontal configuration shown in Fig. 4.7b.

From the preceding discussions, it is evident that the predominance of horizontal contours in natural scene results in better filling-in operation in all three cases considered. This is reflected as more –ve filling-in-value as shown in Fig. 4.6, Fig. 4.7, and Fig. 4.8(in blue). On the other hand, broader distribution of vertical contours results in a more tolerant response in filling-in operation with increasing difference in attributes. This is reflected in the curves (in red) with shallower gradient depicting the changing filling-in-value in Fig. 4.7 and 4.8.

Does the model HPC network predicts filling-in-values in accordance with statistics of natural images it was trained with? To validate these conclusions, we have repeated investigations with misaligned bar stimuli (Fig.4.7) on a network trained on a natural image and its 90-degree rotated version having vertical orientation superiority with asymmetric distribution of contours, which is shown in Fig. 4.10a. The distribution of orientation content of the upper-left image is shown at the bottom of Fig. 4.10a.

We have evaluated the orientation at each pixel (upper left image in Fig. 4.10a) from the direction of the local gradient (of the grayscale image). This was evaluated from the arc tangent of partial derivative (in the kernel) in the vertical direction divided by the value in the horizontal direction.

The distribution reveals the dominance of vertical contours and an asymmetric distribution around the dominant orientation (90 degrees) with a sharper rise (left side) and a slower fall (right side). Training with these two images produced an orientation preference of V1 neurons as shown in Fig. 4.10c, where the neurons are equally sensitive to cardinal orientations and possessed similar distributions around cardinal orientations, which nearly preserved the asymmetries of the original image (Fig. 4.10a). This resulted in an equal filling-in response as shown by the superimposed curves (representing filling-in-values) in Fig. 4.10d. Despite the fact that the distributions are similar, close inspection of Fig. 4.10c reveals that the distributions, centered around cardinal angles, are asymmetric exhibiting a sharper rise at the left side and a comparatively slower fall at the right side. This implies that as long as the moving bar was aligned at $180 - \theta$ ($90 - \theta$) (Fig. 4.10b), the filling-in value altered at a faster rate with the angle and when it was aligned at $180 + \theta$ ($90 + \theta$), the filling-in value altered at a comparatively slower rate. This is reflected in the plot shown in Fig. 4.10d as faster rise on the left and a slower rise on the right side. From these results we conclude that the filling-in-value predicted by the model HPC network is in accordance with the statistics of images used for training, where the absence of anisotropy in the dominance of the contours tuned to cardinal orientations results in equal filling-in response; and similar distribution of cardinal orientations results in similar gradient in the changing filling-in-value with increasing difference in the attributes.

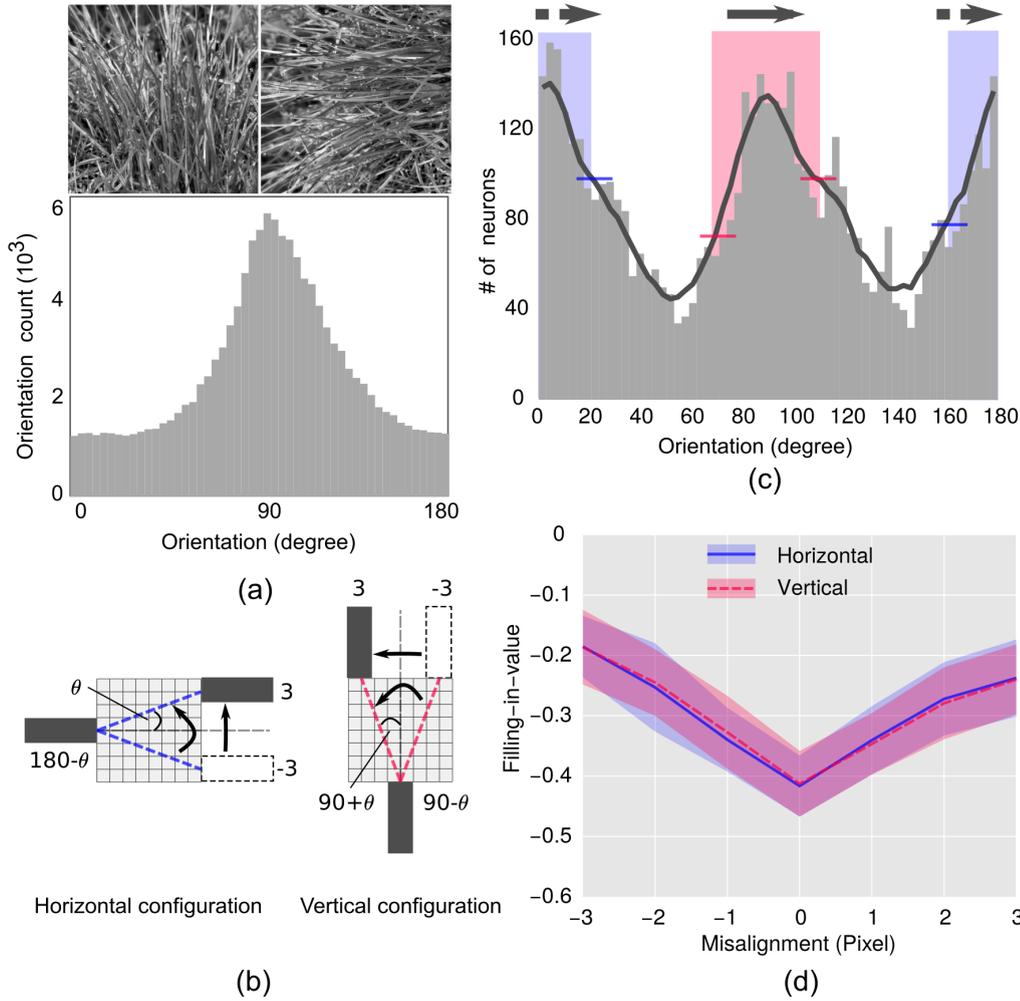


FIGURE 4.10: *Validation investigation.* (a) *Natural images with asymmetric orientation distribution.* The upper-left image mainly possesses contours with a bias towards vertical orientation. The histogram exhibiting this property is shown below. The upper-right image is 90 degrees rotated version of left one (histogram is not shown). (b) *A detailed schematic of the misaligned bar study conducted in horizontal and vertical configuration.* The moving bar was shifted by a maximum amount of 3 pixels on both sides of the mean (aligned) position. For the horizontal configuration it moved upward from the bottom and for the vertical case, it moved leftwards. The angular deviation of the filled-in portion (represented by dotted line inside the BS) can be evaluated from $\theta = \tan^{-1}(\text{position of the moving bar in pixels}/8)$ (the size of BS area is in pixels). (c) *Orientation distribution of trained neurons at level 1.* The continuous line (black) plot is the envelope of the histogram, which was obtained by convoluting the histogram, averaging over 7 bins. The shaded regions around horizontal (in blue) and vertical (in red) orientation indicate the population of neurons that is likely to be activated for filling-in when the moving bar is displaced by an amount degree ($\theta = \tan^{-1}(3/8)$ 20 degrees) around the mean position. The height difference between red lines (blue lines) across this smoothed plot is to indicate the neuronal density difference for the maximum misalignment (20 degrees) around the vertical orientation (horizontal orientation). The arrows above the shadowed regions indicate the direction of the moving bar. (d) *Plots of ‘filling-in-value’ as a function of misalignment obtained from the response of the network.* Convention for lines and the shades are as described in Fig. 4.3.

4.4 Discussions

In this chapter, we have studied the proposition that the properties of filling-in at the blind spot could be understood on the basis of natural image statistics in the backdrop of the predictive coding mechanism. In the first part of this investigation we studied the tolerance of filling-in while in the second part, we investigated the various anisotropies related to the filling-in. Results of our investigations with bar stimuli clearly show that the emergence of tolerance and anisotropies could be understood in terms of the orientation distribution of the features in the natural scene.

This study suggests that natural scene statistics plays a significant role in determining the anisotropy in perceptual filling-in including the anisotropy in tolerance of perceptual filling-in at the blind spot. Over-representation of horizontal contours in natural scene biases the orientation preference of neurons in V1 and that is possibly responsible for the emergence of anisotropy, which is reflected as a horizontal preference in perceptual filling-in operation. The width of the distribution of orientation preference, on the other hand, determines the tolerance and anisotropy in tolerance of filling-in. The broader distribution of vertical contours in natural scene possibly contributes to the greater stability towards vertical orientation in perceptual filling-in operation. Similar assumptions have been made in a previous study [16] to explain the stability of information processing of cardinal orientations and its greater sensitivity to small changes during visual perception.

These results demonstrate that there is a link between the orientation anisotropy in the contours in the natural environment, orientation preference of neurons in V1 and orientation bias in the perceptual filling-in at the blind spot. Our result supports the general speculation [15, 19, 93, 100, 101] that orientation anisotropy in natural scene plays a significant role in determining the anisotropy in the cortex as well as the anisotropy in perceptual orientation preference.

Firstly, we show that the model HPC network, which mimics the prediction-correction computational paradigm of the cortex, is capable of building an internal model of the outside environment by learning the statistics of natural scenes it is exposed to. This is reflected by the fact that the orientation preference, as well as the distribution of orientation preference of model neurons in V1, is very similar to the predominance of horizontal contours and their distribution in the natural environment. The plausibility of this paradigm can be established with the help of several previous findings. In a recent survey [18], in the physiological domain, involving cells in the cat's striate cortex indicate the preferential bias of cells towards horizontal orientation. Imaging studies also revealed [17, 96, 102] the preference of higher percentage of the area of the exposed visual cortex towards horizontal orientation compared to vertical. Innate specification along with prolonged exposure to an anisotropic environment during development is believed to be responsible for the emergence of over-representation of horizontal orientation preference of these neurons. In the psychophysical domain, correspondence between the horizontal bias in human visual processing and the anisotropy in the natural scene has been reported in [19, 98]. A detailed survey in this work also shows the prevalence of horizontal contours in a typical natural scene compared to vertical contours. In a recent study, it has been demonstrated that visual orientation perception reflects the knowledge of environmental statistics [93]. In this work, the estimated internal model of human observers was found to match the orientation distribution measured in photographs of environment though the difference between horizontal and vertical was not addressed.

Secondly, our investigations reveal that the anisotropy in orientation preference (horizontal) of V1 neurons results in the similar anisotropy in the filling-in performance and the distribution (sharper or broader) of cardinal neurons results in the anisotropy of tolerance in filling-in performance. What is the biological plausibility of such a scheme? In an imaging study [16] it has been shown that in V1 the distribution of inputs to the cardinal neurons is narrower compared to those of oblique neurons.

When exposed to selective perturbation induced by adaptation (oriented away from the neuron's preferred orientation), cardinal neurons exhibited greater stability compared to the neurons tuned to oblique orientation. This is attributed to the fact that because of the narrower distribution of local inputs to the cardinal neurons, an adaptive stimulus would stimulate a fewer number of neurons in the vicinity compared to that of the neurons tuned to oblique orientation. This demonstrates that the width of the distribution (of neurons) plays a significant role in determining the responses when stimulated away from the preferred orientation. From a different perspective it indicates that for neurons having narrower distribution, a much greater change in response will be observed with increasing deviation of the stimulus orientation from the neuron's preferred orientation. This implies greater sensitivity and therefore, lesser stability in the present context. Comparatively, neurons having broader distribution will be less sensitive (more stable). This is similar to the findings of our observation. Evidence in favor of larger neural population preferring horizontal orientation (compared to vertical) have also been found in several studies [17, 18, 102], as discussed earlier.

*Truth are illusions whose 'illusioness' is
overlooked ...*

Friedrich Nietzsche

5

General Discussion

5.1 Summary

In this thesis, we have investigated the computational mechanism of filling-in at the blind spot. We postulated that this could be understood by examining the hierarchical predictive processing inside the visual cortex. Moreover, we speculated that the properties of filling-in could be determined by the properties of natural image statistic. We conducted several simulation studies related to the filling-in of bar stimuli in different configurations on a three level HPC model network. The

model network was trained with natural image patches and thereafter a blind spot was emulated in the network by removing the bottom-up feed-forward connection.

In the first part of this thesis, to investigate the computational mechanism of filling-in, we studied the shifting bar filling-in phenomenon. We recorded the response of PE neurons at V1 in BS module while shifting a long bar across the blind spot, and additionally, estimated the corresponding perceptual counterpart using the generative model. We found that the PE neurons in the BS area of model network exhibit the non-linear response similar to the response profile reported in the recent physiological studies [4]. Moreover, we also found that the generated perceptual image corresponding to this response profile represents the course of filling-in reported in psychophysical studies.

In the second part of this thesis, we investigated the role of natural image statistics in determining the properties of the filling-in process. We conducted simulation investigating *tolerance* and *anisotropy* in filling-in (reported in recent psychophysical studies) using bar-pair stimuli of different arrangements (misaligned/disoriented bar-pair) and different configurations (horizontal/vertical bar-pair).

In tolerance investigation, we used bar-pair stimuli with varying in misalignment and orientation across the blind spot. Results derived from the activity of the model network, in response to these stimuli, show that the best filling-in occurs for the completely aligned bars and it continues to occur (though in reduced strength) for a small degree of misalignment but cease to happen after a certain degree of misalignment.

In the investigation related to the anisotropy in the filling-in, we repeated the tolerance investigation for the horizontal as well as vertical bar-pair to study the anisotropy associated to filling-in-tolerance. Moreover, we used expanding bar stimuli to study the anisotropy in filling-in. We found that the model network exhibit better filling-in for horizontal (in compare to vertical) bar-pair for aligned arrangement. However, it exhibits lesser tolerance when it comes to misalignment or disorientation of bar pair.

These results are also in good agreement with the psychophysical results reported earlier [4,6]. We discussed this results in the backdrop of the orientation distribution of the features in the natural images and the distribution of learned receptive field in our model network, which shows prevalence for the horizontal followed by the vertical and then oblique orientation.

5.2 Discussion

This thesis proposes that the general computational principle of hierarchical predictive coding could account for filling-in at the blind spot. Moreover, this thesis also proposes that the properties of filling-in could be the reflection of the properties of natural image statistics.

For an input stimulus around the blind spot, higher areas (V2) generates unified estimate (including the estimate corresponding to blind spot region) of the input stimuli on the basis of the learned statistical regularities of natural images. This estimate remains uncorrected due to the absence of error carrying feed-forward connection in BS region at V1 and therefore, local optimum-estimate is achieved essentially by top-down prediction. Influenced by learned statistical regularities, higher areas predicts a long continuous bar across the blind spot and this results in the perception of completion. The nonlinearity observed in the responses of PE neuron and, hence, the properties of filling-in, result from the degree of similarity between statistics of stimuli around the blind spot and the natural image statistics. Our results support the general suggestions of predictive coding as a general computational principle of visual cortex.

The anisotropy corresponding to the blind spot follows from the anisotropy in the orientation tuning distribution of V1 neurons which was developed via learning in the natural environment. The shape of the orientation distribution (of features)

determines the tolerance in filling-in, where broader (sharper) distribution contributes to more (less) tolerant response and these results in the anisotropy of tolerance of filling-in. The sharper distribution (and higher population) of neurons tuned toward horizontal orientation results into better estimate (hence better filling-in) but more specificity about the estimate they generate and therefore less tolerant to deviations.

In some related studies [103–105], the role of cortico-cortical (V2-V1) interaction in the filling-in of illusory contours and the surfaces was suggested. Neumann [106] proposed that the filling-in of illusory contour could be the outcome of modulation mechanism of the feedback signal from V2, which enhance the favorable response profile of feature detecting neurons, mainly in the superficial layers, at V1, in the context of larger contour coded at V2. This model, therefore, has its limitation in explaining the completion across the blind spot where the activity is mainly found in the deep layer of the V1. In another recent study [107], authors tried to explain the non-linear behavior of neurons in filling-in in terms of the interaction of top-down and bottom-up signal in a multinomial Bayesian model. In this model the interaction between top-down and bottom-up signals has been assumed multiplicatively in nature, and, consequently, feed-forward signals do not represent “prediction errors” but “prediction matches”. However, recently, numerous electrophysiological and neuroimaging findings (for the review see [108, 109]) support the HPC assumption that the feed-forward connection carries the “prediction-error”. In this study, considering feed-forward neurons as error-carrying neurons, we have explained the non-linear response of PE neurons as well as the other properties of filling-in. These PE neurons, in this framework, hypothetically reside in the deep layer of the cortex [59, 110], which is consistent with the physiological findings.

The exact neural circuitry for the implementation of predictive coding is not yet clear to the researchers [109, 111]. Various implementations has been proposed [8, 59, 81, 108]. These proposition do agree on the idea of the separate population of PE neurons and error-carrying neurons, which is recently demonstrated in the physiological

study [112]. These propositions differ each other on the idea of localization of PE neurons and error-carrying neurons. In this context, the deep-layer activities of neurons corresponding to the filling-in could be regarded as a support for the location of PE neurons in the backdrop of standard HPC proposition [59, 110] and the physiological support of deep-layer as a feedback providing layer. So this thesis not only provides the support to the HPC as a general computational principle of visual cortex but also suggest the link between the location of filling-in activity and the localization of the HPC circuitry, especially the location of PE neurons.

In the previous studies [6, 14] on anisotropy in filling-in, it was speculated that there might be different anisotropic processes responsible for different kinds of anisotropy observed in different (misalignment, disorientation, and luminance difference) filling-in investigations e.g., it was speculated that the anisotropy in misalignment experiment might have arisen from the anisotropy in vernier acuity. Here in this study, we have proposed a possible alternative explanation in terms of a unified principle based on the role natural image statistics. We have demonstrated this in filling-in investigations involving misaligned and disoriented bar stimuli. Results of our studies also suggest that the anisotropy in vernier acuity might have its origin in the statistics of natural scenes. Evidence in support of these suggestions can be found in [113], where it was argued that the vernier misalignment can be discussed on the premise that the average orientation of a misaligned pair of abutting lines differs from that of the aligned lines. Vernier acuity preferring horizontal directions over the vertical including the cardinal over the oblique has been demonstrated in this work.

We speculate that the horizontal superiority [6] in the tolerance of luminance difference could be discussed in terms of statistics of the natural scene. Luminance is a surface property, and, therefore, for proper inference, the cortex should be capable of encoding 3D surface information efficiently. In a recent study [114] it has been shown that disparity neurons are capable of encoding statistics of the natural scene.

Studies [115] also show that the pair-wise functional connectivity between the disparity tuned neurons in V1 matches the anisotropic distribution of correlation between disparity signals in the natural scene. Though, these studies mainly concentrated on the cardinal vs non-cardinal aspect of the anisotropy, a close inspection of the plots indicate a broader distribution of the horizontal features. This broader distribution in disparity signal (or pair-wise connectivity) could be linked to the horizontal superiority in the tolerance of luminance difference. Some supportive evidence can be found in a recent work [116] showing that relative luminance and binocular disparity preferences are correlated in accordance with the trends of natural scene statistics. These studies suggest a possible link between the anisotropy in the disparity signal and the relative luminance. In a future work, incorporation of surface representation in the internal model in the HPC framework might explain the anisotropy in luminance difference.

So far, very few psychophysical investigation about anisotropy at the blind spot has been done. These investigations mainly focused on the filling-in of the cardinal bar-pair and do not touch light on the filling-in of a cardinal bar pair. The conclusion of our studies suggests that the, following the distribution of natural features, oblique bar pair would have inferior filling-in in compare to the horizontal and the vertical bar pair. A psychophysical study, using oblique bar pair, could verify this suggestion.

This study does not reject any possible role of intracortical interaction in V1 in filling-in completion. There could be some other (or more than one) prediction-correction pathway within V1, which can contribute to filling-in based on contextual information surrounding the blind spot.

What about the surface or texture filling-in across the blind spot? In this thesis, we mainly investigated the filling-in phenomenon related to the bar stimuli, which in turn mainly governed by the learned statistics of contrast information (edge, boundary, etc.) found in natural scenes. To inference of surface and texture, the network will,

first, need to learn these features from the scene. Some recent works suggest [114,117] that the surface information can be encoded in the activity of disparity neurons in V1. Once a representation of surface and texture is achieved in the probabilistic framework by learning, the same prediction-correction mechanism might be able to account for the filling-in process related to the surface and the texture.

Whether the other filling-in phenomena could be understood in the common framework of the predictive coding? In this thesis, we investigated the filling-in phenomena at the blind spot under the general computational mechanism of predictive coding. As we have discussed the absence of bottom-up feed-forward input the top-down prediction plays an important role in the process of filling-in. This condition may hold for the other filling-in related to the deficit of the input (i.e natural or artificial scotoma) and are expected to be explained in this framework. Other than that the steady prolonged and stabilized retinal images might cause a temporary hindrance for the feed-forward connection and ultimately fit into the condition for filling-in we have discussed above. Then, what about the filling-in related to the illusions (i.e Kanizsa shapes)? where apparently we can't find any absence of the feed-forward input by any means. In such cases, a strong prediction signal might take over the feed-forward signal [118]. These speculations could be verified with a detailed and sophisticated HPC model. Some recent advancement [81,82,119] has been made over the standard predictive coding [8] to accommodate such sophistications. Future studies with such extended model with the larger number of neurons, accommodating non-linearity, surface representation, intra-cortical interaction, color etc. might provide the basis for understanding the filling-in as a whole.

5.3 Conclusion

We demonstrate that the filling-in at the blind spot is the outcome of the prediction-correction mechanism of the cortex. In the absence of feed-forward connection corresponding to the blind spot the prediction made by the higher area dominates the filling-in process and the nature of filling-in is guided by the statistical regularities of the natural scene. Our results also suggest that the over-representation of horizontal contours in natural scene biases the orientation preference of neurons in V1 and that is responsible for the emergence of horizontal preference in perceptual filling-in. The width of the distribution of orientation preference, on the other hand, determines the tolerance and anisotropy in tolerance of filling-in, where the broader distribution of vertical contours in natural scene contributes to the greater stability towards vertical orientation in perceptual filling-in operation.

In short, the results in this thesis suggest that the process of filling-in completion at the blind spot could be a manifestation of a hierarchical predictive coding principle where filling-in performance and associated properties are shaped by the natural scene statistics. Moreover, this work offers new insights into the role of natural scene statistics and suggests what is possibly the first systematic bridge linking anisotropy in three levels: natural environment, visual cortex, and perceptual filling-in at the blind spot.

“Look deep into nature, and then you will understand everything better.” —Albert Einstein



Coding Length of HPC

The coding length can be derived from multivariate Bayesian theorem. For an input signal \mathbf{I} the task is to find the probability of the estimate \mathbf{r} , context driven prediction signal \mathbf{r}^{td} and the associated parameters U . The posterior probability is therefore (after eliminating the normalizing factor $P(I)$ in denominator)-

$$\begin{aligned} P(\mathbf{r}, \mathbf{r}^{td}, U | \mathbf{I}) &\propto P(\mathbf{I} | \mathbf{r}, \mathbf{r}^{td}, U) P(\mathbf{r}, \mathbf{r}^{td}, U) \\ &\propto P(\mathbf{I} | \mathbf{r}, U, \mathbf{r}^{td}) P(\mathbf{r}^{td} | \mathbf{r}, U) P(\mathbf{r} | U) P(U) \end{aligned} \tag{A.1}$$

In a strict hierarchy, \mathbf{I} is independent of \mathbf{r}^{td} . This results $P(\mathbf{I}|\mathbf{r}, \mathbf{r}^{td}, U) = P(\mathbf{I}|\mathbf{r}, U)$. Likewise, $P(\mathbf{r}^{td}|\mathbf{r}, U) = P(\mathbf{r}^{td}|\mathbf{r})$. The statistical independence of \mathbf{r} and U results $P(\mathbf{r}|U)$ into $P(\mathbf{r})$. Therefore the Equ. A.1 can be re-written as-

$$P(\mathbf{r}, \mathbf{r}^{td}, U|\mathbf{I}) \propto P(\mathbf{I}|\mathbf{r}, U)P(\mathbf{r}^{td}|\mathbf{r})P(\mathbf{r})P(U) \quad (\text{A.2})$$

Assuming the probability distributions, $P(\mathbf{I}|\mathbf{r}, U)$ and $P(\mathbf{r}^{td}|\mathbf{r})$ of the gaussian type-

$$P(\mathbf{I}|\mathbf{r}, U) = \exp\left(-\frac{(\mathbf{I} - U\mathbf{r})^2}{\sigma}\right) \quad (\text{A.3})$$

$$P(\mathbf{r}^{td}|\mathbf{r}) = \exp\left(-\frac{(\mathbf{r}^{td} - \mathbf{r})^2}{\sigma^{td}}\right) \quad (\text{A.4})$$

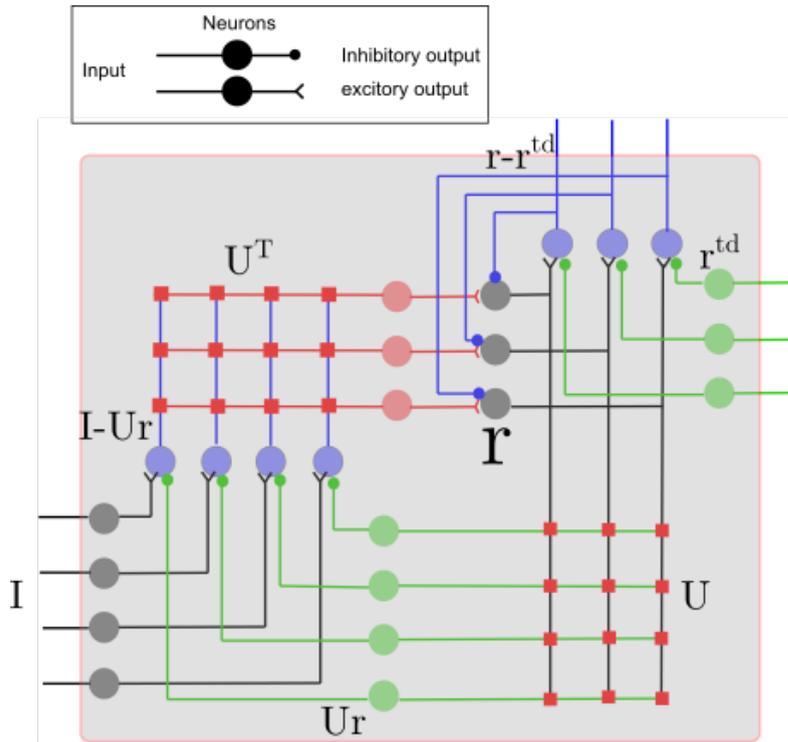
The negative of log of the Eqn.A.2 after substituting in the equations A.3 and A.4 would provide the coding length-

$$E_T = \frac{1}{\sigma^2}(\mathbf{I} - U\mathbf{r})^T(\mathbf{I} - U\mathbf{r}) + \frac{1}{\sigma_{td}^2}(\mathbf{r}^{td} - \mathbf{r})^T(\mathbf{r}^{td} - \mathbf{r}) + g(\mathbf{r}) + h(U) \quad (\text{A.5})$$

B

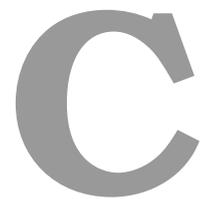
Architecture of the PE Module

Fig. B.1 represents the inner architecture of a PE module. This PE module consist of 3 PE neurons (gray colored solid circles in the upper right). On arrival of an input \mathbf{I} (lower left) the population of PE neurons make the estimate \mathbf{r} of it (also consider the feedback signal from the higher area neurons represented by green color blobs in upper right) and sends the prediction signal $U\mathbf{r}$ via feedback neurons (4 neurons in green) to the lower area. The set of 4 neurons (left blue colored blobs) detect the error $(U\mathbf{r} - \mathbf{r})$ between the input vector \mathbf{I} , with 4 element corresponding to a 2×2 image, and the prediction signal. The 3 error carrying neurons (red colored blobs) then carry the error signal to the PE neurons to update its estimate.

FIGURE B.1: *Inner architecture of a PE module*

The array of the square dot in red color (upper left or lower right) represents connection or weighting matrix. Each row of which represents the efficacy vector (or receptive field in the case of U^T) of a post-synaptic neuron. Each element (red square dot) in an efficacy vector is the connection strength with the pre-synaptic neuron.

As we can see from the above architecture, in this kind of scheme the input image is considered as a vector (not matrix or array) with value representing the gray level. So how an image, which is basically an matrix of intensity pattern (pixel value) converted to a vector? In a such scheme generally an image is coded as a column vector by unfolding the matrix row by row (or column by column). For example an image of size 2×2 is coded by a row vector of size 4×1 that would become the input of I in the Fig. B.1.



Development of Receptive Field in BS area

In this thesis, the network was trained first and thereafter the blind spot was emulated by removing the feed-forward connection in certain region. We argued that active neurons in deep layer (5/6) corresponding to filling-in has been reported to be of binocular type [3,4]. These neurons were found to respond to the inputs from both eye and hence, possess binocular receptive field. It is, therefore, natural to assume that, in normal binocular vision the feed-forward input from the non-BS eye will cause the receptive fields (of these deep layer neurons) to develop.

In this case, one can suggest that these neurons might get relatively reduced input strength in the BS region since these are getting input from only one eye rather than

from both eyes and that can lead to different weighting profile. In this context, we can argue that the strength of the relative input in BS area could be assumed to range from half to the full, depending upon the nature of integration (linear or non-linear) of the input from both eye. The simple linear summation of the input would lead to the half of strength (in BS) while the average of inputs would give the uniform strength. In between these two, one can assume a sigmoid non-linearity over the sum of inputs.

For the simple summation case, Fig. C.1 shows the learned receptive field after taking the reduced strength of the input into the account. As we can see that for the relative strength of 0.5, at level 1, the features at are similar to its surrounding though greatly reduced in strength which recovers greatly with the relative strength of 0.7. The receptive fields at level 2, on the other hand, appears to be unaffected due to these strength reduction of the input in the blind spot. Hence, we argue that the reduction of relative strength of input would not affect the generality of our approach since firstly, relative strength of 0.5 could be extreme, though it captures (up to some extent) the similar features. And secondly, the receptive fields at level 2 seems unaffected which should be obvious by considering theirs larger special extent.

From this result, we can reason that even the reduction of input, for the linear summation case, in a reasonable amount, may not give rise to any qualitative change in the learned receptive fields of the neurons because the nature of the receptive fields is mainly governed by the statistical feature of the input. We, therefore, argue that this situation may not alter the generality of our approach.

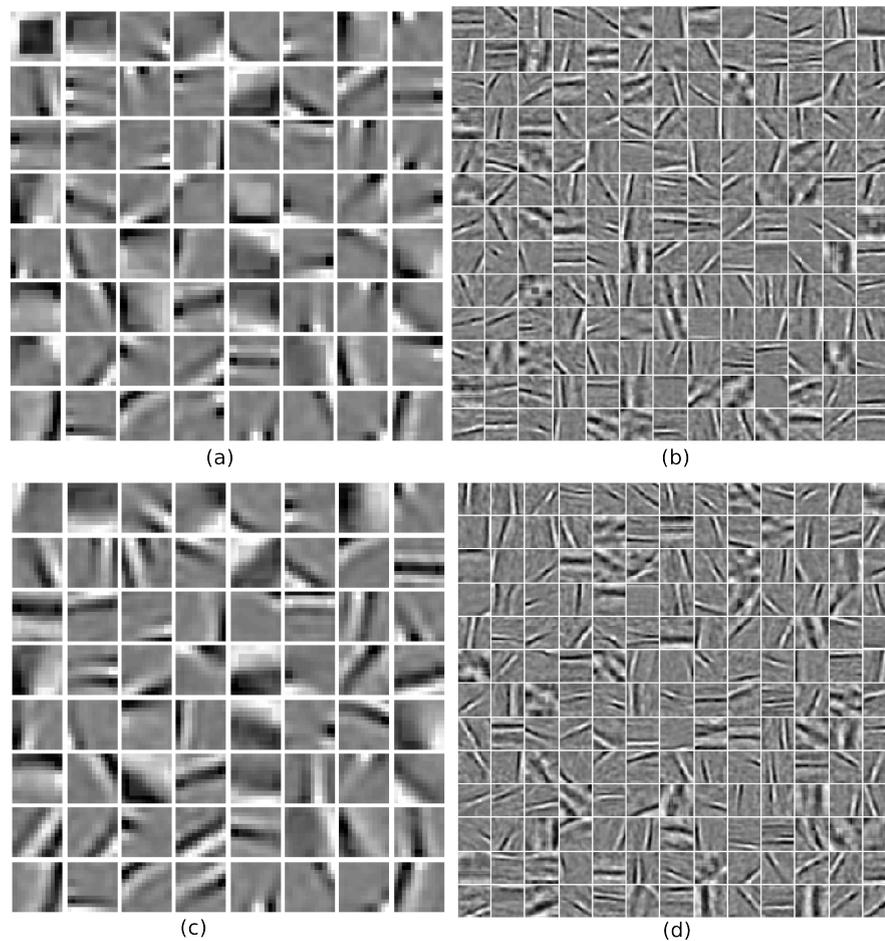


FIGURE C.1: *Learned receptive fields corresponding to different strengthen input at BS region: (a) and (b) show the learned receptive field at level 1 (of size 12×12) and level 2 (of size 30×30) after the BS region (central 8×8 pixel) were exposed to 0.5 of strength of surrounding input. Whereas, (c) and (d) show the receptive fields corresponding to 0.7 of strength of surrounding input.*

D

Filling-in of the natural scene

The natural image shown in FigD.1a is taken as an example scene. We pre-processed it (FigD.1b). From different positions of this natural scene, five image patches of 30×30 pixels were selected after preprocessing as shown in FigD.1b. These patches, shown in FigD.1b, were fed to the HPC model network for investigation.

We measured the responses of neurons at level 1 for these input image patches in both, BS and Non-BS network. The reconstructed images ('perceptual images') corresponding to the response of level 1 neurons is shown in the FigD.1c (second and third row). The images in the second row represent the reconstructed patches when

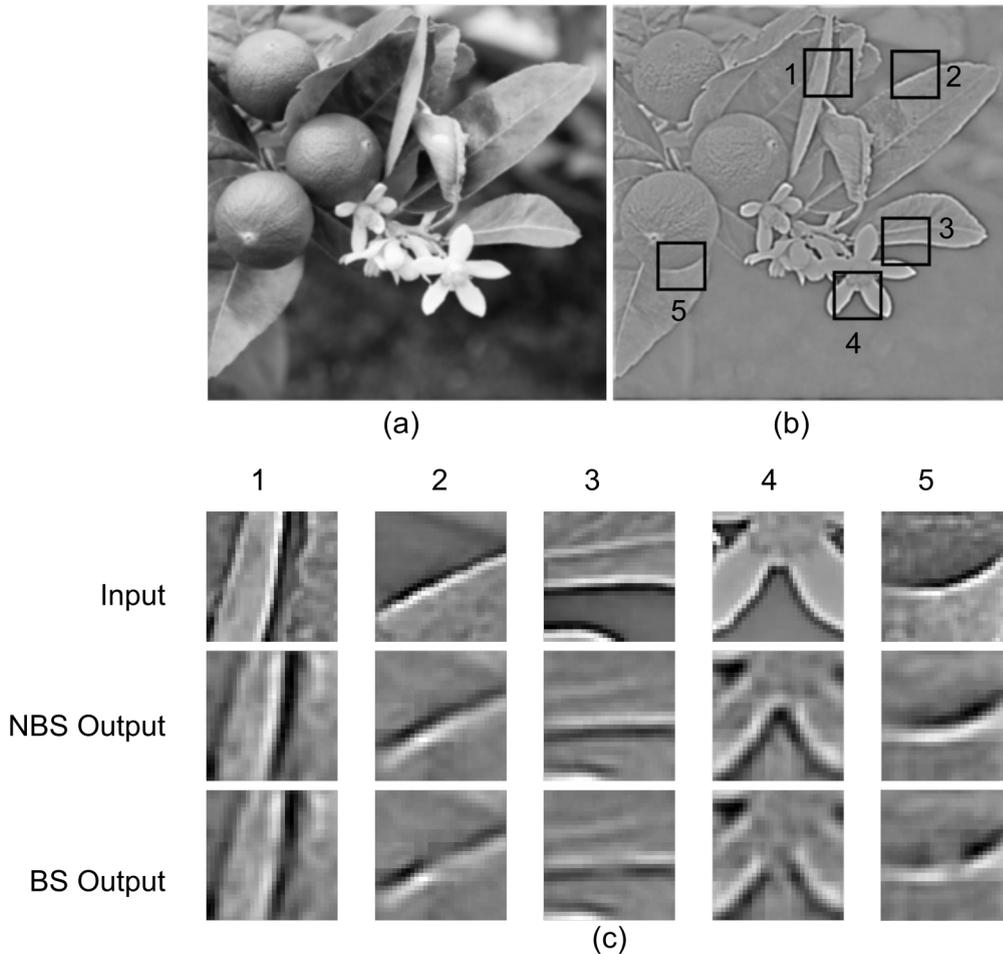


FIGURE D.1: (a) *Natural image*, (b) *Preprocessed natural image*. Black rectangles with number indicates the 30×30 pixel image patches taken from the different positions to feed into the HPC network as inputs, (c) *First row corresponds to the image patches described in (b), where different positions are indicated by column numbers. The images in the second row and the third row are the images reconstructed from the response of level 1 neurons in Non-BS network and BS network respectively.*

Hierarchical Predictive Coding network did not contain a blind spot (non-BS network). The images in the third row, on the other hand, represent the reconstructed patches when the network contained a blind spot (BS network). From the comparison, it is evident that the non-BS network was able to reconstruct the patches efficiently. On the contrary, the BS network could only reconstruct patches that contained simple shapes (e.g. straight bar, slight curved bar in column 1, 2 and 3 in FigD.1c) but failed to reconstruct patches that contained features that were relatively complex in shape (eg. angled bar, curvier bar in column 4, 5 in FigD.1c).

These results are in accordance with major psychophysical observations that the filling-in at the blind spot occurs only for the simple structures and cease to occur for the complex structures (curve and corner [120, 121]). However, there are a few alternative suggestions [90, 122] that the filling-in of curve and angle indeed occur, though extensive studies in support of this hypothesis have not been done till date. While results of this thesis support the first observation, it is not possible to make any final remark. Studies with a Hierarchical Predictive Coding Network incorporating a much more detailed model of V2 neurons may be necessary to resolve this issue because it is generally accepted that neurons at V2 codes higher order structures.

Bibliography

- [1] V. S. Ramachandran, *Blind spots*, *Scientific American* **266** (1992), no. 5 86–91.
- [2] V. S. Ramachandran, *Filling in the blind spot.*, 1992.
- [3] H. Komatsu, M. Kinoshita, and I. Murakami, *Neural responses in the retinotopic representation of the blind spot in the macaque V1 to stimuli for perceptual filling-in.*, *The Journal of Neuroscience* **20** (dec, 2000) 9310–9.
- [4] M. Matsumoto and H. Komatsu, *Neural responses in the macaque V1 to bar stimuli with various lengths presented on the blind spot*, *Journal of Neurophysiology* **93** (2005), no. 5 2374–2387.
- [5] H. Komatsu, *The neural mechanisms of perceptual filling-in*, *Nature reviews. Neuroscience* **7** (mar, 2006) 220–231.
- [6] Y. Araragi and S. Nakamizo, *Anisotropy of tolerance of perceptual completion at the blind spot*, *Vision Research* **48** (feb, 2008) 618–625.
- [7] Y. Araragi, M. Okuma, Y. Ninose, S. Nakamizo, and M. Kondo, *Anisotropy of Perceptual 'Filling In' for Horizontal and Vertical Orientation in the Blind Spot*, *Vision* **16** (2004) 1–12.

-
- [8] R. P. N. Rao and D. H. Ballard, *Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects.*, *Nature Neuroscience* **2** (1999), no. 1 79–87.
- [9] A. Clark, *Whatever next? Predictive brains, situated agents, and the future of cognitive science.*, *The Behavioral and brain sciences* **36** (jun, 2013) 181–204.
- [10] Y. Huang and R. P. N. Rao, *Predictive coding*, *Wiley Interdisciplinary Reviews: Cognitive Science* **2** (2011), no. 5 580–593.
- [11] B. A. Olshausen and D. J. Field, *Natural image statistics and efficient coding.*, *Network: Computation in Neural Systems* **7** (1996), no. 2 333–339.
- [12] B. A. Olshausen and D. J. Field, *Emergence of simple-cell receptive field properties by learning a sparse code for natural images*, *Nature* **381** (1996), no. 6583 607–609.
- [13] B. A. Olshausen and D. J. Field, *Sparse coding with an incomplete basis set: a strategy employed by V1*, *Vision Research* **37** (1997), no. 23 3311–3325.
- [14] Y. Araragi, *Anisotropies of linear and curvilinear completions at the blind spot*, *Experimental Brain Research* **212** (2011), no. 4 529–539.
- [15] D. M. Coppola, H. R. Purves, a. N. McCoy, and D. Purves, *The distribution of oriented contours in the real world.*, *Proceedings of the National Academy of Sciences of the United States of America* **95** (1998), no. 7 4002–4006.
- [16] V. Dragoi, C. M. Turcu, and M. Sur, *Stability of cortical responses and the statistics of natural scenes*, *Neuron* **32** (2001), no. 6 1181–1192.
- [17] D. M. Coppola, L. E. White, D. Fitzpatrick, and D. Purves, *Unequal representation of cardinal and oblique contours in ferret visual cortex*, *Proceedings of the National Academy of Sciences of the United States of America* **95** (mar, 1998) 2621–2623.

- [18] B. Li, M. R. Peterson, and R. D. Freeman, *Oblique Effect : A Neural Basis in the Visual Cortex Oblique Effect : A Neural Basis in the Visual Cortex*, *Journal of neurophysiology* **90** (jul, 2003) 204–217.
- [19] B. C. Hansen and E. a. Essock, *A horizontal bias in human visual processing of orientation and its correspondence to the structural components of natural scenes.*, *Journal of vision* **4** (2004), no. 12 1044–1060.
- [20] T. Gollisch and M. Meister, *Eye Smarter than Scientists Believed: Neural Computations in Circuits of the Retina*, *Neuron* **65** (2010), no. 2 150–164, [[NIHMS150003](#)].
- [21] H. Wässle, *Parallel processing in the mammalian retina*, *Nature Reviews Neuroscience* **5** (2004), no. 10 747–757.
- [22] v. S. Ramachandran and R. L. R. L. Gregory, *Perceptual filling in of artificially induced scotomas in human vision.*, *Nature* **350** (1991), no. 6320 699–702.
- [23] H. J. Gerrits and G. J. Timmerman, *The filling-in process in patients with retinal scotomata.*, *Vision research* **9** (1969), no. 3 439–442.
- [24] M. M. Gassel and D. R. Williams, *Visual function in patients with homonymous hemianopia*, *Brain* **86** (1963), no. 1 1–36.
- [25] H. S. Friedman, H. Zhou, and R. Von Der Heydt, *Color filling-in under steady fixation: Behavioral demonstration in monkeys and humans*, *Perception* **28** (1999), no. 11 1383–1395.
- [26] K. Hamburger, H. Prior, V. Sarris, and L. Spillmann, *Filling-in with colour: Different modes of surface completion*, *Vision Research* **46** (2006), no. 6-7 1129–1138.

-
- [27] L. Spillmann and A. Kurtenbach, *Dynamic noise backgrounds facilitate target fading*, *Vision Research* **32** (1992), no. 10 1941–1946.
- [28] P. De Weerd, R. Desimone, and L. G. Ungerleider, *Perceptual filling-in: A parametric study*, *Vision Research* **38** (1998), no. 18 2721–2734.
- [29] T. Cornsweet, *Visual perception*. Academic press, 2012.
- [30] S. Grossberg and D. Todorović, *Neural dynamics of 1-D and 2-D brightness perception: a unified model of classical and recent phenomena.*, vol. 43. 1988.
- [31] G. Kanizsa, *Organization in vision: Essays on Gestalt perception*. Praeger Publishers, 1979.
- [32] M. A. Goodale and A. D. Milner, *Separate visual pathways for perception and action*, *Trends in neurosciences* **15** (1992), no. 1 20–25.
- [33] E. R. Kandel, J. H. Schwartz, and T. M. Jessell, *Principles of Neural Science*, vol. 3. 2000.
- [34] D. H. Hubel, *Eye, brain, and vision*. Scientific American Library/Scientific American Books, 1995.
- [35] D. J. Felleman and D. C. Van Essen, *Distributed hierarchical processing in the primate cerebral cortex.*, *Cerebral cortex (New York, N.Y. : 1991)* **1** (1991), no. 1 1–47.
- [36] D. Van Essen and E. A. Deyoe, *Concurrent processing in the primate visual cortex*, 1995.
- [37] R. J. Douglas and K. A. Martin, *Neuronal circuits of the neocortex*, *Annual Review of Neuroscience* **27** (2004), no. 1 419–451.
- [38] D. L. Adams and J. C. Horton, *A Precise Retinotopic Map of Primate Striate Cortex Generated from the Representation of Angioscotomas*, *The Journal of Neuroscience* **23** (2003), no. 9 3771–3789.

- [39] B. A. Wandell, *Foundations of vision*. Sinauer Associates, 1995.
- [40] S. W. Kuffler, *Discharge Patterns and Functional Organization of Mammalian Retina*, *Journal of Neurophysiology* **16** (1953), no. 1 37–68.
- [41] D. H. Hubel and T. N. Wiesel, *Receptive fields of single neurones in the cat's striate cortex.*, *Journal of Physiology* **148** (1959) 574–591.
- [42] D. H. Hubel and T. N. Wiesel, *Receptive fields, binocular interaction and functional architecture in the cat's visual cortex*, *The Journal of Physiology* **160** (1962), no. 1 106–154.2.
- [43] G. C. DeAngelis, I. Ohzawa, and R. D. Freeman, *Receptive-field dynamics in the central visual pathways*, *Trends in neurosciences* **18** (1995), no. 10 451–458.
- [44] D. Marr, *Vision: A computational investigation into the human representation and processing of visual information*, .
- [45] R. W. Rodieck, *Quantitative analysis of cat retinal ganglion cell response to visual stimuli*, 1965.
- [46] J. G. Daugman, *Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters.*, *Journal of the Optical Society of America. A, Optics and image science* **2** (1985), no. 7 1160–1169.
- [47] G. C. DeAngelis, I. Ohzawa, and R. D. Freeman, *Spatiotemporal organization of simple-cell receptive fields in the cat's striate cortex. I. General characteristics and postnatal development.*, *Journal of neurophysiology* **69** (1993), no. 4 1091–1117.
- [48] J. P. Jones and L. A. Palmer, *An evaluation of the two-dimensional gabor filter model of simple receptive fields in cat striate cortex*, *Journal of neurophysiology* **58** (1987), no. 6 1233–1258.

- [49] D. L. Ringach, R. M. Shapley, and M. J. Hawken, *Orientation selectivity in macaque v1: diversity and laminar dependence*, *The Journal of neuroscience* **22** (2002), no. 13 5639–5651.
- [50] B. A. Olshausen, *20 Years of Learning About Vision: Questions Answered, Questions Unanswered, and Questions Not Yet Asked*, in *20 Years of Computational Neuroscience*, pp. 243–270. 2013.
- [51] H. Barlow, *Possible principles underlying the transformations of sensory messages*, *Sensory communication* **6** (1961), no. 2 57–58.
- [52] D. M. MacKay, *The epistemological problem for automata*. Automata Studies, Princeton, NJ: Princeton University Press, 1956.
- [53] U. Neisser, *Cognition and reality: Principles and implications of cognitive psychology*. WH Freeman/Times Books/Henry Holt & Co, 1976.
- [54] D. H. Ballard, G. E. Hinton, T. J. Sejnowski, et al., *Parallel visual computation*, *Nature* **306** (1983), no. 5938 21–26.
- [55] D. Mumford, *On the computational architecture of the neocortex - II. The role of the thalamo-cortical loop*, *Biological Cybernetics* **66** (1991), no. 241-251 135–145.
- [56] M. Kawato, H. Hayakawa, and T. Inui, *A forward-inverse optics model of reciprocal connections between visual cortical areas*, *Network: Computation in Neural Systems* (2009).
- [57] P. Dayan, G. E. Hinton, R. M. Neal, and R. S. Zemel, *The Helmholtz machine.*, *Neural computation* **7** (1995), no. 5 889–904.
- [58] K. Friston, *Learning and inference in the brain*, *Neural networks : the official journal of the International Neural Network Society* **16** (nov, 2003) 1325–1352.

- [59] K. Friston, *A theory of cortical responses.*, *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **360** (2005), no. 1456 815–36.
- [60] K. Friston, J. Kilner, and L. Harrison, *A free energy principle for the brain*, *Journal of Physiology-Paris* **100** (2006), no. 1 70–87.
- [61] H. v. Helmholtz, *Theorie der luftschwingungen in röhren mit offenen enden.*, *Journal für die reine und angewandte Mathematik* **57** (1860) 1–72.
- [62] G. E. Hinton, *Training products of experts by minimizing contrastive divergence*, *Neural Computation* **14** (2002), no. 8 1771–1800.
- [63] G. E. Hinton, P. Dayan, B. J. Frey, and R. M. Neal, *The "wake-sleep" algorithm for unsupervised neural networks.*, *Science* **268** (1995), no. 5214 1158–1161.
- [64] P. O. Hoyer, *Non-negative Matrix Factorization with Sparseness Constraints*, *The Journal of Machine Learning Research* **5** (2004) 1457–1469, [[0408058](#)].
- [65] P. O. Hoyer and A. Hyvärinen, *Independent component analysis applied to feature extraction from colour and stereo images.*, *Network: Computation in Neural Systems* **11** (2000), no. 3 191–210.
- [66] D. D. Lee and H. S. Seung, *Learning the parts of objects by non-negative matrix factorization.*, *Nature* **401** (1999), no. 6755 788–91, [[1408.1149](#)].
- [67] H. Barlow, *What is the Computational Goal of the Neocortex?*, 1994.
- [68] T. S. Lee and D. Mumford, *Hierarchical Bayesian inference in the visual cortex*, *Journal of the Optical Society of America A: Optics and Image Science, and Vision* **20** (jul, 2003) 1434–1448.
- [69] A. Yuille and D. Kersten, *Vision as Bayesian inference: analysis by synthesis?*, *Trends in Cognitive Sciences* **10** (2006), no. 7 301–308.

- [70] D. J. Field, *What is the goal of sensory coding?*, *Neural computation* **6** (1994), no. 4 559–601.
- [71] E. P. Simoncelli, *Vision and the statistics of the visual environment*, *Current Opinion in Neurobiology* **13** (2003), no. 2 144–149.
- [72] A. Hyvärinen, J. Hurri, P. O. Hoyer, and H. Glasman-Deal, *Natural Image Statistics*. Springer, 2009.
- [73] Y. Dan, J. Atick, and R. C. Reid, *Efficient coding of natural scenes in the lateral geniculate nucleus: experimental test of a computational theory.*, *The Journal of Neuroscience* **16** (1996), no. 10 3351–3362.
- [74] M. Meister and M. J. Berry, *The neural code of the retina.*, *Neuron* **22** (1999) 435–450.
- [75] B. A. Olshausen and D. J. Field, *Sparse coding with an incomplete basis set: a strategy employed by V1*, *Vision Research* **37** (1997), no. 23 3311–3325.
- [76] E. P. Simoncelli and B. A. Olshausen, *Natural image statistics and neural representation*, *Annual review of neuroscience* **24** (2001) 1193–1216.
- [77] J. J. Atick, *Could information theory provide an ecological theory of sensory processing?*, *Network: Computation in Neural Systems* **3** (1992), no. 2 213–251.
- [78] H. Hosoya and A. Hyvärinen, *A Hierarchical Statistical Model of Natural Images Explains Tuning Properties in V2*, *The Journal of Neuroscience* **35** (2015), no. 29 10412–10428.
- [79] J. F. M. Jehee, C. Rothkopf, J. M. Beck, and D. H. Ballard, *Learning receptive fields using predictive feedback*, *Journal of Physiology-Paris* **100** (2006), no. 1-3 125–132.

- [80] J. F. M. Jehee and D. H. Ballard, *Predictive feedback can account for biphasic responses in the lateral geniculate nucleus*, *PLoS Computational Biology* **5** (2009), no. 5 e1000373.
- [81] M. W. Spratling, *Predictive Coding as a Model of Response Properties in Cortical Area V1*, *The Journal of Neuroscience* **30** (2010), no. 9 3531–3543.
- [82] M. W. Spratling, *Unsupervised learning of generative and discriminative weights encoding elementary image components in a predictive coding model of cortical function*, *Neural Computation* **24** (2012), no. 1 60–103.
- [83] Bubic, *Prediction, cognition and the brain*, *Frontiers in Human Neuroscience* **4** (2010), no. March 1–15.
- [84] M. Fiorani Júnior, M. G. Rosa, R. Gattass, and C. E. Rocha-Miranda, *Dynamic surrounds of receptive fields in primate striate cortex: a physiological basis for perceptual completion?*, *Proceedings of the National Academy of Sciences of the United States of America* **89** (1992), no. 18 8547–8551.
- [85] P. S. Churchland and V. S. Ramachandran, *Filling in: Why dennett is wrong*, *Consciousness in philosophy and cognitive neuroscience* (1994) 65–91.
- [86] H. Awater, J. R. Kerlin, K. K. Evans, and F. Tong, *Cortical representation of space around the blind spot*, *Journal of neurophysiology* **94** (2005), no. 5 3314–3324.
- [87] S. P. Tripathy, D. M. Levi, and H. Ogmen, *Two-dot alignment across the physiological blind spot*, *Vision research* **36** (1996), no. 11 1585–1596.
- [88] P. Andrews and F. Campbell, *Images at the blind spot*, *Nature* **353** (1991) 308.
- [89] D. C. Dennett, *Consciousness explained*. Penguin UK, 1993.
- [90] F. H. Durgin, S. P. Tripathy, and D. M. Levi, *On the filling in of the visual blind spot: Some rules of thumb*, *Perception* **24** (1995), no. 7 827–840.

- [91] F. Campbell, J. Kulikowski, and J. Levinson, *The effect of orientation on the visual resolution of gratings*, *The Journal of physiology* **187** (1966), no. 2 427.
- [92] D. E. Mitchell, R. D. Freeman, and G. Westheimer, *Effect of orientation on the modulation sensitivity for interference fringes on the retina*, *JOSA* **57** (1967), no. 2 246–249.
- [93] A. R. Girshick, M. S. Landy, and E. P. Simoncelli, *Cardinal rules: visual orientation perception reflects knowledge of environmental statistics.*, *Nature neuroscience* **14** (2011), no. 7 926–932, [[NIHMS150003](#)].
- [94] S. Appelle, *Perception and discrimination as a function of stimulus orientation: the " oblique effect" in man and animals.*, *Psychological bulletin* **78** (1972), no. 4 266.
- [95] E. A. Essock, J. K. DeFord, B. C. Hansen, and M. J. Sinai, *Oblique stimuli are seen best (not worst!) in naturalistic broad-band stimuli: A horizontal effect*, *Vision Research* **43** (2003), no. 12 1329–1335.
- [96] D. M. Coppola and L. E. White, *Visual experience promotes the isotropic representation of orientation preference.*, *Visual neuroscience* **21** (2004), no. 1 39–51.
- [97] E. Switkes, M. J. Mayer, and J. A. Sloan, *Spatial frequency analysis of the visual environment: Anisotropy and the carpentered environment hypothesis*, *Vision Research* **18** (jan, 1978) 1393–1399.
- [98] B. C. Hansen, E. a. Essock, Y. Zheng, and J. K. DeFord, *Perceptual anisotropies in visual processing and their relation to natural image statistics.*, *Neural networks : the official journal of the International Neural Network Society* **14** (2003), no. 3 501–526.

- [99] R. Ben-Yishai, R. L. Bar-Or, and H. Sompolinsky, *Theory of orientation tuning in visual cortex.*, *Proceedings of the National Academy of Sciences of the United States of America* **92** (apr, 1995) 3844–3848.
- [100] D. J. Mannion, J. S. McDonald, and C. W. G. Clifford, *Orientation anisotropies in human visual cortex.*, *Journal of neurophysiology* **103** (2010), no. 6 3465–3471.
- [101] R. T. Maloney and C. W. G. Clifford, *Orientation anisotropies in human primary visual cortex depend on contrast.*, *NeuroImage* **119** (oct, 2015) 129–45.
- [102] B. Chapman and T. Bonhoeffer, *Overrepresentation of horizontal and vertical orientation preferences in developing ferret area 17.*, *Proceedings of the National Academy of Sciences of the United States of America* **95** (1998), no. 5 2609–2614.
- [103] H. Neumann, L. Pessoa, and T. Hansen, *Visual filling-in for computing perceptual surface properties.*, *Biological Cybernetics* **369** (2001), no. 5 355–369.
- [104] S. Grossberg, E. M. Center, A. Systems, A. V. Mechanisms, N. R. G. Onr-N, A. F. Office, S. R. G. Afosr, C. Suchta, C. Street, and E. Mingolla, *Neural Dynamics of Form Perception: Boundary Completion, Illusory Figures, and Neon Color Spreading.*, *Psychological Review* **92** (1985), no. 2 173–211.
- [105] S. Grossberg and S. Hong, *A neural model of surface perception: lightness, anchoring, and filling-in.*, vol. 19. jan, 2006.
- [106] H. Neumann and W. Sepp, *Recurrent V1-V2 interaction in early visual boundary processing.*, *Biological Cybernetics* **81** (1999), no. 5-6 425–444.
- [107] H. Hosoya, *Multinomial Bayesian Learning for Modeling Classical and Nonclassical Receptive Field Properties.*, *Neural Computation* **24** (2012), no. 8 2119–2150.

-
- [108] A. M. Bastos, W. M. Usrey, R. A. Adams, G. R. Mangun, P. Fries, and K. J. Friston, *Canonical microcircuits for predictive coding*, *Neuron* **76** (2012), no. 4 695–711.
- [109] P. Kok and F. P. de Lange, *Predictive coding in sensory cortex*, in *An introduction to model-based cognitive neuroscience*, pp. 221–244. Springer, 2015.
- [110] R. P. N. Rao and T. J. Sejnowski, *Predictive Coding, Cortical Feedback, and Spike-Timing Dependent Plasticity*, *Probabilistic models of the brain* (2002) 297.
- [111] P. Kok, *Perceptual inference: A matter of predictions and errors*, *Current Biology* **26** (2016), no. 17 R809–R811.
- [112] A. H. Bell, C. Summerfield, E. L. Morin, N. J. Malecek, and L. G. Ungerleider, *Encoding of stimulus probability in macaque inferior temporal cortex*, *Current Biology* **26** (2016), no. 17 2280–2290.
- [113] G. Westheimer, *Anisotropies in peripheral vernier acuity.*, *Spatial vision* **18** (2005), no. 2 159–167.
- [114] Y. Zhang, X. Li, J. M. Samonds, and T. S. Lee, *Relating functional connectivity in V1 neural circuits and 3D natural scenes using Boltzmann machines*, *Vision Research* **120** (2016) 121–131.
- [115] X. Li, J. M. Samonds, Y. Liu, and T. S. Lee, *Pairwise interaction of V1 disparity neurons depends on spatial configural relationship between receptive fields as predicted by 3D scene statistics*, in *Society of Neuroscience Conference Abstract*, 2012.
- [116] J. M. Samonds, B. R. Potetz, and T. S. Lee, *Relative luminance and binocular disparity preferences are correlated in macaque primary visual cortex*,

- matching natural scene statistics.*, *Proceedings of the National Academy of Sciences of the United States of America* **109** (2012), no. 16 6313–8.
- [117] S. W. Zucker, *Stereo, shading, and surfaces: Curvature constraints couple neural computations*, *Proceedings of the IEEE* **102** (2014), no. 5 812–829.
- [118] A. Bartels, *Visual perception: early visual cortex fills in the gaps*, *Current Biology* **24** (2014), no. 13 R600–R602.
- [119] M. Boerlin, C. K. Machens, and S. Denève, *Predictive Coding of Dynamical Variables in Balanced Spiking Networks*, *PLoS Computational Biology* **9** (2013), no. 11.
- [120] V. S. Ramachandran, *Filling in gaps in perception: Part i*, *Current Directions in Psychological Science* **1** (1992), no. 6 199–205.
- [121] V. S. Ramachandran, *Filling in gaps in perception: Part ii. scotomas and phantom limbs*, *Current Directions in Psychological Science* **2** (1993), no. 2 56–65.
- [122] J. J. Gibson, *The senses considered as perceptual systems.*, .