

Computational Modelling of Visual Illusions

Astrid Zeman, B.Eng (Software Hons I) Dip. Eng. Prac

Department of Cognitive Science

ARC Centre of Excellence in Cognition and its Disorders

Faculty of Human Sciences

Macquarie University, Sydney, Australia

This thesis is presented for the degree of Doctor of Philosophy (PhD)

February, 2015

Table of Contents

Computational Modelling of Visual Illusions.....	1
Thesis Abstract.....	5
Statement of Candidate	6
Acknowledgments	7
1 Introduction.....	9
<i>1.1 General overview</i>	<i>10</i>
<i>1.2 Defining illusions</i>	<i>14</i>
1.2.1 A general definition	14
1.2.2 Illusions as source reconstructions	16
<i>1.3 Categorising illusions</i>	<i>19</i>
<i>1.4 Aims, strengths and limitations of models</i>	<i>22</i>
1.4.1 Limitations of Models	22
1.4.2 From theories to models and back.....	25
<i>1.5 Marr's different levels of description.....</i>	<i>26</i>
<i>1.6 Biological analogies</i>	<i>28</i>
1.6.1 Historical influences from biology	28
1.6.2 Biologically inspired systems.....	31
<i>1.7 Computational models of vision.....</i>	<i>32</i>
1.7.1 Deterministic versus probabilistic	32
<i>1.8 Pre-cortical models.....</i>	<i>34</i>
1.8.1 Historical context	34
1.8.2 Current models applied to illusions	36
1.8.3 Our model selection: Exponential filter family model	38
<i>1.9 Ventral stream models</i>	<i>40</i>
1.9.1 General properties	40
1.9.2 Hierarchical models in historical context	42
1.9.3 Our model selection: feed-forward model HMAX	43
1.9.4 Feedback models	45
<i>1.10 Existing computational models of visual illusions</i>	<i>46</i>

1.10.1	How to model illusions.....	46
1.11	Scope of this thesis	48
1.12	Thesis layout	49
1.13	References	49
2	Study 1	60
	<i>Abstract</i>	61
2.1	<i>Introduction</i>	61
2.2	<i>Methods</i>	65
2.2.1	HMAX Layer Descriptions	66
2.2.2	Task Description.....	68
2.2.3	Experimental Setup	68
2.3	<i>Results</i>	70
2.3.1	Experiment I: Control.....	70
2.3.2	Experiment II: Illusion Effect.....	71
2.3.3	Experiment III: Illusion Strength Affected by Angle.....	73
2.4	<i>Discussion</i>	76
2.5	<i>References</i>	84
2.6	<i>Appendix: Determining whether low spatial frequency information may be influencing the SVM</i>	87
2.6.1	Stage I: Extracting the highest weights entering the SVM layer	87
2.6.2	Stage II: Identify the spatial scale of the top contributing features in C2.	89
3	Study 2	92
3.1	<i>Abstract</i>	93
3.2	<i>Introduction</i>	94
3.3	<i>Materials and Methods</i>	99
3.3.1	Computational model: HMAX	99
3.3.2	Stimuli: Training and test sets (Control and Müller-Lyer).....	100
3.3.3	Procedure: Learning, parameterization, illusion classification	103
3.4	<i>Results</i>	104
3.4.1	Experiment I: Classification of ML images after each level of HMAX	104
3.4.2	Experiment II: HMAX classification of ML images with reduced variance	110

3.5	<i>Discussion</i>	112
3.6	<i>References</i>	117
4	Study 3	121
	<i>Abstract</i>	122
4.1	<i>Introduction</i>	123
4.2	<i>Material & Methods</i>	129
4.2.1	Stimuli	129
4.2.2	Model	132
4.3	<i>Results</i>	138
4.4	<i>Discussion</i>	144
4.5	<i>References</i>	154
5	Discussion & Conclusion	158
5.1	<i>Chapter overview</i>	159
5.2	<i>Summary and integration of studies</i>	159
5.2.1.	Each of the studies in summary	159
5.2.2.	Common threads between studies	161
5.3	<i>Our studies in the context of wider academic literature</i>	165
5.3.1.	Can illusions manifest in artificial brains?	165
5.3.2.	Alternatives to HMAX and associated illusion predictions	167
5.3.3.	Alternatives to the exponential model and associated illusion predictions	169
5.4	<i>Possible improvements and future studies</i>	173
5.4.1.	Evaluating theories using computer models	173
5.4.2.	Extensions and limitations of the HMAX model	176
5.4.1.	Extensions and limitations of the exponential filter model	187
5.5	<i>Applications for computer modelling of illusions</i>	190
5.5.1.	Autonomous systems and engineering applications	190
5.5.2.	Links with psychological disorders	192
5.6	<i>Closing Remarks</i>	196
5.7	<i>References</i>	198

Thesis Abstract

Illusions reveal some of the sophisticated, underlying neural mechanisms that often remain hidden in our day-to-day visual experience. Illusions have traditionally been studied using psychophysical methods, which quantify overall, system-level effects observable at the highest layer of the visual hierarchy. This thesis applies the relatively new technique of computational modelling to the study of visual illusions, to quantify bias and uncertainty within various levels of our visual system. The method adopted in this thesis merges statistical inferences, obtained from exposure to image subsets, with filtering operations that mimic visual neural processing from layer to layer. Previous computational models of visual illusions have considered these in isolated arrangement. This dissertation highlights the benefits of combinatorial modelling, which includes separating out the contribution of neural operations from potential statistical influences.

The first study in this dissertation investigates a well-known line-length illusion in a benchmark model of the visual ventral stream, demonstrating that a model imitating the structure and function of our cortical visual system is susceptible to illusions. In the second study, we further scrutinise this line-length illusion inside each layer of the benchmark model, observing magnitudes of uncertainty and bias that propagate through each level. In the third and final study, we introduce a new model based on exponential filters inspired by contrast statistics of natural images. We apply a suite of lightness illusions to this new model and demonstrate that low-level kernel operations can account for a large set of these illusions. In summary, this thesis shows that combining filtering functions with natural image statistics not only allows for illusory bias and uncertainty to be imitated in artificial neural network models, but it also provides further evidence for and against some proposed theories of visual illusions.

Statement of Candidate

I certify that the work in this thesis entitled “Computational modelling of visual illusions” has not previously been submitted for a degree nor has it been submitted as part of requirements for a degree to any university or institution other than Macquarie University.

I also certify that the thesis is an original piece of research and it has been written by me. Any help and assistance that I have received in my research work and the preparation of the thesis itself have been appropriately acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

.....

Astrid Zeman (Student ID: 41959639)

February, 2015

Acknowledgments

I am indebted first and foremost to my principal supervisor Kevin Brooks (MQ) who has invested much of his time and effort in guiding me through my PhD. My co-supervisor Oliver Obst (CSIRO) provided a lot of support for me over the years and I'm happy to have worked together with him for nearly a decade. I was privileged to work directly with Sennay Ghebreab (UvA) in Amsterdam for 6 months, who helped me to greatly expand my knowledge and who demonstrated by example how to best manoeuvre along the intersection between machine learning and neuroscience.

At Macquarie University, I was fortunate enough to be a part of the cognitive science department, which, when I started, was headed by Max Coltheart. I have made many friendships in the department, some that are lifelong. I've been inspired by many female researchers at Macquarie, some whom make excellent and much needed role models for women in science. These women notably include Kiley Seymour, Susan Wardle and, most inspiring of all, Doris McIlwain.

My desire to pursue an academic career began back in 2003, with large thanks to Mary-Anne Williams. She was instrumental in expanding robotics research at UTS and inspiring young undergraduates to explore new technologies and make the most of their imagination. Lecturer Steve Murray provided me with consistent encouragement and glowing references, spurring me onto a path of research.

From 2005 to 2009, I very much enjoyed working with the autonomous systems lab at the CSIRO. Here I worked under Mikhail Prokopenko, who instilled in me the value of producing high quality science. The late Don Price and many other gems in CSIRO upheld the quality of science in all of their work. I had the joy of working with very talented and intelligent people with quirky and enjoyable personalities, including Matthew Chadwick, Joe Lizier, Rosalind X. Wang, Piraveenan Mahendrarajah, Ying Guo, Vadim Gerasimov, Scott Heckbert, Cameron Fletcher and many more.

Finally, I'd like to acknowledge all of my cohorts who have enjoyed and suffered through the pursuit of research. These include pod-mates Shahd, Rob, Shu, Trudy, Lisa, Anne, Adam, Kellie, Yatin & Steve. Gödel, Escher, Bach groupies Vince, Lincoln, Jordan & more. Dancers Shiree, Kris, Nathan & Yvette. Acrobat Britta. Crossword enthusiasts Sam & Peter. Dinosaur loving Gen. Admin angels Lesley & Teri. The effervescent Monica, Lisi & Linda. Other layabouts Enzo, Jacopo, Vishnu, Xenia, Serje, Nora, Rochelle, Huachen, Thushara, Jonathan, Em, Brendan, Mel, Loes, Regine, Marissa, & many, many more.

To all of my biological and non-biological family who have provided support, thank you.

About this Thesis

This thesis has been prepared in accordance with the Macquarie University Journal article format thesis guidelines. Each chapter has been written in the format of a self-contained journal article. Where possible, all attempts have been made to minimise any referencing and stylistic inconsistencies between the chapters. Some small amendments have been made to the chapters contained within this thesis that differ from the published works. Most notably, Chapter 2 now includes an Appendix (Section 2.6).

Study 1 has been published as:

Zeman, A., Obst, O., Brooks, K. R., & Rich, A. N. (2013). The Müller-Lyer Illusion in a computational model of biological object recognition. *PLoS ONE*, 8(2), e56126. DOI:10.1371/journal.pone.0056126.

Study 2 has been published as:

Zeman, A., Obst, O., & Brooks, K. R. (2014). Complex cells decrease errors for the Müller-Lyer illusion in a model of the visual ventral stream. *Frontiers in Computational Neuroscience*, 8(112). doi:10.3389/fncom.2014.00112

Study 3 has been published as:

Zeman, A., Brooks, K. R & Ghebreab, S. (2015). An exponential filter model predicts lightness illusions. *Frontiers in Human Neuroscience*, 9(368), doi: 10.3389/fnhum.2015.00368.

Author Contributions

Study 1: Conceived and designed the experiments: AZ ANR KRB OO. Performed the experiments: AZ. Analysed the data: AZ OO. Wrote the paper: AZ ANR KRB OO.

Study 2: Conceived and designed the experiments: AZ KRB OO. Performed the experiments: AZ. Analysed the data: AZ KRB. Wrote the paper: AZ KRB.

Study 3: Conceived and designed the experiments: AZ, SG. Performed the experiments: AZ. Analysed the data: AZ, SG. Wrote the paper: AZ KRB, SG.

1 Introduction

1.1 General overview

“It may be necessary to invent imaginary brains — by constructing functional machines and writing computer programs to perform perhaps much like biological systems. In short, we may have to simulate to explain; though simulations are never complete or perfect.”

(Gregory, 1963)

Every day we encounter visual illusions. From the moment we switch on our phones, televisions or computer screens, we perceive continuous image motion instead of a series of flickering static visual frames. Our brains are able to effortlessly stitch together a series of still images into one smooth moving percept, by making continuous sensory predictions that may or may not accurately reflect our external environment. This example is one of many that highlights the ubiquity of visual illusions in our day-to-day lives and the rich insights that they offer in further understanding our minds.

Illusions have long been recruited as a method for uncovering hidden neural processes. The method of choice in experiments to date has been psychophysical, measuring system-level effects that quantify biases that are present when viewing a range of stimuli. In the past couple of decades, neuroimaging techniques have gained popularity in examining where and when biases occur. A very recent and under-utilised technique, computational modelling, allows researchers to explore how illusions might occur by specifying the nature of neural computations that could bring about certain biases. This thesis applies specific computational models to the study of visual illusions and highlights some advantages of adopting such an approach.

In this dissertation, we demonstrate the versatility of computational modelling by applying it to two types of illusion: one type of illusion that deals with line-length discrimination and another type that deals with lightness judgements. We present two different types of models: a well-known hierarchical feature model that approximates the visual ventral stream, and an in-house model of early visual processing. By recruiting two different models that represent either the cortical or pre-cortical stages of visual processing, we show that illusions can be modelled at multiple stages of the visual hierarchy.

In the first study, we demonstrate a well-known line-length illusion, the Müller-Lyer illusion, in a benchmark model of the visual ventral stream known as HMAX (Serre et al., 2005; Mutch and Lowe, 2008). The Müller-Lyer illusion (Figure 1-1a) occurs when a line with arrowheads or arrow-tails appears shorter or longer respectively (Müller-Lyer, 1889). The underlying cause of the illusion has been under contention since its inception (Müller-Lyer, 1889; Heymans, 1896; Lewis, 1909; Pieron, 1911) with prominent contenders including statistical correlations of line configurations in the environment (Howe and Purves, 2005b), the top-down application of size constancy scaling rules learnt by exposure to natural images (Gregory, 1963), a reliance on low spatial frequency information (Carrasco et al, 1986), and bottom-up neurological mechanisms, such as lateral inhibition (Coren, 1970). The model produces a larger bias when classifying Müller-Lyer images with more acute fin angles, consistent with human observers. Our training images are all artificially constructed, demonstrating that exposure to natural images is not a necessary condition for bringing about the illusion. We also find that there is no additional reliance on the outputs of low spatial frequency filters over those of high spatial frequency, showing that this is also an unnecessary condition. Following from these experiments, we conducted another study to delve more deeply into the potential underlying causes of the Müller-Lyer illusion and measure how bias is transformed within the model.

In the second paper, we extend the first study to examine the contribution of complex and simple cell operations in HMAX towards Müller-Lyer bias and uncertainty. This study is essentially a lesioning analysis, where layers of the model are removed one by one to observe the effect this has on classifying Müller-Lyer images. We discover that there is an initial bias present in the input images, representative of the possible influence of image statistics in generating the illusion (Howe and Purves, 2005b). Our study reveals that any processing of Müller-Lyer images within the model reduces bias when compared to classifying images directly. We hypothesise that increasing the variance of line positions in input images would engage complex cell operations and therefore reduce bias levels. We find that, in particular, complex cell operations reduce levels of uncertainty and bias in 87.5% of cases.

In the third and final study, we move away from the aforementioned cortical model of visual processing to a model based on pre-cortical mechanisms. This model uses filters optimised for natural images that also approximate horizontal cell operations found in the retina. In addition, the model incorporates normalisation functions that are based on contrast gain control principles found in the LGN. We test a battery of lightness illusions using this model, some of which are purported to involve higher-level processing. The model is able to accurately predict the direction of 24 out of 27 illusion, accounting for a large proportion of lightness illusions using solely low-level mechanisms. As with the previous study, we demonstrate that a combination of statistical influences and filtering operations are able to bring about a bias commensurate with human perceptual experience.

The studies included in this thesis showcase sophisticated modelling techniques that are robust enough to apply to a large breadth of illusory stimuli. These papers supplement and complement existing neurological and psychophysical studies. Moreover, they generate predictions for studies using other methods. Each paper provides further evidence for and

against proposed theories behind each illusion. We are able to rule out some of the necessary causes of certain illusions and tease apart the separate contributions of different causes leading to particular illusions. Using feed-forward models, we identify the contribution of low-level, bottom-up mechanisms towards each of these illusions (as distinct from higher-level, top-down influences). We lesion layers to determine the progression of illusory bias and uncertainty layer to layer within a network. We are also able to manipulate the parameters of the models that represent or simulate various neural properties, such as levels of lateral inhibition or the sizes of certain neural populations, to determine how this affects accuracy and precision. Not only are we able to show the separate contributions of low-level versus higher-level mechanisms, but we are able to tease apart and quantify the contribution of specific mechanisms found within the various levels of the visual hierarchy (such as filtering versus contrast gain normalisation) towards illusory effects.

This thesis highlights that illusory effects are brought about by a complex interplay between statistical influences obtained from images in the environment, with neural operations that transform an image into its sensory representation. The studies suggest that neural operations shaped by image statistics are responsible for bringing about illusory effects, as put forward by Coren (1970), Bertulis and Bulatov (2001), Howe and Purves (2005) and Corney and Lotto (2007), to name just a few. It is evident that there is a complex integration of both external and internal factors that bring about a range of illusory effects. This has implications for artificial systems that mimic their biological counterparts, demonstrating their susceptibility to certain illusions and suggesting that these biases can be mitigated by the careful selection of training images. Furthermore, this thesis also impacts on human perception, bringing about further understanding of some of the likely origins of certain illusions and demonstrating the power and capability of computational modelling as a tool to better inform our understanding of illusions.

1.2 Defining illusions

1.2.1 A general definition

Visual illusions have long been established as an aid to uncover some of the fundamental mechanisms that underlie our visual perception. Illusions are defined as “systematic visual and other sensed discrepancies from simple measurements with rulers, photometers, clocks and so on” (Gregory, 1997, p.2). Gregory’s definition is apt in that it highlights two key points about illusions. Firstly, he uses the term *systematic*, to demonstrate that perceptual discrepancies for illusions are not simply one-off events, but repeatable biases. Secondly, Gregory refers to “visual and sensed discrepancies” that differ from other cognitive discrepancies such as belief disorders.

This thesis deals with *visual illusions* that are known to occur through mechanisms in the brain, as distinct from *optical illusions* that are caused by properties of the light and eye. Visual illusions help to reveal underlying assumptions and algorithms that the brain uses (Gilchrist, 2003). Optical illusions, on the other hand, are useful for studying physical light properties. For example, a straw appears bent in water due to light being refracted by travelling through two different mediums of water and air. In the interests of revealing potential neural mechanisms behind particular illusions, this dissertation focuses on visual illusions.

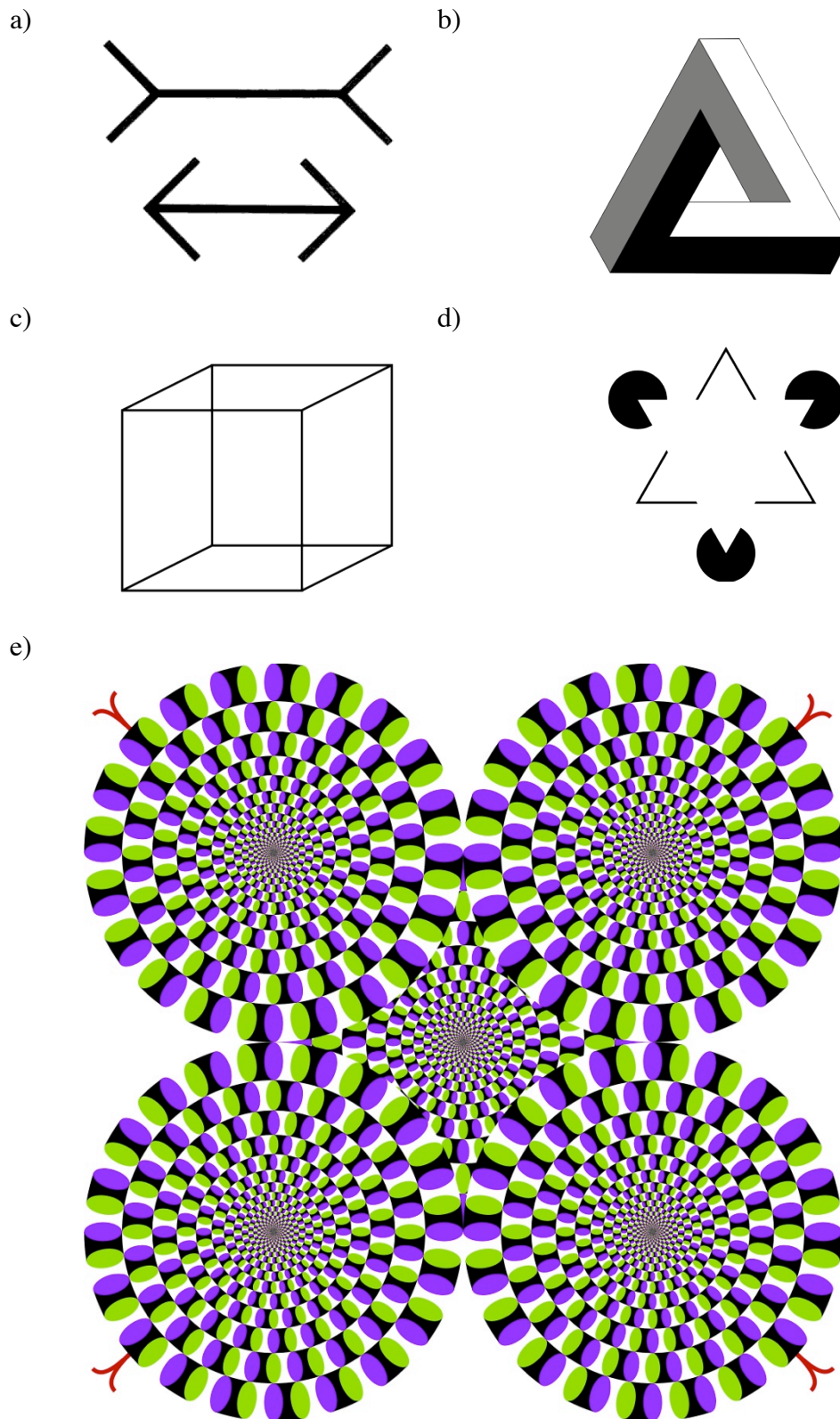


Figure 1-1: Illusions in this chapter in order of mention a) Müller-Lyer (1889) b) Penrose triangle (Penrose & Penrose, 1958) c) Necker cube (Necker, 1832) d) Kanisza (1976) triangle e) Kitaoka's (2003) rotating snakes

While it is common to define illusions as a discrepancy between perception and physical reality, some researchers argue that it is difficult to formulate an adequate description of reality, especially when an illusion is presented in an artificially contrived situation (Rogers, 2014). For example, take the Penrose Triangle (Penrose & Penrose, 1958, Figure 1-1b), which is an impossible object that would not exist in reality only insofar as we assume that all edges are straight and converge. Finding a way to measure this object in the environment, or construct a veridical description of it, would be difficult if not impossible. For many illusions however, it is possible to directly measure the discrepancy between reality and perception, by using physical tools to quantify an object in the external (to the observer) world, and psychophysics to quantify the internal (to the observer) perceptual experience.

1.2.2 Illusions as source reconstructions

Defining illusions as a discrepancy between physical reality and internal perceptual experience is sometimes referred to as an “error” in perception. The use of this term can be misleading, because it may be that your perceptual system is not actually making any “mistakes” at all, in terms of computing what is presented to it, but is instead making the best inferences or predictions possible given the limited information available. In many cases, illusions are constructed to leave out information that would otherwise be normally present in the natural world, where we would rely on context for interpreting stimuli. Taking the Penrose triangle example again (Penrose & Penrose, 1958), we know that each individual corner of the object is presented in a plausible manner in 3 dimensional space, yet the object as a whole is implausible. Here it is evident that we are simply making the best inference possible, given a stimulus that is locally consistent but globally contradictory.

There are many examples that demonstrate the idea that our visual system is making the best conceivable estimation given impoverished, ambiguous or inconsistent stimuli. This has led Weiss *et al.* (2002) and others to refer to illusions as “optimal percepts”, where instead of

making an “error” in judgment, our perceptual system is behaving optimally overall and when viewing an illusion, prior assumptions lead the system to the inappropriate conclusion. Our visual system reconstructs the most likely stimulus that would produce our current percept, commonly referred to as the ideal source. To discover the ideal source, our brains would first estimate the probabilities of certain features or properties inherent in natural images that we have been exposed to over our previous experience. The underlying patterns that our brain extracts may relate to co-occurring features or image properties, such as contrast distributions (Field, 1987; Zhu and Mumford, 1997). After establishing the patterns that arise in natural image statistics, our perceptual system will apply these probabilities to fill in the gaps of information that is missing in the stimulus, essentially reconstructing the most probable real-world source of the retinal image (Dakin and Bex, 2003; Corney and Lotto, 2007; Brown and Friston, 2012).

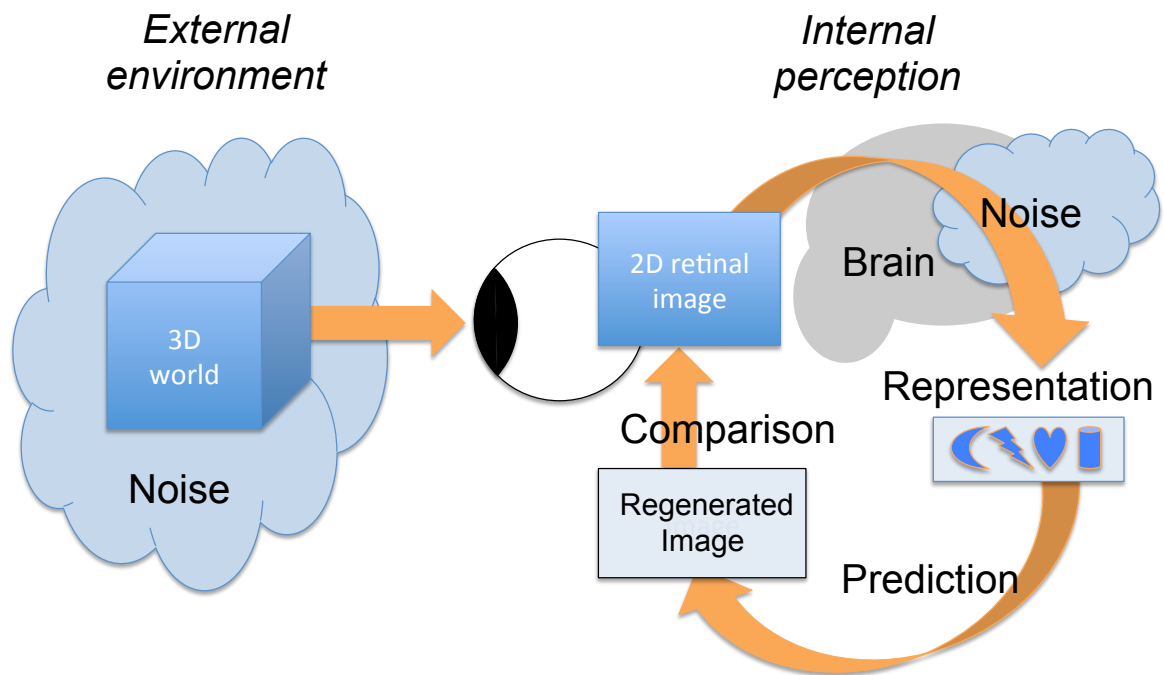


Figure 1-2. The inverse problem regenerates an image from representation

Regenerating the best source by exploiting typical underlying patterns in natural images is referred to as the inverse optics problem (Purves and Lotto, 2010) or just as the inverse problem (Hill and Johnston, 2007). In all descriptions of the inverse problem, there are some clear elements that define the process of information flow over time. Firstly, information flows from a retinal image through hierarchical layers of the brain where a representation is formed as a set of neural activations at each level. A representation is “a set of measurements which collectively encode the geometry and other visual qualities” (Edelman, 1997). For example, a set of lines and shaded areas viewed by the eye may trigger a representation in higher levels of the brain. Once a representation is generated, this allows for information to flow back down the hierarchy to regenerate or best predict the initial stimulus. During the execution of these two steps, time will have passed and slight changes in the environment will have occurred. This means that the prediction that is generated will have altered slightly compared to a new input image due to internal noise, (such as stochastic variations in neural firing), or external noise (such as changes in lighting, etc). Comparing predictions that are generated from the brain (source reconstructions) with the updated stimulus determines the amount of error between prediction and reality. Figure 1-2 illustrates the information flow of the inverse problem.

The inverse problem originated from Helmholtz’s idea of “unconscious inference” – that visual impressions are formed in an involuntary fashion, a process oblivious to the observer (Helmholtz, 1867). The concept of unconscious inference highlighted that observers continuously make predictions about sense data without even being aware of the absence or degradation of information. Helmholtz’s idea has since been explored at the neural level, looking at how information can be encoded, transported and recoded to reconstruct the input. Our visual system will often receive input that is noisy or missing, and yet is able to

reconstruct a perceived stimulus source despite this.

Helmholtz's idea of unconscious inference is nowadays more formally interpreted as being an early example of predictive coding. Predictive coding was highlighted as a general principle in low-level vision, whereby noise within a system would be reduced by exploiting existing patterns found in natural scenes (Srinivasan et al, 1982). This principle was also extended to higher-level visual areas (Rao and Ballard, 1999). Predictive coding has been popularly implemented into generative computer models, to reconstruct “fantasies” of the most likely stimuli based on probabilities of simultaneously occurring features (Hinton and Zemel, 1994; Mumford, 1994; Lee and Mumford, 2003; Kersten *et al.*, 2004; Friston, 2005). These networks are sometimes referred to as hierarchical predictive coding models. While Clark (2013) reviews the evolution of these models and proposes their extension into the future, section 1.9 includes further elaboration on this class of model.

1.3 Categorising illusions

Through the decades there have been numerous attempts to systematise illusions, either by their causes or by their appearance, to identify some of the potential common causes that may underlie some illusions. Classifying illusions by their appearance has a potentially far greater chance of consensus among researchers compared to classifying illusions by their causes, since the origin of many illusions is still in dispute. For now, we outline some of the proposed classifications of illusions that include defining causes as a distinguishing factor.

One of the best-known and most robust taxonomies of illusions was proposed by Richard Gregory (Gregory, 1997). He proposed a tentative classification of illusions, creating a 4 x 4 matrix of appearances and causes (Gregory, 1997). Appearances were sub-divided into four kinds: ambiguities e.g. Necker Cube (Necker, 1832) (Figure 1-1c), distortions e.g. Müller-

Lyer Illusion (Müller-Lyer, 1889) (Figure 1-1a), paradoxes e.g. Penrose Triangle (Penrose & Penrose, 1958) (Figure 1-1b), and fictions e.g. Kanisza triangle (Kanisza, 1976) (Figure 1-1d). Each illusion was then classified along another dimension based upon its proposed aetiology, using four categories: optics, signals, rules, and objects. These four causes could be grouped into those that are physical, as a result of light or optical disturbances, or cognitive, involving the misapplication of rules or specific knowledge. Gregory's specification of physical causes is straightforward and presents less room for contention. Attributing illusions to cognitive aspects is less clear, which we elaborate on below.

Gregory (1997) defines knowledge as being specific to particular classes of stimuli (using his example, that faces are convex), whereas rules are general principles that are applied to all objects and scenes (such as the Gestalt laws). Gregory also differentiates knowledge from rules in terms of the direction of information flow, such that knowledge flows "top-down" and rules are applied "sideways". Let's take for example, Gregory's (1970) hollow face or hollow mask illusion, where a mask that is rotated so that the concave side is facing the observer, is seen as a convex object. While it is clear that the hollow mask illusion is stronger in the upright position for faces compared to when it has been inverted (Hill and Bruce, 1993), it is also clear that our perception of convexity applies to other objects, creating illusions such as the hollow potato illusion (Hill and Bruce, 1994). Hill and Johnston (2007) show that the hollow illusion extends to other objects besides the humble potato, and that for objects with a canonical orientation, the illusory strength for an upright concave object is greater than for its inverted configuration. These studies indicate that any specific knowledge that is purportedly applied to the convexity of faces is also applicable to a variety of objects. From studies such as Hill and Johnston (2007), we can infer that it is not face-specific "knowledge" that is applied, but more generic object-based "rules" that apply to this illusion.

Changizi et al (2008) presented a more recent attempt to systematise illusions, this time into a 7 x 4 matrix of 28 classes, with one dimension being the property that is manipulated and the other dimension being the perception it affects. The authors distinguish 24 illusion classes based on the effects of (1) size, (2) speed, (3) luminance contrast, (4) distance, (5) eccentricity, and (6) vanishing point, on perceived (A) size, (B) speed, (C) luminance contrast, and (D) distance. They also present another 4 classes that do not clearly fit into this framework, bringing the number of illusion categories up to 28. They put forward that a single visual information processing mechanism adequately explains all of the illusions in their framework. This work was criticised shortly afterward by Briscoe (2010), who upheld Gregory's taxonomy as a more versatile grouping. Briscoe (2010) argues that the task of systematising illusions is perhaps an illusion in itself, drawing attention to the work of Coren, Girgus and Day (1973) who state that visual illusions are "multiply caused and maintained by a number of different peripheral and central factors" (p. 504). In other words, illusions may not be linearly separable by their causes. Rather than relying on one-to-one mappings between illusions and their most probable cause, it is useful to employ set theory to link illusions with their multiple causes. Set theory is a branch of mathematical logic where collections of objects form sets and an object can belong to multiple sets. Allowing for illusions to belong to multiple sets would lead to a new logical and visual configuration that taxonomises illusions with more than one aetiology. To date, illusion classifications have only been presented in tables, automatically constricting each effect to having only a one-to-one mapping with its most probable or the current most broadly accepted cause.

This thesis specifically selects illusions where there is no current consensus in the literature, with many researchers suggesting different causes behind each effect. These illusions may purportedly be the result of a number of factors, necessitating their placement across multiple groupings. Using computational networks, it is possible to separate the individual

mechanisms or operations that are applied to a stimulus to determine the necessary causes of an illusion and therefore its appropriate grouping within a taxonomy. By testing some of the necessary causes of each illusion using computational models, we can quantify the potential contributions towards each effect and build more accurate classifications.

1.4 Aims, strengths and limitations of models

1.4.1 Limitations of Models

“Since any fit of a model to data is never more accurate than the data themselves, a model is only worthwhile when it can describe data sufficiently accurately using far fewer parameters than the number of data points modeled.”

(Zhaoping, 2014, p. 2)

A model is an abstract representation of an existing, real system. This thesis looks specifically at models that emulate how information is processed from the retina to LGN and along the brain’s visual ventral stream. When constructing a model, there is always a trade-off between explanatory power (the ability to explain a phenomena with as few parameters as possible) and complexity, with the former providing greater links to proposed theories and the latter providing greater fidelity to the real system (Meeter *et al.*, 2007; Zhaoping, 2014; Low-Décarie *et al.*, 2014). This trade-off forms a sliding scale that encapsulates a range of models, where cognitive models generally aim to provide greater explanatory power, and computational neuroscience models aim to provide a precise description of data. This scale does not say that one approach is more sophisticated or difficult than the other, but emphasizes that these different approaches have separate goals in mind.

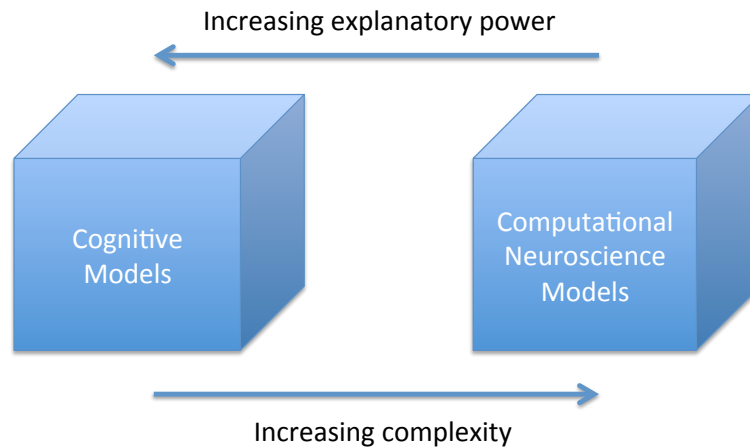


Figure 1-3: The trade-off between cognitive and computational models.

To implement a theory in a computational model, a researcher needs to consider a number of questions. What flows into and out of the system? Is information stored in memory? How would this system interact with other systems? How is information represented between levels? The process of analysing information flow and transformation inside a model implements the details of a theory and makes explicit any assumptions made. Theory and models provide the link between neurophysiology and visual perception (Zhaoping, 2014), often imitating biological architecture and neuronal functions to predict system-level behaviours.

It is essential to point out that all models come with their own set of assumptions and inbuilt limitations. As expressed by Box and Draper (1987, p. 424): “all models are wrong, but some are useful”. For example, many early artificial neural network models are incorrect since they contain artificial neurons that communicate numerical values rather than discrete spikes of activity. It is therefore important to firstly address the primary motivation behind building or using a model. Many models aim for fidelity first and foremost, which would be the dominating incentive for the majority of computer vision systems. The execution speed of the model in producing a prediction can also be a dictating factor, especially for real-time

systems. Ease of implementation is also considered valuable, as well as reducing the number of free parameters. These goals generally constitute the motivating factors for computer scientists and engineers. An equally important goal for some modellers is explanatory power: the ability of a model to provide simplified descriptions of real-life phenomena so that it can be linked to theories. This would be a more influential motivating factor for psychologists, who wish not only to predict psychophysical data, but also to explain it.

It may seem that a model that captures entire system dynamics faithfully and completely would be the perfect model. A number of current projects exist with this aim in mind, such as the “Blue Brain Project” (Markram, 2006), the Human Brain Project (Markram, 2011) and the BRAIN Initiative (Alivisatos *et al.*, 2013). However, a model that demonstrates complete and precise emulation of a real system will not necessarily provide any more information about the inner workings of that system than the original. A faithful replica of the brain would still require well-designed tests to measure the success of one theory against another. Simpler abstractions may be just as informative and require much less time and effort to implement.

In selecting the models for this thesis, we propose to find the appropriate balance between complexity and explanatory power (Figure 1-3). We select models that are able to provide adequate explanatory power while also showcasing high levels of accuracy. The main motivation for selecting each model is to test and support at least one particular theory or explanation for a selected illusion. However, there is one caveat when selecting any model: that models are not isomorphic with theories. Models can be considered as formalisations of theories, but as Norris (2005) states: ‘there is rarely a straightforward one-to-one mapping between model and theory’. There are a number of additional steps required to bridge this gap, which we describe below.

1.4.2 From theories to models and back

This thesis does not explicitly build a model to test one particular theory. Instead, we first select a particular illusion known to occur in the visual areas that are being modelled. The illusion or illusions that we select have competing theories for their underlying causes. We choose an appropriate model that emulates the functioning of the principal visual areas where a particular illusion is thought to occur, providing a platform to test one theory against another. Each study presents quantitative results that not only demonstrate the relevant model's ability to predict the direction of bias, but also measures the magnitude of the illusion, allowing these to be directly compared against human psychophysical data.

We emphasise that evidence of a particular cause in bringing about an illusion in a model does not equate to whether that cause can generalise to humans. Simulations can be compared to and linked with predictions made by theories, allowing us to assess some of the existing explanations of illusions. However, considering that our models are not exact facsimiles of the brain, we cannot explicitly generalise results from our model to human causes. In a model, we are able to hint at various factors as being necessary to bring about an illusion. These may hint about some of the likely causes of an illusion in humans and may provide support for one explanation over another. Nevertheless, these explanations are satisfactory only for models and not for humans. Our studies identify what cause is not necessary to bring about an illusion across all systems.

Many existing models are built to support one particular theory over all others in explaining a particular effect. For instance, let us look at existing models of the Müller-Lyer illusion (MLI), an effect where the perceived length of a line is contracted (or elongated) with arrowheads (or arrow-tails). Bertulis and Bulatov (2001, 2005) implemented a model based purely on filtering mechanisms, with the motivation to test only the theory of lateral inhibition

in explaining the MLI. Howe and Purves (2002, 2004, 2005a and 2005b) look only at the statistical relationships that are theorised to be responsible for the effect. Until Zeman *et al.* (2013, 2014), no attempt had been made to combine image filtering with statistical biases and quantify the level of contribution of multiple factors. As mentioned previously, illusions are most often underpinned by a number of causes (Coren, Girgus and Day, 1973).

We can see that models provide a useful device for testing theories against one another. Models allow for particular simulations that are simply not possible to conduct in humans (this is also possible using TMS and other neuroimaging techniques and we elaborate on the added benefits of models here). For example, models allow for lesioning studies, where sections of the model can be knocked out, which can also be achieved using TMS but only for outer cortical regions. Models also allow for parameters to be adjusted, such as levels of lateral inhibition, which cannot be achieved in neuroimaging or TMS. The effectiveness of computational modelling is further highlighted by looking at how models can better inform and potentially improve theories. The process of building or selecting a computational model may expose assumptions and details in a theory that have not been fully enunciated, leading to a deeper understanding of the proposed explanations. Most of all, models can provide quantifiable predictions that theories alone are not able to produce themselves. In other words, they are the ‘crucial link between theory and data’ (Norris, 2005).

1.5 Marr’s different levels of description

David Marr released his seminal work in 1982, which described different levels of analysis for a computational problem (Marr, 1982). Building on work with Tomaso Poggio (Marr & Poggio, 1977), which describes four levels of description, Marr compressed these into three, namely the “computational level”, the “algorithmic level” and the “hardware implementation

level”. The computational level addresses the end goal of the system in order to predict a certain perceptual experience, for example, to identify whether a given image contains a cat or a dog. The algorithmic level addresses how to implement the computation, describing input and output representations and the operations that transform input into output. Finally, the hardware implementation level addresses the physical system used to perform the computations.

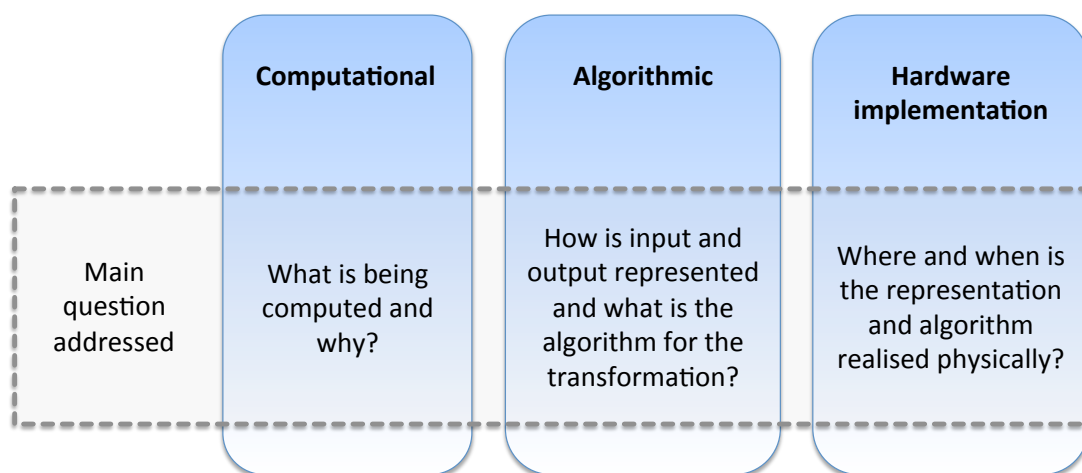


Figure 1-4 Marr's 3 levels with the corresponding research question that each level addresses (reproduced from Marr, 1982)

Marr's levels of description allow us to essentially separate computers from their computations. We know that a computer program can run on separate computers, and so the hardware implementation is independent from the program. For example, we can run a text editor on a Unix machine, a Windows machine or a Mac. Likewise, the algorithms underlying the operations can be independent from the hardware and the required task. Taking our previous example, different text editor programs could use different algorithms to manipulate text, with the Unix-based vi command-line editor being distinctly different from screen-based

text editors. Finally, the computational task can be separately defined from the hardware and the algorithmic levels, such that a text-based program can be used to write a letter, a poem, or a card.

The three levels of description address separate questions in regards to the system being modelled. The computational theory level addresses the overall question of what is being computed and why. The algorithmic level addresses how input and output are represented and the transformation process from input to output. The hardware implementation level looks at the physical realisation of the algorithm. Figure 1-4 illustrates the core question that is raised and addressed through the adoption of each of the three levels (reproduced from Marr, 1982). A model that is able to encapsulate some degree of comparison across multiple levels – the physical, the computational and the algorithmic, showcases a very strong measure of biological plausibility.

1.6 Biological analogies

1.6.1 Historical influences from biology

Over the past five decades, neurophysiology has exerted a massive influence on computational models of vision. Work in the lab of Stephen Kuffler, in the 1950s and '60s, was key in influencing the type and function of artificial neurons, their hierarchical arrangement and how information is modified and transformed within each neuron. Kuffler himself was instrumental in establishing the functional organisation of the retina, publishing landmark studies on the receptive fields of retinal ganglion cells (Kuffler, 1952; Kuffler, 1953). Kuffler showed that the maximal response rate of retinal ganglion cells occurred when light was presented in a dark surround, (for units known as on-centre cells), or when a dark centre was surrounded by a light surround (for units known as off-centre cells). It was around

this time that three soon-to-be influential researchers joined Kuffler, building on the work that he had established: David Hubel, Torsten Wiesel and Horace Barlow.

It was in Kuffler's lab that David Hubel developed a new technique using microelectrodes that allowed for extracellular recordings of single cells in various cortical areas (Hubel, 1959). Hubel then collaborated with Torsten Wiesel to pioneer work on single-cell recordings in the visual areas of cats (Hubel and Wiesel, 1959). They discovered that cells in the visual cortex responded vigorously to lines and edges instead of the spots and circles to which retinal ganglion cells respond preferentially (Wiesel, 1960). Hubel and Wiesel applied this new method to measure the firing rate of single cells in response to a range of visual stimuli, in cat striate (Hubel and Wiesel, 1962) and extra-striate visual areas (Hubel and Wiesel, 1965). They identified two types of cells: "simple" cells that fire rapidly when presented with edges and gratings of a particular orientation in a specific location within the cell's receptive field, and "complex" cells that respond to edges and gratings of a particular orientation anywhere in within their receptive field.

At around the same time, Horace Barlow began suggesting why certain sensory information was transmitted to formulate a set of communication principles (Barlow, 1961). Earlier, Barlow had investigated single cell responses in the frog's retina (Barlow, 1953). It was from these experiments, and from Hubel and Wiesel's work, that Barlow began to question not just what information was being computed at each neuron (the calculations), but what information was being transmitted neuron to neuron (the connections or relays). Barlow wanted to identify what information was deemed important to pass on and why some information was propagated and other information was not. Barlow proposed 3 hypotheses behind the transmission of sensory information. His first hypothesis was that neurons transmitted key information that led directly to certain behavioural responses and that irrelevant signals were

rejected (the “Password Hypothesis”). His second hypothesis was that relays regulated or controlled the flow of sensory information (the “Controlled Pass-Characteristic Hypothesis”). This would prevent the system being bombarded with excess information and becoming overloaded. His third hypothesis was that connections recoded sensory messages to reduce redundancies (the “Redundancy-Reducing Hypothesis”). Barlow phrased this final hypothesis as “recoding to reduce the redundancy of our internal representation of the outer world”. In other words, the signals transmitted by neurons were either aimed at being important (communicating the most influential information), transmitting select information (reducing overload in the system) or at economising sensory information (removing repetitive signals).

During this period in Kuffler’s lab, Barlow began to take a subtly divergent approach to researching the visual system from that of Hubel and Wiesel. Hubel and Wiesel were predominantly interested in investigating single cell responses at each stage of the visual hierarchy and mapping out these stimulus-response pairings. The work of Hubel and Wiesel demonstrated that certain aspects of the visual system were essentially deterministic, showing that a certain input to a cell would produce a well-defined output. In contrast, Barlow was interested in the motivating principles behind why such stimulus-response pairings took place. He turned to information theory as a means to quantify and better understand the underlying communication principles between neurons. Barlow formulated his Redundancy-Reducing Hypothesis (Barlow, 1961) using the formal definition of “information” - a quantitative measure dependant upon the prior probability of previous messages (Shannon, 1948). A message that is completely noisy will not carry any information. Also, a message that could already be predicted based on prior probabilities would carry no additional information. Using Shannon’s principle, it becomes clear that improbable events carry more information than probable ones (Prokopenko *et al.*, 2007). Turning to statistical approaches, Barlow focused on how information was being conveyed over an imperfect, noisy communication channel.

Rather than focusing on neural operations themselves, which appeared to be deterministic, he honed in on the probabilistic nature of neurons firing in a noisy environment.

This divergence in fundamental approaches continues to this day, with proponents of computational modelling showing either an inclination towards statistical modelling, or towards pre-defined filtering operations. That is not to say that these two approaches cannot exist together in a unified model, but that many researchers have a favoured approach for analysing computer vision problems, leaning more favourably towards a probabilistic or a deterministic approach. There are advantages and disadvantages in taking either approach. Deterministic models are easier to implement and test and do not require any learning periods, since kernel operations are hard-coded. On the other hand, probabilistic models provide greater sophistication but also require learning periods to extract prior probabilities for which to predict future events.

When modelling the visual system, computer scientists do not only need to decide whether it is more appropriate to take a probabilistic approach or a deterministic one. Modellers also need to consider the difficult question of where to define the beginning and end of the visual system. Is it necessary to go back as far as the retina and model functions representing rod and cone populations? Likewise, is it necessary to incorporate decision-making models to emulate the abilities of pre-frontal cortex? Many visual ventral stream models only emulate cortical areas V1 to IT and bypass retinal and LGN cell functions. In the following subsections we review pre-cortical (section 1.8) and cortical ventral stream (section 1.9) models of vision.

1.6.2 Biologically inspired systems

Many information technology researchers turn to biology for inspiration since it provides

living solutions to computational problems. Notably, in scenarios that involve distributed processing and/or decision-making, many biological systems provide exemplar cases. Biologically inspired systems, which are also referred to as bioinspired or biologically plausible, attempt to model the biological entity that they are emulating, rather than simply to reproduce the end result (Helms *et al.*, 2009; Flammang & Porter, 2011).

One point of contention in using the term bioinspired, is the level of fidelity to the biological counterpart being emulated (Flammang & Porter, 2011). Specifically, at what point does an artificial visual system become biologically plausible? If a researcher decides to implement artificial neurons, how many should be simulated? Do we need to simulate down to the level of ion interactions? Shall we simulate from V1 or earlier at the thalamus or the retina or even the physical wave properties of light? It is best to describe biological plausibility on a sliding scale, rather than just as a binary distinction, in order to encapsulate these varying levels. Biological plausibility may not just occur at a structural level, but at a system level or a temporal level or at a micro-level. To address the different levels of computation that are needed to model vision, we look to Marr's (1982) treatise mentioned in Section 1.5.

1.7 Computational models of vision

1.7.1 Deterministic versus probabilistic

Computational models that simulate vision are roughly divided into those that are deterministic and those that are probabilistic. Deterministic models produce the same result for every simulation run given a pre-defined input with pre-defined filters. After receiving an input image, these models then convolve the image with one or more kernels to produce a filtered image. Here the terms filter and kernel are used interchangeably. Probabilistic models,

on the other hand, are influenced by patterns embedded within large sets of natural images. Common properties such as contrast and spatial structure (Geisler, 2008) are extracted to produce a set of image statistics that represent the typical properties of stimuli that occur naturally. Statistics are either manually pre-specified or learnt before run-time, during which a stochastic element is involved. Once a set of statistical patterns is determined from a series of images, these patterns or templates are then applied to new images that the model is exposed to. Taking an image and analysing whether it conforms to a template removes any information that is not in line with the template. In this way, noise is removed from images since it does not conform to any underlying pattern. Following on from this, probabilistic models are generally better at handling noise (Srinivasan *et al.*, 1982).

Models that use learning are referred to as machine learners. Learning usually happens before run-time execution, but can also occur during run-time. Learning allows for connection strengths, referred to as weights, to be adjusted within an artificial neural network. In supervised learning, a machine learner is presented with labeled examples that it then uses to infer an underlying function that will be used to predict future examples. Each example input is provided with a desired output in supervised learning. In contrast, unsupervised learning is where there is no clear well-defined output for each input. In this case, a machine learner extracts patterns or structure from examples presented to it. Deterministic models, having a fixed input to output mapping, do not use any learning. Probabilistic models may recruit learning, although this is not necessary if probabilities are pre-defined. Probabilistic models may use supervised learning, unsupervised learning or a combination of the two.

It is worth noting that although some models can be easily classified as deterministic (e.g. Bertulis & Bulatov, 2001, 2005) or probabilistic (Howe and Purves, 2002; Howe and Purves, 2004; Corney & Lotto, 2007; Brown and Friston, 2012), there are some models that combine

both techniques (Dakin & Bex, 2003; Serre et al., 2007). We opt for a combined approach, where filters can be learnt from images that the model is trained on (Serre et al., 2007) or where image statistics are gleaned from filtered images instead of original source images (Dakin & Bex, 2003). In the coming sections we highlight a variety of deterministic and probabilistic approaches that are subdivided into pre-cortical and cortical models. We focus on the initial pre-cortical stages of image processing before addressing processing along the entire visual ventral stream. We divide these two categories of models, pre-cortical models and ventral stream models, into two separate sections of this thesis as outlined below. With each section, we highlight historical context, some of the models that have recently been applied to illusions and the modelling approach we take.

1.8 Pre-cortical models

1.8.1 Historical context

Early models of the retina and LGN were deterministic, designed to receive input as a map of stimulus intensities that would be transformed by a weighted sum (Kuffler, 1953; Rodieck, 1965). The weighted sum would allow for intensities falling in the ON regions to be positively weighted and those falling in the OFF regions to be negatively weighted, resembling the functionality of centre-surround cells (refer to the left side of Figure 1-5A). Any function that transforms an input to output using a reweighting scheme is referred to as filtering. Filtering the stimulus using both positive and negative weights results in an output containing negative values. For the output to correspond to firing rates, it needs to be converted to only positive values. Once responses have been filtered, where the result could include negative values, these were transformed into only positive values corresponding to firing rates. Linear rectification allows for a simple positive transformation, with values below

a threshold being output as zero and values above threshold being output as a linear function (Carandini & Ferster, 2000). Figure 1-5A illustrates this process (reproduced from Carandini, 2004).

In pre-cortical models, centre-surround receptive fields are usually modelled using a difference of Gaussians (DOG), first proposed by Rodieck (1965). DOG filters are based on two different sized Gaussians, where a Gaussian with smaller standard deviation is subtracted by a Gaussian with larger standard deviation. These 2D Gaussians are rotated about the mean, forming an isotropic filter. Marr and Hildreth (1980) put forward a Laplacian of Gaussians (LoG) as a more effective filter function for edge detection. LoG filters are also isotropic, showing no orientation selectivity. DOG is commonly used as an efficient approximation for LoG (Burger and Burge, 2013, p. 325).

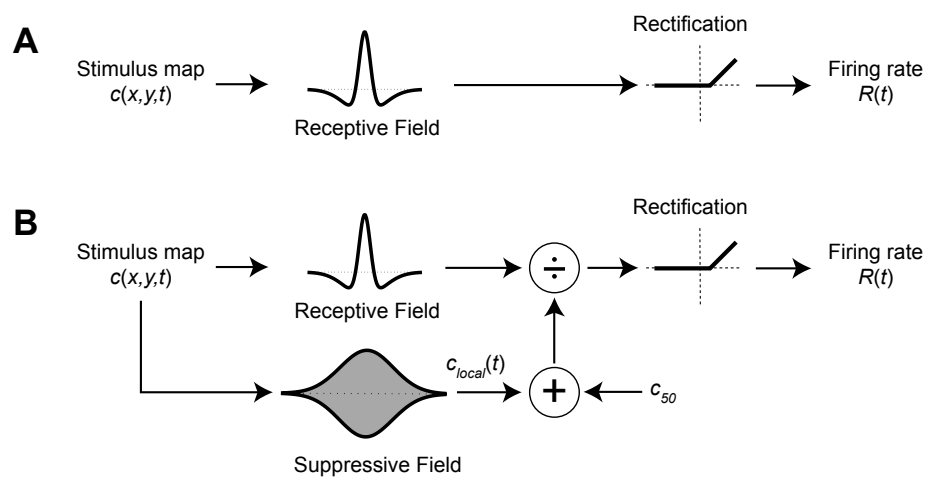


Figure 1-5: Models of the LGN, reproduced from Carandini (2004). A) Traditional model that maps stimulus $c(x,y,t)$ to firing rate using centre-surround filtering and linear rectification. B) Bonin et al.'s (2004) model incorporating a suppressive field with divisive gain control.

Carandini (2004) highlights shortcomings of the traditional filtering model presented in Figure 1-5A and advocates a revised model illustrated in Figure 1-5B (Bonin et al., 2004).

The revised model incorporates a suppressive field (in line with work from Levick *et al.*, 1972), providing a mechanism not just for driving inputs (increasing overall output) but also for modulating them (increasing output relative to other values). The receptive field (combining excitatory and inhibitory zones) and the suppressive field are incorporated together using divisive gain control (Freeman *et al.*, 2002; Solomon *et al.*, 2002). Gain is the rate at which response grows as a function of input magnitude. Divisive gain control modulates the output of the neuron by taking into account the range of signals output by the suppressive field. The model in Figure 1-5B that integrates receptive and suppressive fields provides an accurate model of LGN neurons for both the cat and monkey (Carandini, 2004). It is worth noting that LoG filters are not wholly representative of retinal processing and Daugman (1988) describes stimuli that are invisible to these filters, including texture discrimination, motion perception and pattern detection. The study we propose, which involves lightness illusions (Chapter 4), does not involve texture discrimination or any of the limitations listed by Daugman (1988). Wallis (2001) provides an in-depth comparison between LoG, DoG and Gaussian functions, and their relation to monkey physiology.

1.8.2 Current models applied to illusions

One of the most extensively investigated pre-cortical models applied to illusions is the DOG model (Blakeslee & McCourt, 1997). This deterministic model successfully demonstrated the Simultaneous Contrast Illusion (SCI), where a grey patch with a light surround appears darker than an otherwise identical grey patch with dark surround (Chevreul, 1839) (Figure 1-6). Subsequent versions of the DOG model required oriented filters to account for a larger variety of illusions (Blakeslee & McCourt, 1999, 2001, 2004; Blakeslee *et al.*, 2005). The introduction of an orientation component into DOG filters makes the model more closely matched to V1, and therefore it no longer strictly fits into our pre-cortical model category.

More recent versions of the model have removed the orientation component and reintroduced isotropic DOG filters (Cope *et al.*, 2013, 2014a, 2014b).

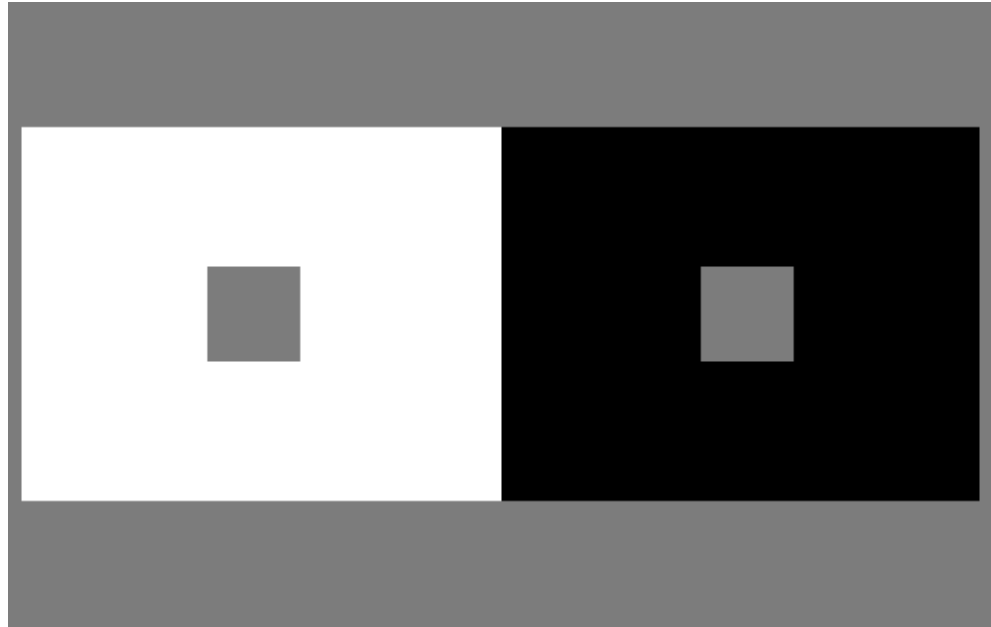


Figure 1-6: Simultaneous contrast illusion (SCI), The left grey patch surrounded by white appears darker than the right grey patch surrounded by black (Chevreul, 1839)



Figure 1-7: Craik, Cornsweet, O'Brien Illusion (CCOB) by Fibonacci (2007)

Probabilistic models have also been applied to low-level illusions (Williams *et al.*, 1998a, 1998b; Nundy & Purves, 2000; Purves and Lotto, 2003; Dakin and Bex, 2003), although the mapping of these models to pre-cortical areas of the brain is not intended or shown. Some probabilistic models however, do emphasise pre-cortical involvement in processing illusions. Dakin and Bex (2003) use an approach that combines filtering and image statistics. They apply a bank of LoG filters to a set of images to find common underlying distributions of filter amplitudes over a range of spatial frequencies. Dakin and Bex pay particular attention to the Craik–Cornsweet–O’Brien effect (CCOB), (O’Brien 1958; Craik 1966; Cornsweet 1970), illustrated in Figure 1-7. The CCOB illusion is where introducing an edge (a light-dark transition) between monochrome grey areas makes the region adjacent to the lighter luminance values of the edge appear lighter (right of Figure 1-7) and the region adjacent to darker values of the edge appear darker (left of Figure 1-7). After applying LoG filters to the CCOB, Dakin and Bex find that low spatial frequencies (SFs) are responsible for this effect. By boosting noise in low SFs, effectively eliminating the low SF structure, the illusion can be eradicated. Conversely, introducing noise in high SFs maintains the illusion. This result was tested and confirmed in human experiments. From these results, Dakin and Bex propose that restoring responses of a filter bank to those that we expect when viewing a natural image (through reweighting SFs), could be extended to account for other lightness illusions.

1.8.3 Our model selection: Exponential filter family model

Taking a similar approach to Dakin and Bex (2003), we implement a filter model that also takes into account contrast distributions found in natural images. We adopt an in-house model that uses a family of exponential filters described in Basu & Su (2001). These filters are designed to optimise the encoding of typical stimulus details as established using image statistics. Looking at contrast distributions over a large set of natural images, it is evident that

a similar underlying pattern is present, well described by a histogram with high kurtosis (having a sharp peak) and a heavy tail (Field, 1987; Simoncelli & Olshausen, 2001). The distribution that would best fit this description would be an exponential function (Zhu & Mumford, 1997). Zhu and Mumford put forward that Gaussian filters are not appropriate for extracting high spatial frequency information such as edges. Based on their analysis of images, they propose that exponential filters provide a better solution than Gaussian filters for preserving image structure. Taking an approach inspired by natural image statistics, we apply a set of different size and shape exponential filters to an image and then optionally normalise the result using divisive gain control (Carandini, 2004), which is described within this section above and illustrated in Figure 1-5. Our model follows the same process as that shown in Figure 1-5B, replacing the receptive field by the exponential function.

We emphasise that our exponential filter model is not intended to closely mimic biology. Our motivation is to take effective image filtering techniques that have been found in computer vision and use these to explore the effect that other types of filters may have on the processing of illusions. Nevertheless, it is possible to draw analogies between our model and neurobiology. To determine what brain areas or cell types are best represented by our exponential filter model, we identify some of the relationships between this model and neurobiology. We find that the model is best matched to pre-cortical visual areas in the following ways:

1. Exponential filters with high kurtosis have been identified in H1 horizontal retinal cells (Packer and Dacey, 2002, 2005).
2. Exponential filters with medium kurtosis form a Gaussian function. Gaussian differences (DOG) and Gaussian derivatives (LoG) are commonly used to model the LGN and retina (Kuffler, 1952; Kuffler, 1953).

3. From a functional standpoint, where the purpose of DOG and LoG filters is to extract edges from an image, Basu and Su (2001) find that exponential filters are well-suited for this purpose
4. Exponential filters are able to deal well with increasing amounts of noise (Basu and Su, 2001), which is believed to be an underlying principle behind the function of inhibition in the retina (Srinivasan, 1982)
5. The method of normalisation we use, divisive gain control, is shown to be present in LGN (Carandini, 2004).

1.9 Ventral stream models

1.9.1 General properties

The majority of visual ventral stream models are hierarchical in structure, with each layer corresponding to activations over increasing receptive field sizes with increasing generalisation across image features (Serre, 2014), (see Figure 1-8). Features are specific structures in an image such as points, edges or shapes. The presence or absence of a feature is determined by applying a filter to the image. For example, by applying a DOG filter, we can determine whether a circular feature of a stipulated size is present at a specified image location.

As the hierarchy is traversed, each layer stores a higher-level representation of the previous layer. These systems usually process information in a linear-non-linear chain of alternating cell types (Serre, 2014). Linear transformations represent processing by a bank of filters or a feature dictionary and non-linear functions pool information over multiple filters. Features at the lower levels are usually points and edges, compared to more complex mid and high-level features that encapsulate combinations of features, such as edge intersections, or object parts.

As information flows from the early retinal layers to higher neural layers such as the inferotemporal cortex (IT), representations generally become more compact (the number of neurons per layer decreases) and neurons respond with increasing invariance to stimulus location. These efficient representations reduce the number of connections required, increasing what is termed *sparsity*. Increasing sparsity helps to reduce the level of energy consumption needed for maintaining connections.

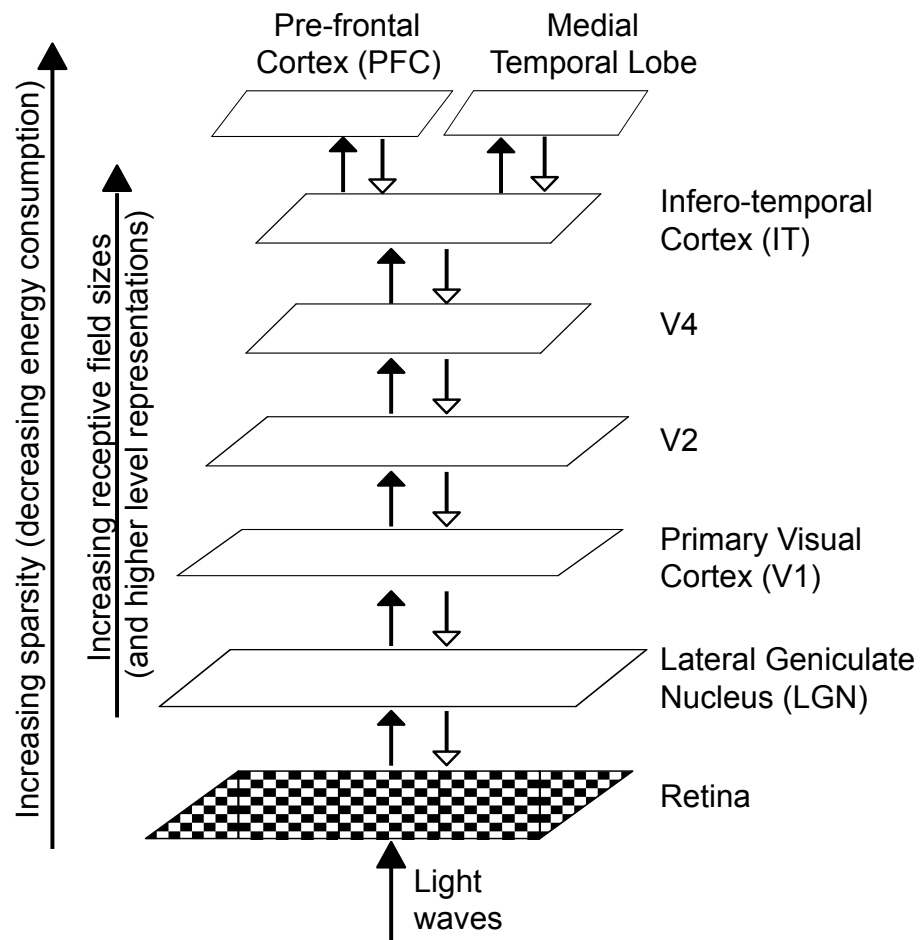


Figure 1-8: Layers of the visual system that form the ventral pathway. Black arrows represent feed-forward connections and white arrows represent feedback connections.

Connections not represented here are those made within layers and also between remotely-connected layers.

1.9.2 Hierarchical models in historical context

The Neocognitron is a multi-layered neural network consisting of feature-extracting S-cells and feature-combining C-cells. It was one of the first models of the ventral visual stream to include both “simple” (S) and “complex” (C) cells inspired by physiology (Fukushima, 1980). Fukushima (1988) designed the Neocognitron for recognising digital characters, successfully demonstrating resilience to changes in position, scale and deformation (Fukushima, 1988). Building on the simple-complex cell architecture from the Neocognitron, LeCun and colleagues later demonstrated the ability of a similar network structure to recognise handwritten digits (LeCun *et al.*, 1989a). This was successfully applied to postal-code recognition (LeCun *et al.*, 1989b) and digit recognition for handwritten cheques (LeCun *et al.*, 1997). LeCun’s (1989a, 1989b) model, called LeNet, has been recruited for a variety of applications speech and time-series prediction as well as image classification (LeCun and Bengio, 1995). Over the past couple of decades, a number of hierarchical models have been designed that employ a simple-complex architecture for image recognition (Wallis and Rolls, 1997; Riesenhuber and Poggio, 1999; Thorpe *et al.*, 2001; Ullman *et al.*, 2002; VanRullen, & Thorpe, 2002; Serre *et al.*, 2007; Masquelier and Thorpe, 2007). A thorough review of these models can be found in Serre (2014).

Most models to date have adopted a simple-complex cell architecture because of its excellent robustness to noise, translation, scaling and rotation. This common architecture can be roughly mapped to parts of the human brain involved in object recognition (see Figure 1-8), namely the ventral regions of the human visual processing hierarchy. This stream of processing begins at the retina, from whence information is propagated forwards (and backwards) through layers towards the pre-frontal cortex (PFC), where an object is identified. PFC is the area of the cerebral cortex covering the front part of the frontal lobe usually associated with decision making. Usually in computer models, areas V1 to IT are replaced

with stacked alternating layers of simple and complex artificial cells. A decision making algorithm is applied at the top-most layer to represent the role of PFC. Cortical models usually bypass modelling LGN and the on- and off-centre cells therein. Instead, they first perform filtering at the V1 level using, most commonly, Gabor filters. The model layers corresponding to visual processing modules from V1 to IT represent a “dictionary of features” (Mutch and Lowe, 2008) that are often drawn upon to distinguish object categories. The highest layer, representing PFC, is responsible for ultimately deciding how an object is recognised or categorised.

Models of vision are generally sub-divided into feed-forward and feedback models. Feed-forward models attempt to emulate the first 100-130ms of visual processing that has been shown to produce accurate object categorisation in monkeys (Hung *et al.*, 2005) and humans (Kirchner and Thorpe, 2006). One of the most widely researched and applied feed-forward models is HMAX, which we describe in detail below. We adopt HMAX as our model of choice among the available ventral stream models, predominantly because there is a large body of evidence showcasing its biological plausibility (Serre *et al.*, 2010). We also discuss feedback models to contrast with our choice of HMAX in order to highlight some of the advantages and disadvantages of these two different approaches.

1.9.3 Our model selection: feed-forward model HMAX

The aim of this model is to replicate the first feed-forward sweep of visual processing by implementing a feed-forward architecture consisting of simple and complex cell layers. Simple (S) cells fire with fine-tuned selectivity towards particular stimuli whereas complex (C) cells respond invariantly despite variance in the stimulus, such as spatial location (Serre *et al.*, 2005b). In other words, simple cells achieve discrimination in detecting features at

specific locations in the visual field. Complex cells allow for generalisation by pooling activations across simple cells.

The HMAX system has been extensively tested using a variety of paradigms, one of the first being a set of experiments inspired by Logothetis *et al.* (1995) where monkeys were trained to recognise novel paperclip objects. Logothetis and co-authors found cells in IT that were selectively responsive to different scaled, translated and rotated views of individual paperclips that were previously unknown. Similar experiments were performed using an early version of the HMAX model by Riesenhuber & Poggio (1999). The model was trained with specific views of paperclips and was later tested with both the same paperclips and other distractor objects. Artificial units within the model demonstrated a high response to the trained stimulus and a decreasing response as the view was rotated, scaled or translated. Over certain ranges of transformations, the response of the unit was greater to the preferred stimulus than to any of the distractors. Response properties of the artificial cell matched those from single cell recordings of the monkey after it was trained on the paperclip task.

Later versions of the model were tested in a rapid animal versus non-animal recognition task (Serre, 2007). The results for the model are directly comparable to those for human observers performing the same recognition task using a brief image display followed by a mask to limit feedback. The model performed at an accuracy level similar to human subjects when the delay between stimulus and mask was around 50ms, indicating that the HMAX model provides a satisfactory description of the feed-forward path of visual object recognition in humans.

The model has been further extended in a number of ways to account for other visual phenomena. An extensive list of model capabilities is listed in (Serre *et al.*, 2007). The HMAX model is able to demonstrate that object recognition is largely feed-forward in the

first 100-200ms (Serre *et al.*, 2007). The results of the system are comparable with biological and human psychophysical data (Serre and Poggio, 2010), allowing for direct comparison to the physical, algorithmic and computational levels proposed by Marr (1982). These multi-level comparisons give greater evidence to the plausibility of the model.

HMAX, however, does show some limitations. Being a model that is restricted to feed-forward processing, it does not include any of the back-projections found in visual cortex. Without feedback, this model cannot account for visual phenomena involving top-down attentional mechanisms. Modelling feedback would be necessary if we wish to model perception within the visual ventral stream beyond the first 100-200ms.

1.9.4 Feedback models

Feedback models are those in which information flows in two directions, where signals traverse upwards along the hierarchy and back down it. From the suite of possible feedback models, probabilistic generative models have recently gained momentum as increasingly plausible models of perception (Friston, 2010; Bastos *et al.*, 2012; Clark, 2013). These models are referred to under a variety of names, including Bayesian inference models, free energy networks, deep Boltzmann machines and deep sparse networks. Their underlying architecture is hierarchical and focuses heavily on feedback. Using Bayesian inference, higher levels predict patterns that emerge from lower levels, where differences between expectations and sensory readings are minimised from the top down. The gap between bottom-up signals and top-down predictions is called the error term. Clark (2013) provides an overview on the evolution of these architectures.

Historically, training very deep (multi-level) feedback networks was a difficult problem until

Hinton *et al.* (2006) and Hinton (2007) demonstrated a technique for minimising error at lower levels of the architecture. Compared to feed-forward networks, feedback networks generally take a long time to train. However, one of the biggest advantages in using probabilistic generative models is that all features at multiple levels are learnt in an unsupervised manner, emerging from patterns in the sensory input. Because features are learnt, input to the system can be in any form – sound, touch, smell, temperature, etc. This property of probabilistic generative models lends them to a wide variety of applications, from acoustic speech recognition (Mohamed *et al.*, 2013) to handwritten digit recognition (Ciresan *et al.*, 2010).

Probabilistic generative models are still relatively recent and have had much less comparison with their neurophysiological counterparts, in contrast to feed-forward models that more clearly map to neurophysiology. Although Friston (2012) has shown that probabilistic models are theoretically possible, there has been no neurological evidence to demonstrate this. Probabilistic generative models have been applied to visual illusions with success (Brown and Friston, 2012; Brown *et al.*, 2013). For the purposes of this thesis, we limit computational modelling to feed-forward models in order to explore the extent to which they alone can emulate illusory effects. By modelling illusions in feed-forward networks we can clearly separate explanations that rely on feedback mechanisms from those that are only feed-forward driven.

1.10 Existing computational models of visual illusions

1.10.1 How to model illusions

We introduced illusions as a sensed discrepancy that differs from the physical measurement of the stimulus source (Section 1.2). The question then arises: if we use computer model to

emulate an illusion, how is it possible to measure this sensed discrepancy? Furthermore, if we replicate a systematic bias in a computer model, how do we know this is an illusion versus a mistake in the machine learner? We consider each of these questions in turn.

Firstly, we address how it may be possible to measure an illusion within a computer model. In order to do this, we turn to our definition of illusions and reflect on what information can be measured in a computer model versus what information is measured externally to the model. Our previous definition of illusions was as a discrepancy between what is sensed and the physical stimulus. Taking this definition, it is possible to demonstrate an illusion as being a mismatch between external measurements of the stimulus and representations of that stimulus inside a model. The representation taken from within the computer model would be at the final layer or from previous layers.

Secondly, how do we demonstrate that an illusion, and not a general error, is being replicated within a computer model? Is it possible to classify any error as an “illusion” in a model? In order to establish this, we propose a method to determine whether the classifier is able to perform accurately using features from the model before exposing it to possible biases. This process contains two necessary steps: 1. Propose a control task, demonstrating accurate performance; 2. Demonstrate a pattern of errors in response to specific illusory stimuli that is reproducible and not random. By demonstrating that the machine learner is first capable of handling a control task, for which no bias is present, then the programmer can be confident that the algorithm is capable of handling the input and producing a valid result. It is worth noting that a model contains many parameters that can be adjusted to ensure correct performance. The control task is also useful for this purpose, in assuring that the parameter settings of the model are correct.

1.11 Scope of this thesis

This thesis looks at two different types of illusion in two separate feed-forward models. We first focus on emulating illusory bias and precision in a benchmark model of the visual system, demonstrating that models of visual cortex are susceptible to illusions and can be used to further inform our understanding of the causes of illusions. In order to cover the full gamut of modelling information from the retina, the second model emulates pre-cortical neural operations. This is to emphasize the important influence that early visual processing has on our perception of certain illusory images, which is bypassed in the first model which only simulates neurons from V1 onwards.

We select illusions that are static (no motion), greyscale (no color), monocular (occur even when viewing with only one eye) and are known to be mediated by processes in the visual ventral pathway, starting from the retina. We deliberately select illusions that are subject to ongoing debate about the role of low level versus high level mechanisms, or about the influence of bottom-up versus top-down information. By purposely selecting feed-forward models to simulate these illusions, we are able to determine some of the requirements for bringing about a particular illusion and separate low level from high level influences.

The first model we select is HMAX, and within HMAX, we choose to model the Müller-Lyer illusion, for which there is plentiful debate regarding its causes being statistically-driven (Gregory, 1963; Howe and Purves, 2005b) or filter-driven (Müller-Lyer, 1889; Coren, 1970). In this architecture, there is no modelling of lower-level areas such as the retina or LGN. Because of this, the second model in this thesis focuses on pre-cortical areas that feed into the visual ventral stream. Lightness illusions are proposed to be mediated by processing within very early visual areas (Blakeslee and McCourt, 1999, 2001, 2004), although there are theories that suggest higher-level influences impact on our lightness perception (Gilchrist,

1977; Knill and Kersten, 1991; Anderson, 1997). This second model simulates a range of lightness illusions that have been previously tested in other filtering models of early visual processing (Blakeslee & McCourt, 2004; Robinson et. al, 2007) to show whether high level visual areas are necessary in bringing about a range of lightness effects.

1.12 Thesis layout

This thesis consists of five chapters. The first introductory chapter includes a literature review of illusions, computational models of the visual system and how these subjects are combined together to bring about further understanding of the theories behind particular illusions. Chapters 2, 3 and 4 consist of 3 extensive published or submitted studies. The second chapter demonstrates the Müller-Lyer illusion in HMAX, a state-of-the-art model of visual cortex, ruling out the necessity of some of the proposed explanations behind the illusion. The third chapter delves deeper into HMAX simulations of the Müller-Lyer, quantifying bias and uncertainty layer-to-layer and identifying key mechanisms that bring about the effect. The fourth chapter looks at a pre-cortical model of vision and emulates a suite of lightness illusions in this model. The fifth and final chapter provides a discussion on how computational modelling, and in particular the experiments included in this thesis, have provided a deeper understanding of the necessary factors that drive visual illusions.

1.13 References

- Alivisatos, A. P., Chun, M., Church, G. M., Greenspan, R. J., Roukes, M. L., & Yuste, R. (2012). The brain activity map project and the challenge of functional connectomics. *Neuron*, 74(6), 970 – 974.
- Anderson, B. L. (1997). A theory of illusory lightness and transparency in monocular and binocular images: The role of contour junctions, *Perception*, 26(4), 419–454.
- Barlow, H. B. (1953). Summation and inhibition in the frog's retina. *J Physiol*, 119(1), 69-88.
- Barlow, H. B. (1961). Possible principles underlying the transformation of sensory messages. W.A. Rosenblith, (ed.), *Sensory communication*, MIT Press, Cambridge, MA, 217–234.

- Barlow, H. B. (1972). Single units and sensation: A neuron doctrine for perceptual psychology? *Perception*, 1(4), 371 – 394.
- Barlow, H. B., Kaushal, T. P. & Mitchison, G. J. (1989). Finding minimum entropy codes. *Neural Computation*, 1, 412-423.
- Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P., & Friston, K. J. (2012). Canonical microcircuits for predictive coding. *Neuron*, 76(4), 695-711.
- Basu, M. and Su, M. (2001). Image smoothing with exponential functions, *International Journal of Pattern Recognition and Artificial Intelligence*, 14(4), 735–752.
- Bertulis, A., & Bulatov, A. (2001). Distortions of length perception in human vision. *Biomedicine (Lithuania)*, 1(1), 3 - 25.
- Bertulis, A. & Bulatov, A. (2005). Distortions in length perception: Visual field anisotropy and geometrical illusions. *Neuroscience and Behavioral Physiology*, 35(4), 423-434.
- Blakeslee, B. and McCourt, M. E. (1997). Similar mechanisms underlie simultaneous brightness contrast and grating induction. *Vision Research*, 37(20), 2849–2869.
- Blakeslee, B. and McCourt, M. E. (1999). A multiscale spatial filtering account of the White effect, simultaneous brightness contrast and grating induction. *Vision Research*, 39, 4361–4377.
- Blakeslee, B. and McCourt, M. E. (2001). A multiscale spatial filtering account of the Wertheimer-Benary effect and the corrugated Mondrian. *Vision Research*, 41(19), 2487–2502.
- Blakeslee, B. and McCourt, M. E. (2004). A unified theory of brightness contrast and assimilation incorporating oriented multi-scale spatial filtering and contrast normalization. *Vision Research*, 44(21), 2483–2503.
- Blakeslee, B., Pasieka, W., and McCourt, M. E. (2005). Oriented multiscale spatial filtering and contrast normalization: a parsimonious model of brightness induction in a continuum of stimuli including White, Howe and simultaneous brightness contrast. *Vision Research*, 45(5), 607–615.
- Bonin, V., Mante, V., & Carandini, M. (2004). Nonlinear processing in LGN neurons. In: *Advances in Neural Information Processing Systems*, 16. S. Thrun, L. Saul, and B Schölkopf, (eds.), MIT Press.
- Box, G. E. P., & Draper, N. R. (1987). *Empirical Model Building and Response Surfaces*, John Wiley & Sons, New York, NY.

- Brown, H., & Friston, K. (2012). Free energy and illusions: the Cornsweet Effect. *Frontiers in Psychology*, 3(43).
- Brown, H., Adams, R. A., Parees, I., Edwards, M., & Friston, K.. (2013). Active inference, sensory attenuation and illusions. *Cognitive Processing*, 14(4), 411-427.
- Burger, W. and Burge, M. J. (2013). Principles of Digital Image Processing – Advanced Methods. Springer Undergraduate Topics in Computer Science, New York.
- Carandini, M., and Ferster, D. (2000). Membrane potential and firing rate in cat primary visual cortex. *Journal of Neuroscience*, 20, 470-484.
- Carandini, M. (2004). Receptive Fields and Suppressive Fields in the Early Visual System. In: *The Cognitive Neurosciences*. Gazzaniga, M. S., (ed.), MIT Press, 313 – 326.
- Carrasco M, Figueroa JG, Willen JD (1986) A test of the spatial-frequency explanation of the Müller-Lyer illusion. *Perception*, 15, 553–562.
- Changizi, M.A., Hsieh, A., Nijhawan R., Kanai R., Shimojo S. (2008). Perceiving the Present and a Systematization of Illusions, *Cognitive Science*, 32, 459–503.
- Chevreul, M. E. (1839). De la loi du contraste simultané des couleurs et de l’assortiment des objets colorés. Translated into English by C. Martel as *The principles of harmony and contrast of colours*. (English Second Edition: Longman, Brown, Green and Longmans (1855)).
- Ciresan, D., Meier, U., Gambardella, L. and Schmidhuber, J. (2010). Deep, big, simple neural nets for handwritten digit recognition. *Neural computation*, 22(12), 3207–3220.
- Clark A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science *Behavioral & Brain Sciences*, 36(3),181-204.
- Cope, D., Blakeslee, B. & McCourt, M. E. (2013). Analysis of multidimensional difference-of-Gaussians filters in terms of directly observable parameters. *Journal of the Optical Society of America A*, 30(5), 1002-1012.
- Cope, D., Blakeslee, B., and McCourt, M. E. (2014a). Modeling lateral geniculate nucleus response with contrast gain control. Part 1: formulation. *Journal of the Optical Society of America A*, 30(11), 2401-2408.
- Cope, D., Blakeslee, B., and McCourt, M. E. (2014b). Modeling lateral geniculate nucleus response with contrast gain control. Part 2: analysis. *Journal of the Optical Society of America A*, 31(2), 348-362.

Coren S. (1970). Lateral inhibition and geometric illusions. *The Quarterly Journal of Experimental Psychology*, 22, 274–278.

Coren S., Girgus J.S., & Day, R. H. (1973). Visual Spatial Illusions: Many Explanations. *Science*, 179:4072, 503-504.

Cornsweet, T. N. (1970). *Visual perception*. New York: Academic.

Craik, K. J. W. (1966). *The nature of psychology: a selection of papers, essays and other writings by the late K. J.W. Craik*. Cambridge University Press.

Dakin, S. C. & Bex, P. J. (2003) Natural Image Statistics Explain Brightness "Filling-In". *Proceedings of the Royal Society of London, Biological Sciences*, 270(1531), 2341-2348.

Daugman, J. G. (1988) Pattern and motion vision without Laplacian zero crossings. *Journal of the Optical Society of America A* 5(7), 1142– 1148.

Edelman, S. (1998). Representation is Representation of Similarities, *Behavioral and Brain Sciences* 21, 449-498.

Fibonacci. (2007). Cornsweet illusion - Own work. Licensed under CC BY-SA 3.0 via Wikimedia Commons, Last accessed: 18th February 2015.
http://commons.wikimedia.org/wiki/File:Cornsweet_illusion.svg#mediaviewer/File:Cornsweet_illusion.svg

Field, D. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America A*, 4(12), 2379-2394.

Flammang, B. E. & Porter, M. E. (2011). Bioinspiration: Applying Mechanical Design to Experimental Biology. *Integrative and Comparative Biology*, 51(1), 128–132.

Freeman, T. C. B., Durand, S., Kiper, D. C., and Carandini, M. (2002). Suppression without inhibition in visual cortex. *Neuron*, 35, 759-771.

Friston, K. (2005) A theory of cortical responses. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 360(1456):815–36.

Friston K. (2010). The free-energy principle: a unified brain theory? *Nat Rev Neuroscience*, 11(2), 127-38.

Friston K. (2012). A Free Energy Principle for Biological Systems. *Entropy*, 14, 2100-2121.
 doi:10.3390/e14112100.

- Fukushima, K. (1980). Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, 36(4), 193–202.
- Fukushima, K. (1988). Neocognitron: A hierarchical neural network capable of visual pattern recognition. *Neural Networks*, 1(2), 119–130.
- Geisler, W. S. (2008). Visual Perception and the Statistical Properties of Natural Scenes. *Annual Review of Psychology*, 59, 167–192.
- Gilchrist, A. L. (1977). Perceived lightness depends on perceived spatial arrangement, *Science*, 195(4274), 185–187.
- Gilchrist, A. (2003) The importance of errors in perception. In: *Colour Perception: Mind and the Physical World*. R. Mausfield & D. Heyer, (eds.), Oxford, Oxford University Press, 437-452.
- Gilchrist, A. (2014). A gestalt account of lightness illusions. *Perception*, 43(9), 881 – 895.
- Gregory, R. L. (1963). Distortion of visual space as inappropriate constancy scaling. *Nature*, 199, 678–680.
- Gregory R. L. (1966). *Eye and Brain: the psychology of seeing*. London: Weidenfeld & Nicolson; 5th edition 1997, Oxford University Press/Princeton University Press.
- Gregory, R. L. (1970). *The Intelligent Eye*. London: Weidenfeld and Nicolson.
- Gregory, R. L. (1997). Knowledge in perception and illusion, *Phil. Trans. R. Soc. Lond. B*, 352, 1121–1128
- Gregory, R. L. (2005). The Medawar Lecture 2001: Knowledge for Vision: Vision for Knowledge. *Philosophical Transactions: Biological Sciences*, 360(1458), 1231-1251.
- Helmholtz, Hermann von, *Treatise on physiological optics*. New York: Dover, 1962. (English translation by J. P. C. Southall from the 3rd German edition of *Handbuch der physiologischen Optik*. Hamburg: Voss. First published in 1867, Leipzig: Voss).
- Helms, M., Vattam, S. S., & Goel, A. K. (2009). Biologically inspired design: process and products. *Design Studies*, 30(5), 606-622.
- Heymans, G. (1896). Untersuchungen über das “optischen Paradoxen”. *Zeitschrift für Psychologie und Physiologie der Sinnesorgane*, 11, 66-67.
- Hill, H., & Bruce, V. (1993). Independent effects of lighting, orientation, and stereopsis on the hollow-face illusion. *Perception*, 22(8), 887 – 897.

- Hill, H., & Bruce, V. (1994). A comparison between the hollow-face and 'hollow-potato' illusions. *Perception*, 23(11), 1335 – 1337.
- Hill, H. & Johnston A. (2007). The hollow-face illusion: object-specific knowledge, general assumptions or properties of the stimulus? *Perception*, 36(2), 199 - 223.
- Hinton, G. E. & Zemel, R. S. (1994). Autoencoders, minimum description length and Helmholtz free energy. In: *Advances in neural information processing systems 6*. J. Cowan, G. Tesauro & J. Alspector (eds.), Morgan Kaufmann, 3-10.
- Hodgkin, A. L. & Huxley, A. F. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *The Journal of Physiology*, 117(4), 500–544.
- Hubel, D. H. (1959). Single unit activity in striate cortex of unrestrained cats. *The Journal of Physiology*, 147, 226-238.
- Hubel, D. H., & Wiesel, T. N. (1959). Receptive fields of single neurones in the cat's striate cortex. *Journal of Physiology*, 148, 574– 591.
- Hubel, D. H., Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology*, 160, 106-154.
- Hubel, D. H., Wiesel, T. N. (1965). Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat. *Journal of Neurophysiology*, 28, 229-289.
- Hung, C. P., Kreiman, G., Poggio, T., & DiCarlo, J. J. (2005). Fast readout of object identity from macaque inferior temporal cortex. *Science*, 310(5749), 863–866.
- Howe, C. Q. and Purves, D. (2002). Range image statistics can explain the anomalous perception of length. *Proceedings of the National Academy of Sciences*, 99(20), 13184-13188.
- Howe, C. Q. and Purves, D. (2004). Size Contrast and Assimilation Explained by the Statistics of Natural Scene Geometry. *Journal of Cognitive Neuroscience*, 16(1), 90-102.
- Howe, C.Q. and Purves, D. (2005a). Perceiving Geometry: Geometrical Illusions Explained by Natural Scene Statistics. New York: Springer.
- Howe C. Q., Purves D. (2005b). The Müller-Lyer illusion explained by the statistics of image–source relationships. *Proceedings of the National Academy of Sciences*, 102(4), 1234–1239.

- Kanisza, G. (1976). Subjective contours. *Scientific American*, 234, 48-52.
- Kersten, D., Mamassian, P., & Yuille, A. (2004). Object perception as Bayesian inference. *Annual Review of Psychology*, 55, 271-304.
- Kirchner, H., & Thorpe, S. J. (2006). Ultra-rapid object detection with saccadic eye movements: Visual processing speed revisited. *Vision Research*, 46(11), 1762– 1776.
- Kitaoka, A. (2003). “Rotating snakes.” <http://www.psy.ritsumei.ac.jp/~akitaoka/rotsnakes13e.html>.
- Knill, D. C. and Kersten, D. (1991). Apparent surface curvature affects lightness perception. *Nature*, 351(6323), 228-30.
- Kuffler S. W. (1952). Neurons in the retina; organization, inhibition and excitation problems. *Cold Spring Harbor Symposia on Quantitative Biology*, 17, 281-292.
- Kuffler, S. W. (1953). Discharge patterns and functional organization of mammalian retina. *J. Neurophysiology* 16, 37-68.
- Kuffler S. W. (1973). The single-cell approach in the visual system and the study of receptive fields. *Investigative Ophthalmology*, 12(11), 794-813.
- LeCun, Y., Jackel, L. D., Boser, B., Denker, J. S., Graf, H. P., Guyon, I., Henderson, D., Howard, R. E. & Hubbard, W. (1989a). Handwritten digit recognition: Applications of neural network chips and automatic learning. *IEEE Communications Magazine*, 27(11), 41 – 46.
- LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W. & Jackel, L. D. (1989b). Backpropagation applied to handwritten zip code recognition. *Neural Computation* 1, 1, 541–551.
- Y. LeCun and Y. Bengio. (1995). Convolutional networks for images, speech, and time-series. In M. A. Arbib, (Ed.), *The Handbook of Brain Theory and Neural Networks*. MIT Press.
- Y. LeCun, L. Bottou, and Y. Bengio. (1997). Reading checks with graph transformer networks. In: *International Conference on Acoustics, Speech, and Signal Processing*, 1, 151-154, IEEE.
- Lee, T. S. & Mumford, D. (2003) Hierarchical Bayesian inference in the visual cortex. *Journal of Optical Society of America, A*, 20(7), 1434 – 48.
- Lewis, E. O. (1909) Confluxion and contrast effects in the Müller-Lyer illusion. *British Journal of Psychology*, 3(1-2), 21-41.

- Levick, W. R., Cleland, B. G., and Dubin, M. W. (1972). Lateral geniculate neurons of cat: retinal inputs and physiology. *Investigative Ophthalmology*, 11, 302-311.
- Logothetis, N.K., Pauls, J. & Poggio, T. (1995). Shape representation in the inferior temporal cortex of monkeys. *Current Biology*, 5, 552–563.
- Low-Décarie, E., Chivers, C. and Granado, M. (2014). Rising complexity and falling explanatory power in ecology. *Frontiers in Ecology and the Environment*, 12, 412–418.
- Markram, H. (2006). The Blue Brain Project. *Nature Reviews Neuroscience*, 7, 153 – 160.
- Markram, H., Meier, K., Lippert, T., Grillner, S., Frackowiak, R., Dehaene, S., Knoll, A., Sompolinsky, H., Verstreken, K., DeFelipe, J., Grant, S., Changeux, J-P., & Saria, A. (2011). Introducing the human brain project. *Procedia Computer Science*, 7, 39-42.
- Marr, D. and Poggio, T. (1977). From understanding computation to understanding neural circuitry. In: *Neuronal Mechanisms in Visual Perception, Neurosciences Res. Prog. Bull.*, 15(3). E. Poppel, R. Held and J.E. Dowling, (eds.), 470-488.
- Marr, D. & Hildreth, E. (1980). Theory of edge detection. *The Royal Society of London B*, 207, 187 – 217.
- Marr, D. (1982). *Vision*. San Francisco, CA: Freeman.
- Masquelier T., & Thorpe, S. J. (2007). Unsupervised learning of visual features through spike timing dependent plasticity. *PLoS Computational Biology*, 3(2): e31. doi: 10.1371/journal.pcbi.0030031
- Meeter, M., Jehee, J. F. M & Murre, J. M. J. (2007). Neural models that convince: Model hierarchies and other strategies to bridge the gap between behavior and the brain. *Philosophical Psychology*, 20(6), 749 – 772.
- Mohamed, A., Dahl, G., and Hinton, G. (2012). Acoustic modeling using deep belief networks. *IEEE Transactions on Audio, Speech, and Language Processing*, 20(1), 14–22.
- Müller-Lyer, F. C. (1889). Optische Urteilstäuschungen. *Archiv für Anatomie und Physiologie*, 2, 263–270.
- Mumford, D. (1994) Neuronal architectures for pattern-theoretic problems. In: *Large-scale theories of the cortex*. C. Koch & J. Davis, (eds.), pp. 125–52. MIT Press.
- Mutch J, Lowe DG (2008) Object class recognition and localization using sparse features with limited receptive fields. *International Journal of Computer Vision* 80: 45–57.
- Necker, L. A. (1832). Observations on some remarkable optical phaenomena seen in Switzerland; and on an

optical phaenomenon which occurs on viewing a figure of a crystal or geometrical solid. *London and Edinburgh Philosophical Magazine and Journal of Science*, 1(5), 329–337. doi:10.1080/14786443208647909

Norris, D. (2005) How do computational models help us build better theories? Chapter 20. In A. Cutler, (Ed.) *Twenty-First Century Psycholinguistics: Four Cornerstones*.

Nundy, S., & Purves, D. (2000). A probabilistic explanation of brightness scaling. *Proceedings of the National Academy of Science*, 99, 14482–14487.

O'Brien, V. (1958). Contour perception, illusion and reality. *Journal of the Optical Society of America*, 48(2), 112–119.

Packer, O. S. and Dacey, D. M. (2002), Receptive field structure of H1 horizontal cells in macaque monkey retina, *Journal of Vision*, 2(4), 272–292.

Packer, O. S. and Dacey, D. M. (2005), Synergistic center-surround receptive field model of monkey H1 horizontal cells, *Journal of Vision*, 5(11), 1038–1054.

Penrose, L. S., & Penrose, R. (1958). Impossible objects: A special type of visual illusion. *British Journal of Psychology*, 49(1), 31–33.

Pieron, H. L. (1911). L'illusion de Müller-Lyer et son double mechanisme. *Revue Philosophique A*, 71, 245 - 284.

Poggio, T., & Bizzi, E. (2004). Generalization in vision and motor control. *Nature*, 431, 768-774.

Prokopenko, M., Boschetti, F., & Ryan, A. J. (2007). An Information-Theoretic Primer on Complexity, Self-Organization, and Emergence. *Complexity*, 15(1), 11-28.

Purves, D., Shimpf, A., & Lotto, R. B. (1999). An Empirical Explanation of the Cornsweet Effect. *The Journal of Neuroscience*, 19(19), 8542–8551.

Purves, D. & Lotto, R. B. (2010) (Second edition, November 5). *Why We See What We Do Redux*. Sinauer Associates, Inc. Publishers Sunderland, Massachusetts, U.S.A.

Rao and Ballard, 1999. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*. 1999. 2:79–87

Riesenhuber, M., and Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience*, 2, 1019–1025.

- Robinson, A. E., Hammon, P. S., and de Sa, V. R. (2007). Explaining brightness illusions using spatial filtering and local response normalization. *Vision Research*, 47(12), 1631–1644.
- Rodieck, R. W. (1965). Quantitative analysis of cat retina ganglion cell response to visual stimuli. *Vision Research*, 5, 583-601.
- Rogers, Brian, 2014, "Delusions about illusions" *Perception* 43(9) 840 – 845
- Serre, T., Wolf, L., and Poggio T. (2005a). Object recognition with features inspired by visual cortex. In: Proceedings of 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005). San Diego: IEEE Computer Society Press, 886–893.
- Serre, T., Kouh, M., Cadieu, C., Knoblich, U., Kreiman, G., and Tomaso Poggio. (2005b). A theory of object recognition: computations and circuits in the feedforward path of the ventral stream in primate visual cortex. Technical report, Massachusetts Institute of Technology, Cambridge, MA, December 2005.
- Serre, T., Oliva, A., & Poggio, T. (2007) A feedforward architecture accounts for rapid categorization. *Proceedings of the National Academy of Science*, 104(15), 6424–6429.
- Serre, T., & Poggio, T. (2010). A neuromorphic approach to computer vision. *Communications of the ACM (online)*, 53(10), October 2010.
- Serre, T., Wolf, L., & Poggio, T. (2005). Object recognition with features inspired by visual cortex. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005), San Diego. IEEE Computer Society Press.
- Serre, T. (2014). Hierarchical Models of the Visual System. In *Encyclopedia of Computational Neuroscience* (pp. 1-12). Springer New York.
- Shannon, C. E. (1948). A Mathematical Theory of Communication. *The Bell System Technical Journal*, 27, 379–423, 623–656.
- Simon, H. (2008). Alternative Views of Complexity. In: *Emergence: Contemporary readings in philosophy and science*. M. A. Bedau & P. Humphreys, (eds.), MIT Press, 249-258.
- Solomon, S. G., White, A. J., and Martin, P. R. (2002). Extraclassical receptive field properties of parvocellular, magnocellular, and koniocellular cells in the primate lateral geniculate nucleus. *Journal of Neuroscience*, 22, 338-349.
- Srinivasan, M. V., Laughlin, S. B. and Dubs, A. (1982). Predictive coding: a fresh view of inhibition in the retina. *Proceedings of the Royal Society of London B., Biological Sciences*, 216(1205), 427-59.

- Stokes, D. (2013). Cognitive Penetrability of Perception. *Philosophy Compass*, 8(7), 646-663.
- Tan H-R. M., Lana, L., & Uhlhaas P. J. (2013). High-frequency neural oscillations and visual processing deficits in schizophrenia. *Frontiers in Psychology*, 4:621.
- Thorpe, S., Delorme, A., & VanRullen, R. (2001). Spike-based strategies for rapid processing. *Neural Networks*, 14(6-7), 715-725.
- VanRullen, R. & Thorpe, S. J. (2002). Surfing a spike wave down the ventral stream. *Vision Research*, 42(23), 2593-2615.
- Ullman, S., Vidal-Naquet, M., and Sali, E. (2002). Visual features of intermediate complexity and their use in classification. *Nature Neuroscience*, 5(7), 682-687.
- Wallis, G. (2001). Linear models of simple cells: Correspondence to real cell responses and space spanning properties. *Spatial Vision*, 14(3), 237-260. doi:10.1163/156856801753253573
- Weiss, Y., Simoncelli, E. P., & Adelson, E. H. (2002). Motion illusions as optimal percepts. *Nature Neuroscience*, 5(6), 598-604.
- Williams, S. M., McCoy, A. N., & Purves, D. (1998a). The influence of depicted illumination on perceived brightness. *Proceedings of the National Academy of Sciences of the United States of America*, 95, 13296-13300.
- Williams, S. M., McCoy, A. N., & Purves, D. (1998b). An empirical explanation of brightness. *Proceedings of the National Academy of Sciences of the United States of America*, 95, 13301-13306.
- Zeman, A., Obst, O., Brooks, K. R., & Rich, A. N. (2013). The Müller-Lyer Illusion in a computational model of biological object recognition. *PLoS ONE*, 8(2), e56126. doi:10.1371/journal.pone.0056126
- Zeman, A., Obst, O., & Brooks, K. R. (2014). Complex cells decrease errors for the Müller-Lyer Illusion in a computational model of the visual ventral stream. *Frontiers in Computational Neuroscience*, 8, 112. doi:10.3389/fncom.2014.00112
- Zeman, A., Brooks, K. R & Ghebreab, S. (in submission). An exponential filter model predicts lightness illusions. *In submission*.
- Zhaoping, L. (2014). *Understanding Vision, Theory, Models, and Data*. Oxford University Press, United Kingdom.
- Zhu, S. C., and Mumford, D. B. (1997). Learning generic prior models for visual computation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*: June 17 - 19, San Juan. Puerto Rico, ed. IEEE Computer Society, 463-469. Los Alamitos, CA : IEEE Computer Society.

2 Study 1

The Müller-Lyer Illusion in a computational model of biological object recognition.

Abstract

Studying illusions provides insight into the way the brain processes information. The Müller-Lyer Illusion (MLI) is a classical geometrical illusion of size, in which perceived line length is decreased by arrowheads and increased by arrowtails. Many theories have been put forward to explain the MLI, such as misapplied size constancy scaling, the statistics of image-source relationships and the filtering properties of signal processing in primary visual areas. Artificial models of the ventral visual processing stream allow us to isolate factors hypothesised to cause the illusion and test how these affect classification performance. We trained a feed-forward feature hierarchical model, HMAX, to perform a dual category line length judgment task (short versus long) with over 90% accuracy. We then tested the system in its ability to judge relative line lengths for images in a control set versus images that induce the MLI in humans. Results from the computational model show an overall illusory effect similar to that experienced by human subjects. No natural images were used for training, implying that misapplied size constancy and image-source statistics are not necessary factors for generating the illusion. A post-hoc analysis of response weights within a representative trained network ruled out the possibility that the illusion is caused by a reliance on information at low spatial frequencies. Our results suggest that the MLI can be produced using only feed-forward, neurophysiological connections.

2.1 Introduction

Visual illusions have the potential to offer great insight into our visual perception. Illusions have been extensively studied by psychologists, as a method of deducing the assumptions that the brain makes and how we process visual information. One classical illusion known to induce misjudgement, is the Müller-Lyer Illusion (MLI). In the MLI, the perceived length of a line is affected by arrowheads or arrowtails placed at the ends of the line (Müller-Lyer, 1889). Specifically, the line appears elongated in the arrowtails and contracted with arrowheads (see

Figure 2-1A). Behavioural studies have shown that the strength of the illusion is correlated with factors including shaft length (Fellows, 1967; Brigell & Uhlarik, 1979), fin angle (Dewar, 1967) and inspection time (Coren & Porac, 1984; Predebon, 1997).

Although many theories have been put forward to explain the MLI (reviewed in Bertulis & Bulatov, 2001), there is ongoing debate as to the source of the MLI. Originally, the illusion was explained as a combination of two opposing factors: ‘confluxion’ and ‘contrast’ (Müller-Lyer, 1896a, 1896b). These terms were later interpreted into more modern concepts of lateral inhibition and contour repulsion (Coren, 1970). Higher weighting placed on low spatial frequency information has also been investigated as a possible contributing factor towards the MLI (Carrasco et al., 1986; Ginsburg, 1978). Explanations based on spatial filtering properties have been investigated further in computer models and have been found to produce a Müller-Lyer effect (Bertulis and Bulatov, 2001). It is possible that these mechanistic explanations may not provide a full explanation of the illusion and we may need to look beyond explanations that purely involve bottom-up neural computation. Gregory was the first researcher to suggest that the images in our environment could influence our perception of the MLI and introduced another type of explanation based on misapplied size constancy scaling. Size constancy scaling refers to our visual system’s ability to perceive an object as being of a constant size, even though changes in viewing distance change the size of its retinal image. To deduce the real-world size of an object, we take into account the perceived distance when scaling the retinal image size. When the depth of an image is misperceived, the scaled size judgement will also be erroneous. Gregory (1963) proposed that implicit depth cues in the arrowtails image imply that this object is more distant than the arrowheads image, such that their identical retinal sizes produce unequal perceived sizes.

Explaining the illusion has proven difficult because the effect persists even when the wings of the illusory figure are replaced with other terminating shapes, such as circles or squares

(Figure 2-1B). Even without the shaft (Figure 2-1), the perceptual effect remains. These variants demonstrate the persistence of line length misjudgement and rule out simple explanations for the cause of the illusion.

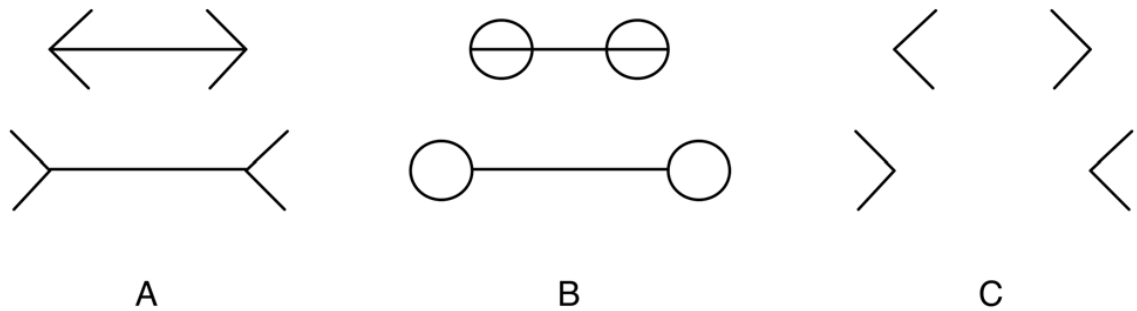


Figure 2-1 The ML illusion in various forms. A: The classical four-wing form illustrates the perceptual effect of the top line appearing shorter than the bottom line, even though the lines are of equal length. B: Terminating circles still induce a perceptual effect of line length misjudgment. C: The effect persists even when shafts are removed from the original figure.

Although there is disagreement on what causes the MLI, there is some consensus on where the illusion occurs in the brain. It is commonly accepted that visual information is processed via two pathways (Goodale & Milner, 1992): the ventral stream or ‘what’ pathway, which extends from striate cortex to infero-temporal lobe and the dorsal stream or ‘where’ pathway, which extends from occipital to parietal cortex. A recent fMRI study shows increased blood oxygen level-dependant signal strength in the Lateral Occipital Cortex (LOC) when participants viewed the MLI versus a control image (Weidner & Fink, 2007). An MEG study has demonstrated results consistent with the previous fMRI data, showing strong activation along the ventral visual pathway in lateral occipital areas and the inferior temporal cortex (Weidner *et al.*, 2010). Therefore, a number of studies support the proposal that the ventral stream plays a dominant role in processing the MLI. We hypothesised that as the MLI occurs

within the ventral stream of visual processing, then a model that imitates the structure and functionality of this region should be able to demonstrate this perceptual effect.

Currently, a number of biologically plausible image recognition models exist that computationally mimic visual cortex. To date, the majority of these have been concerned with correct object identification or classification. In this paper we apply these models to a task known to produce an illusion in human observers. Here, we seek to demonstrate a similarity to human perception, not simply by reproducing a poor level of overall performance, but further by producing a specific predictable pattern of errors. We highlight several advantages for researchers from different fields who adopt this novel approach of mimicking visual ‘errors’ in computational object recognition models. For perceptual psychologists, a model that imitates illusory perception would allow for the isolation and testing of factors thought to contribute to an illusion. Errors of perception have been extremely informative in demonstrating how the human brain works. Working with a computational model opens up possibilities for conducting experiments that are difficult, if not impossible, to do in humans. These types of experiments include parameter changes (such as the level of inhibition), the modification of learning stimuli and exploration of the effect caused by artificial lesions. For the computer scientist, classification that matches human error patterns increases the biological psychological plausibility of a model. Identifying illusions can enable computers to reject interpretations of the world that yield impossible objects or paradoxes. Classification experiments may also reveal elements of neural information processing that have yet to be uncovered and lead to improved object recognition and categorisation. Thus, we can use models to test explanations of well-studied geometrical illusions from a new perspective.

This paper outlines a set of experiments conducted in HMAX, a well-established, biologically plausible model of object recognition (Serre *et al.*, 2005). The main goal is to analyse performance of the model in judging relative line lengths for control stimuli versus Müller-

Lyer stimuli. Essentially, we want to assess whether a feed-forward object recognition model, with no exposure to natural images, can ‘perceive’ the MLI. We found a consistent pattern of errors that demonstrated a Müller-Lyer effect in HMAX after training on a non-natural set of images.

2.2 Methods

Our experiments required a model that was biologically plausible, in that it could be functionally mapped to the human visual ventral stream. A number of models currently exist which have been inspired by neurophysiology, pioneered by systems such as the Neocognitron (Fukushima, 1980) and convolutional networks (LeCun *et al.*, 1989, LeCun & Bengio, 1995). From these biologically plausible options, we selected the model that has demonstrated much evidence consistent with neurological and psychological data. The HMAX model, with features inspired by visual cortex (Serre *et al.*, 2005) has not only shown results congruent with psychological and neurological experiments, but it has also made correct predictions of biological phenomena (Serre & Poggio, 2010). We selected a version of the HMAX model that exclusively models the ventral visual stream and has successfully demonstrated multi-class categorisation (Mutch & Lowe, 2008).

The five-layer architecture setup is similar to that described in Mutch & Lowe (2008), where input to the network is fed through an image layer and then processing flows sequentially through the other four layers. These layers alternate in their primary functionality, dedicated to either template matching or convolution. The behaviour of these artificial cells is said to model the Simple (‘S’) and Complex (‘C’) neuronal functionality discovered by Hubel and Wiesel in cat striate cortex (Hubel & Wiesel, 1959). Simple cells demonstrate higher levels of activation in response to a specific, preferred stimulus, whereas Complex cells demonstrate invariance through high response levels across varied but related inputs. Figure 2-2 illustrates

the set of layers within the model which are described in further detail below.

2.2.1 HMAX Layer Descriptions

Image Layer. Input to the model is fed via the image layer, which receives a 256x256 pixel greyscale image. An image pyramid with 10 levels is constructed using bicubic interpolation, with each level $2^{1/4}$ smaller than the previous. We therefore have the image duplicated at scales of 215x215, 181x181, 152x152, 128x128, 108x108, 91x91, 76x76, 64x64 and 54x54 pixels. This forms a multi-scale representation of the input image.

S1 Layer (Gabor filter). Output from the image layer is received by the S1 layer, which employs Gabor filters at every position and scale. Twelve orientations are used for the Gabor filters which are 11x11 in size and are applied to all levels of the 4D pyramid, before the results are normalised.

C1 Layer (Local invariance using hard MAX). This layer pools the response of nearby S1 units to create position and scale invariance at a local level. The range of a C1 unit forms the shape of a pyramid spanning an area 10x10 units across the base with a height of 2 levels. The response R_C of a C unit is the maximum value of all S units X_l to X_n that fall within the filter range. This max filter achieves subsampling by moving around each S1 orientation pyramid in steps of five with an overlap of 2 positions and scales. The resultant C1 output is a convolved and compressed representation of S1 units. Note that the max function is not applied over different orientations, hence the C1 layer maintains a 4D pyramid structure.

S2 Layer (Learned intermediate features). This layer performs template matching at every position and scale in the C1 layer. A patch of C1 units centered at each position and scale is compared with a prototype patch d . These prototypes are randomly sampled from the C1 layers of the training images in the initial feature learning stage. After feature learning is

complete, each of these prototypes can now be seen as an additional convolution filter which is run over C1.

C2 Layer (Global invariance using hard MAX). This layer constructs a d - dimensional vector, where each element is the maximum response to one of the model's d prototype patches anywhere within the image. All orientation information is collapsed into one representation. At this stage of the model, all position and scale information has been removed, so it is now a 'bag of features'.

SVM Layer (Decision making module). Finally, classification of the image is performed using an all-pairs linear SVM. C2 vectors are normalised before being fed into the classifier. The majority-voting method is used to assign test images to categories.

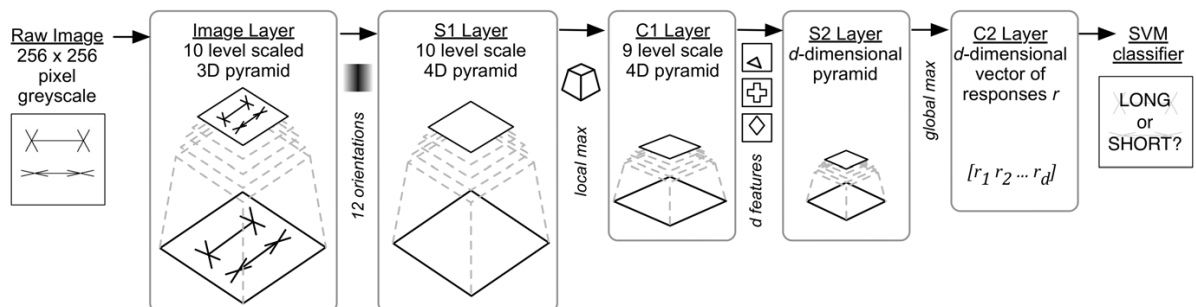


Figure 2-2 HMAX model architecture. Information flows unidirectionally through the hierarchical layers. Input to the system is a 256x256 greyscale image and the output is a classification of the image as LONG or SHORT. The input image is first transformed onto multiple scales via the Image Layer. The following four layers alternate in their functionality, dedicated to template matching (S layers) or feature pooling (C layers). The final SVM layer performs binary classification.

2.2.2 Task Description

The task in these experiments was to perform a two choice category task on a set of images. This task mimics a psychophysical yes-no length discrimination procedure. The classifier had to decide whether the top line in a given image was longer (L) or shorter (S) than the bottom line. Examples of images from each category are illustrated in Figure 2-3. All images fed into the model were 256x256 pixels in size, with black lines drawn using a 262 pixel pen on a white background. For the L condition, the top line had randomised line length between 120 and 240 pixels. For the S condition, the bottom line length was randomised also between 120 and 240 pixels. The comparator line length was randomised to be between 2 and 62 pixels shorter than the top (or bottom) line for the L (or S) condition. The vertical position of the top line was randomised between 48 and 108 pixels from the top of the image while bottom line's vertical position was randomly placed between 148 and 208 pixels. This forced the machine learner to rely on invariant properties rather than on absolute positional information for classification.

2.2.3 Experimental Setup

We ran each experiment in two stages: a training stage and a test stage. The model consisted of interleaved S and C layers, with a support vector machine (SVM) on top to perform final classification (see Methods Section for details). For the training period, we exposed the network to a set of 450 images to learn features at different positions and scales. Features were only learnt in the S2 and C2 cell layers; S1 and C1 have a fixed set of features (refer to Methods Section). Once the C2 vectors were built for the training set, the SVM was trained to perform the L/S classification task. For the test phase, C2 vectors were built for the test set of images which were then classified using the SVM.

Cross Fin (XF) images (Figure 2-3 Column 1) were used for training, since they contain

features present in both control and test stimuli and they do not induce any illusory effects. Fin lengths were randomised between 15 and 40 pixels (measured from the end of the shaft to the tip of the fin). Fin angles were randomised between 10 and 90 degrees for both top and bottom lines. This was to prevent the classifier from relying on the end positions of fins or on bounding box information to make a length judgment. Essentially, we wanted to confirm that the machine learner was making its decision based only on the length of the inside lines (shafts) while also allowing it to be exposed to other irrelevant features.

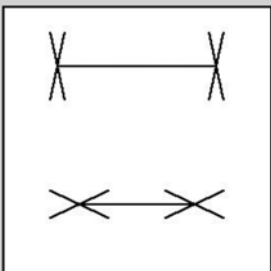
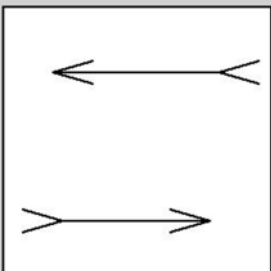
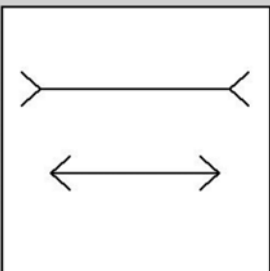
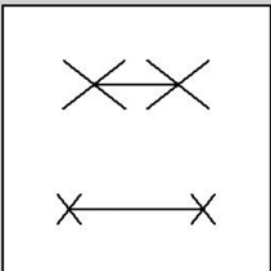
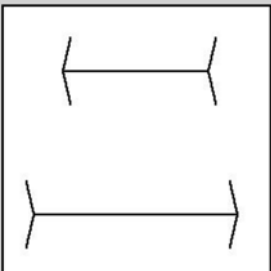
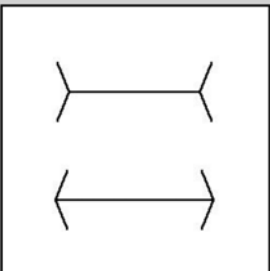
	CROSS FIN (XF) training images	CONTROL (LR) test images	ILLUSION (ML) test images
LONG (L) category examples			
SHORT (S) category examples			

Figure 2-3 Images presented to the model, categorised as LONG (top row) or SHORT (bottom row). The consistent manipulation across all three conditions (training, control and illusion) is the difference between top and bottom line lengths (while the size and orientation of fins changes as well as the distance between top and bottom lines. Column 1: Cross Fin (XF) images are used for training in all experiments. Column 2: Control (LR) images are used to test accuracy levels for a standard stimulus. Column 3: Illusion (ML) images are used to test performance levels for images that induce human perceptual error.

2.3 Results

2.3.1 Experiment I: Control

The first experiment we ran was to ensure that the classifier was able to distinguish long from short images at an acceptable level of accuracy and precision for a set of control stimuli. The control stimuli we used are illustrated in Figure 2-3 Column 2, where the top line has arrows pointing to the left and the bottom line has arrows pointing to the right. Fin angles were randomised between 10 and 70 degrees. We selected these control stimuli (annotated LR) because they contain the same number of features as those present in our illusion test stimuli.

As expected, performance results for the experiment were affected by the size of the network. We varied the number of S2 units (corresponding to the number of learned features) and measured the accuracy of classification as the average of performance (% correct) in each of 10 runs with 150 test images per category. Figure 2-4 illustrates these results, with error bars marking standard error of the mean between runs. When the network size reached 1000 S2 cells, performance exceeded 90%. With network sizes larger than 1000, performance did not substantially improve. We therefore chose to use this network size for all subsequent experiments so as to achieve high accuracy while minimising computational expense. For our following experiments, the critical comparison was between our control and illusion conditions.

With a network size of 1000 S2 cells, we achieved an overall accuracy of 90.3% for our control. We noticed a slight bias between our LONG category (89.2%) and our SHORT category (91.47%), however this was not statistically significant (using a two- tailed paired t-test, $p < 0.05$).

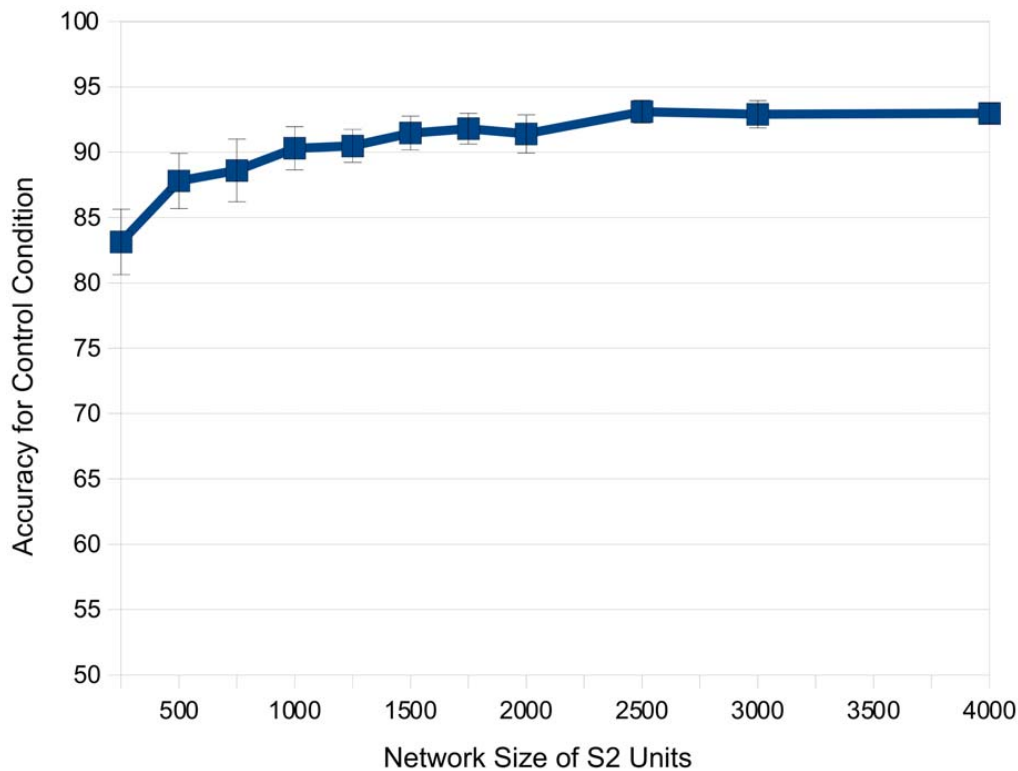


Figure 2-4 Experiment I: Control Results. Accuracy for the control condition versus the network size of S2 units. Values shown are the average of 10 runs. Error bars show standard error of the mean.

2.3.2 Experiment II: Illusion Effect

The second experiment compared the results from the control experiment against those obtained using illusory Müller-Lyer (ML) images. The ML images we tested are shown in Figure 2-3 Column 3, where the top line always has arrowtails and the bottom line always has arrowheads. The fin length and fin angle were varied in the same way as for the control images. If the top line always has arrowtails for every single test image, the top line will appear perceptually elongated. The bottom line always having arrowheads will appear contracted. For a human observer, this means that when the two lines are objectively of equal length, the top line will appear longer. When humans are presented with any of these ML images, they will therefore classify them as ‘long’ on more occasions than when control images are used.

If the model is not susceptible to the illusion, accuracy levels should be similar to those shown in Experiment I. However, if the model is susceptible to the illusion, then we should expect to see two effects. Firstly, for the LONG category, we would expect to see the model classifying these above the accuracy level in the control condition (89.2%). Secondly, for the SHORT category, we expect to see the classifier perform worse than the control condition (91.47%). Because of the consistent configuration of the test images, the machine learner would classify images as ‘long’ more often than the control condition. This would cause it to overclassify for the LONG category and underclassify in the SHORT category.

Figure 2-5 shows the accuracy (in terms of % correct) of ML image classification plotted alongside the control condition from Experiment I. Values displayed are the average of 10 runs for 150 test images per category and error bars indicate standard error. S2 network size was set to 1000 as in the control condition. As we can see from the figure, the ML condition shows classification accuracy above the control condition for the LONG category, however this difference was only trending towards significance (using a two- sample, equal variance t-test, $p=0.0674$). The inverse effect is shown in the SHORT category, where the ML condition performs under the classification accuracy of the control condition. The difference between the ML and Control conditions for the SHORT category was significant (using a two-sample, equal variance t-test, $p=0.000027$). This indicates that the model is indeed susceptible to the MLI.

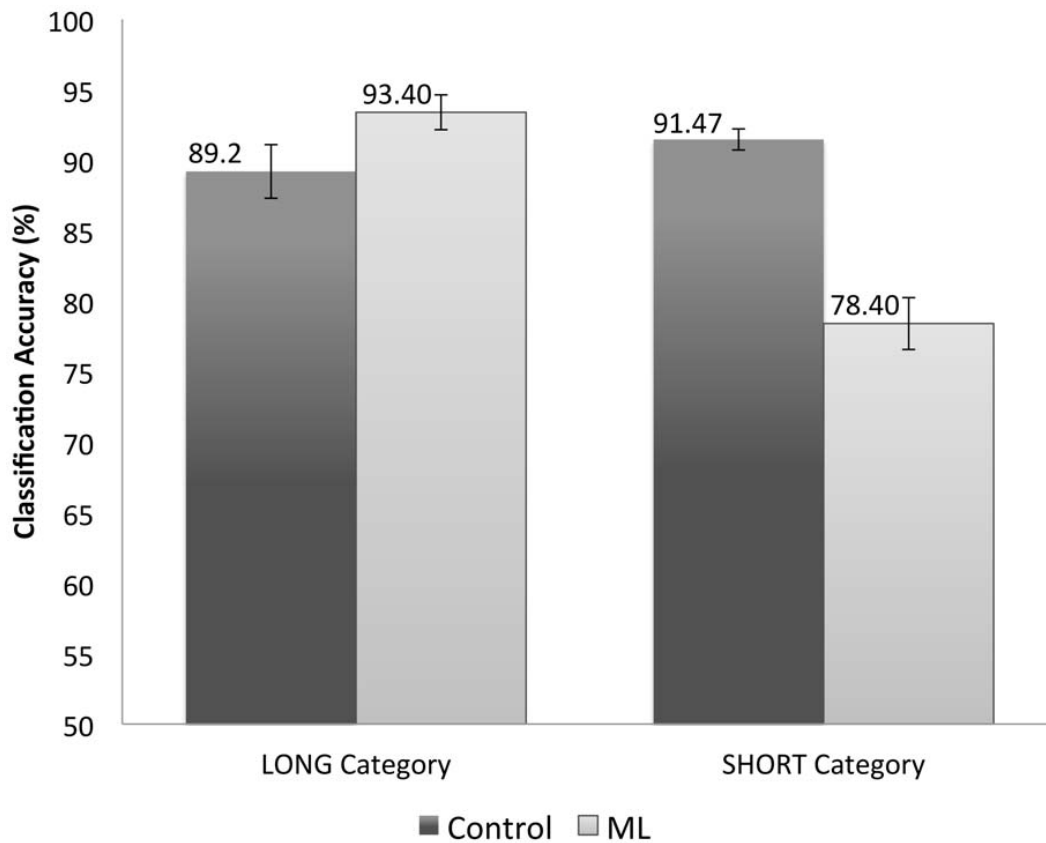


Figure 2-5 Experiment II: Control vs. Illusion Results. Accuracy (in terms of % correct) for the control versus ML images. Values shown are the average of 10 runs. Error bars indicate standard error of the mean.

2.3.3 Experiment III: Illusion Strength Affected by Angle

The results shown in Experiment II demonstrate errors consistent with an illusory effect; however they do not provide a detailed picture of classification performance using HMAX for control versus illusory data. We can obtain a better picture of the illusory effect within HMAX by measuring classification across incremental line length differences. By plotting classification results as a psychometric function, we are able to extract information such as the Point of Subjective Equality (PSE), for the illusory stimulus. Furthermore, we can separate out factors known to affect the strength of the illusion, such as the fin angle size or fin length, and observe consequent changes in the PSE.

Figure 2-6 shows results for the control condition versus illusion conditions with three separate fin angles, plotted as psychometric functions. Looking along the x-axis, negative values on the left indicate the SHORT category, and positive values on the right represent the LONG category. The y-axis indicates the percentage of images classified as LONG. If a classifier was always able to correctly identify the line length categories, we would see a sharp step function that takes the value of 0% on the left and 100% on the right, with a sharp transition at a line length difference of zero. Instead, what we see is a series of sigmoid functions indicating that when line length difference is large (in either negative or positive direction), it is easier for the system make a correct classification judgement. Sigmoid curves such as these are typical when mapping human psychophysical responses.

We first plotted the control condition with all angles collapsed. When there were large differences in line lengths (60 pixels), HMAX was able to categorise near ceiling for both the LONG category (far right) and the SHORT category (far left). When classification was at 50%, indicating that the top and bottom lines were judged to be the same length (i.e. the PSE), the line length difference was zero, indicating no bias. However, ML figures with 40 degree fins showed a PSE of -12.5 pixels. This indicates that with 40 degree fins, the top line must be 12.5 pixels shorter for HMAX to regard the two lines to be of equal length. Illusory lines with 20 degree angle fins demonstrated a much smaller PSE of -41 pixels. Considering human data, 20 degree angle fins would create an illusory bias of 26% (Restle & Decker, 1977). For our lines of 120 to 240 pixels, this would create an average PSE of 46.8 pixels. Therefore the PSE for 20 degree angle fins in HMAX is relatively congruent with human data. Illusory lines with 60 degree angle fins no longer demonstrated an illusory effect, indicated by a PSE of zero.

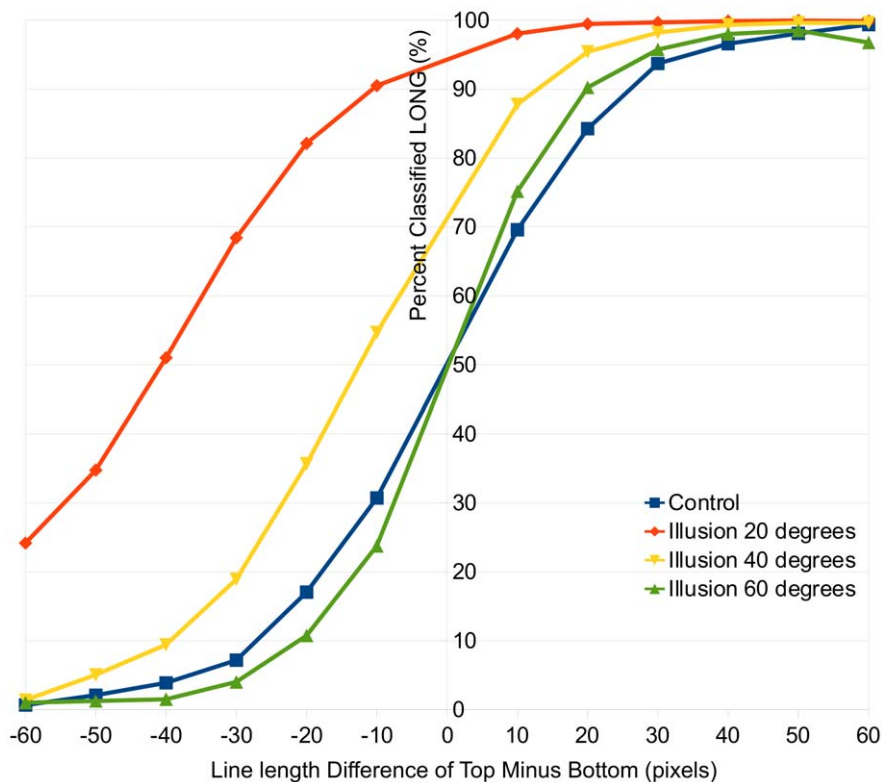


Figure 2-6 Experiment III: Illusion Strength Affected by Angle. Results here are plotted as psychometric curves with values on the left representing the SHORT condition, and values on the right representing the LONG condition. The control condition with all angles collapsed shows no bias. For illusory lines with 40 degree fins we see a PSE of approximately 12 pixels. Illusory lines with 20 degree fins show a larger PSE, congruent with human data. Illusory lines with 60 degree fins no longer demonstrate an illusory effect, indicated by intersection of the curve through 50% when the line length difference is zero.

Human data for the Müller-Lyer Illusion shows smaller effects as the angle becomes larger (Restle & Decker, 1977), which was also demonstrated by HMAX. For 20 degree data, HMAX performance matched human performance closely. However, as fin angles were increased, the illusory effect tapered off earlier in HMAX compared with human data. The 40 degree data showed a smaller effect than expected, while the 60 degree data show no effect at all, whereas humans are known to experience a Müller-Lyer effect with angles up to 80 degrees (Restle & Decker, 1977). So although we observed an overall illusory effect and a

degradation of illusory strength with an increase in fin angles, the illusory effect decreased faster for HMAX compared to humans.

2.4 Discussion

In this paper, we devised a set of experiments to measure the classification performance for an ML stimulus versus a control, in a biologically plausible model of object recognition. The task was to classify images as SHORT or LONG based on the relative lengths of top and bottom lines in an image. We trained the model using a set of cross fin images that do not induce any illusion in humans and that contain all features present in test stimuli. In Experiment I, we explored different network sizes to achieve an overall classification accuracy level of 90% for our control condition. We then compared these results to an illusory stimulus in Experiment II, where we observed a respective increase and decrease in classification accuracy for the long and short conditions. This indicates that, as for human observers, this computational model of object recognition shows skewed performance levels when judging relative line lengths for Müller-Lyer stimuli. In Experiment III, we further investigated the strength of the illusion within the model by manipulating fin angle. We observed a smaller PSE for illusory stimuli with more acute fin angles, indicating a larger illusory effect. As fin angles increased, the PSE increased. This suggests that the HMAX computational model of object recognition is able to emulate the human MLI in two ways: 1) by demonstrating an overall bias in line length classification with illusory stimuli and 2) by demonstrating a larger Müller-Lyer effect with more acute fin angles.

Although HMAX is able to demonstrate an illusory effect, our results are not identical to patterns seen in human data. In particular, one possible reason for this is that even though HMAX is a biologically plausible model, it does omit a number of features present in the human visual system, most notably the notion of feedback or recurrent connections. Because HMAX is fundamentally a feed-forward model, to make a fair comparison between the

illusion in HMAX and the illusion in humans, results from the model should be compared with human results obtained using a backward masking paradigm or repetitive transcranial magnetic stimulation (rTMS). Human psychophysical experiments performed on the MLI have, to date, not included methods that eliminate feedback processing, such as backward masking or rTMS. We plan to run further experiments using backward masking in human subjects to allow for this comparison.

Careful consideration was applied to selecting our control test stimuli. We ruled out the use of straight fin images (having wings orthogonal to the shaft) because they contain a smaller number of features compared to ML stimuli. We also ruled out the possibility of using different combinations of terminating fins because the Müller-Lyer illusion exists in many forms. We discovered that the best control stimuli were a combination of left and right arrowheads. These control images not only contain the same number of features as the ML stimuli, but also allow us to directly compare accuracy levels with varying fin angles and fin lengths.

Misclassification of the ML images, as shown in Figure 2-5, indicates that this computational model is susceptible to perceptual errors similar to those experienced by humans. These experimental results add to the plausibility of models that adopt a simple- complex architecture. Not only is the HMAX model able to achieve accuracy levels on par with humans in performing rapid object categorisation (Serre et al., 2007), we now show that this model can mimic aspects of human performance in misclassifying illusory stimuli.

The other significant and perhaps most surprising finding from these experiments is that the illusion was generated in a model that includes only feed-forward processing. No feedback connections are present in the HMAX model, and apart from initial feature training in the learning stage of the model, weights and connections are fixed during normal operation.

Information in the system flows in one direction, from the initial image layer through simple and complex layers to the SVM. This implies that ML line length misjudgement can occur from feedforward connections alone.

Shortcomings of HMAX, including the lack of recurrent connections, may not be the sole explanation for the gap between model and human data. Another possibility for this mismatch is the use of constrained training images, consisting entirely of thin black lines on a white background. Including natural scenes as part of the training set, for example, may improve the match with human data. Each of these points could be addressed separately by testing other models or by training HMAX with other image sets. Our results provide a baseline for further comparisons and the analysis of other potential explanations of the MLI.

The images used for training the model allow us to further assess proposed explanations of the MLI. The image set we used for training was inherently two dimensional in nature, consisting only of straight black lines on a white background (see XF images in Figure 2-3). In order to verify Gregory's (1963) misapplied size constancy scaling theory, we would need to train the model on images taken of 3D scenes. Gregory (1963) argues that illusory figures are 'flat projections of typical views of objects lying in three-dimensional space'. Given that our model exhibited an illusory effect without training on any 3D images, we can be confident that misapplied size constancy scaling is not a necessary factor in causing the MLI in our model, and to the extent that this model mimics human visual processing of ML figures, it may not be necessary to explain the behaviour of human subjects. Our training image set further suggests that the ML illusion can occur in the absence of statistics of image-source relationships. Howe and Purves (2005) propose that the ML illusion is caused by the "statistical relationship between retinal images and their real- world sources". For our experiments, we did not train HMAX on any natural images and maintained a consistent

number of features across our training images. Our results suggest that the Müller-Lyer illusion can be caused even without information embedded in natural images.

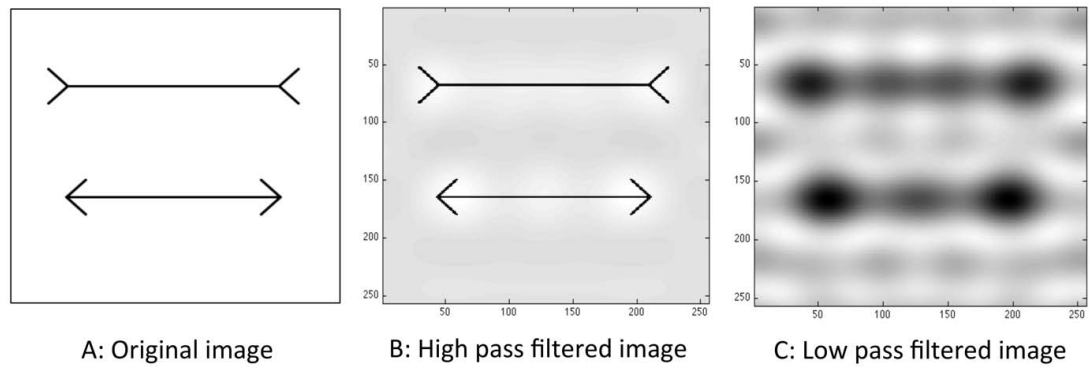


Figure 2-7 An image of the Müller-Lyer Illusion high and low pass filtered. A: the original image B: The image high pass filtered at 5 cycles per image. C: The image low pass filtered at 5 cycles per image.

Ginsburg (1978) suggested that in human observers, the MLI is caused primarily by stronger weighting of low spatial frequency information, later supported by results from Carrasco (1986). When Müller-Lyer figures are low pass filtered, a physical difference manifests, elongating the wings-out figure (see Figure 2-7). If HMAX were to give stronger weighting to information flowing from units representing larger spatial scales, this would be expected to produce a similar effect. To investigate this possibility, we conducted a post-hoc analysis on one of the trained networks. We first extracted how information is weighted within the SVM layer of the model and then mapped these weights to spatial frequencies. Within the HMAX model, there is a direct relationship between spatial frequency information and receptive field size (Serre & Riesenhuber, 2004). We were able to graph the bounding box sizes of the top 20 most influential features used by the SVM in order to make a classification decision (out of 1000 available). Figure 2-8 shows that the majority of highly weighted features fed into the SVM contain high spatial frequency information. This is inconsistent with the potential

explanation that low spatial frequency information is highly influential in driving the MLI in humans. We can therefore rule out the possibility that the illusion in the network is caused by stronger weighting of low spatial frequency information.

We have demonstrated that a Müller-Lyer effect can arise in an artificial model of neural information processing. This provides an opportunity to test the extent to which hypothesised underlying neural mechanisms contribute to the illusion. For instance, lateral inhibition has been proposed as an explanation for the MLI (Coren, 1970). We initially explored how changing lateral inhibition levels within the HMAX architecture affects classification performance, but altering lateral inhibition levels affected the accuracy of classification of control stimuli, which was maximal at the default parameter settings. Since we measured the Müller-Lyer effect by comparing classification performance for illusory images against control images, we therefore decided to keep the default lateral inhibition levels where the control accuracy was highest. It may be useful to further examine the role of lateral inhibition in the future. Other possibilities include the isolation of information at different orientations to assess their relative contributions to the size of the illusion. Although beyond the scope of the current study, these have the potential to be useful tests of contributing mechanisms.

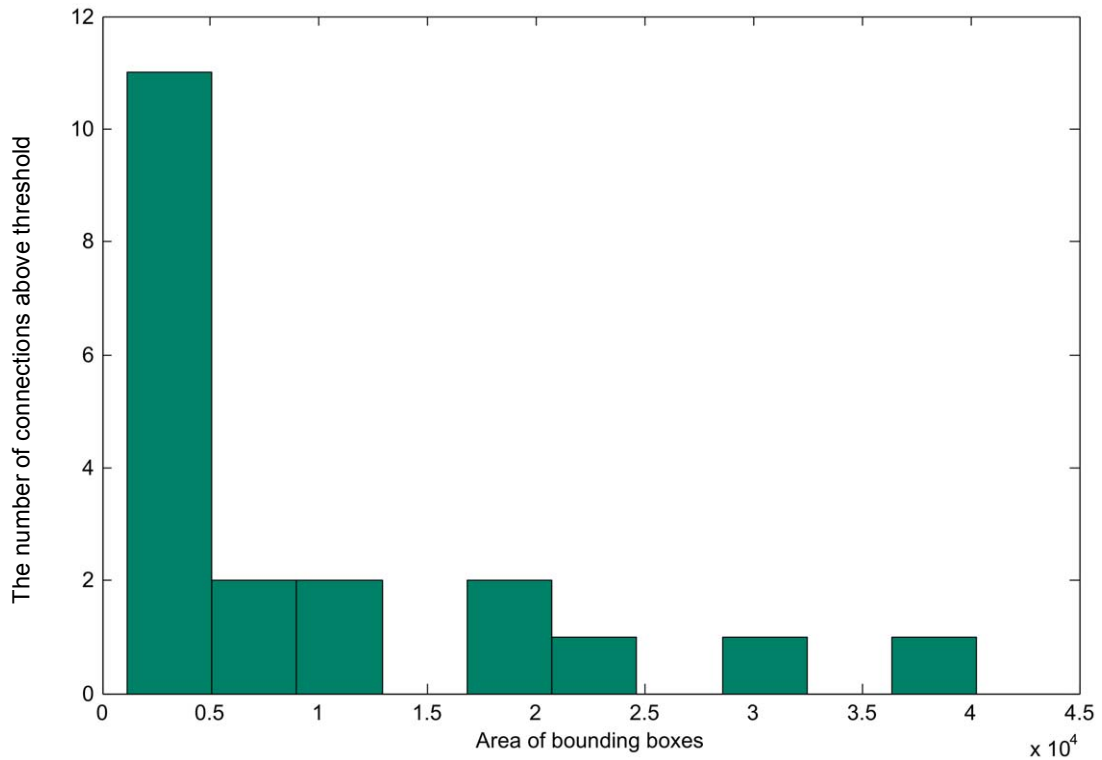


Figure 2-8 The twenty most influential features used by the SVM layer in a representative trained network, ordered by bounding box size. A post-hoc analysis of a trained network showed the 20 most influential features used to make a classification decision out of 1000 available. Stronger weighting is placed on features that have small bounding boxes and hence contain high spatial frequency information.

To date, there have been relatively few studies where artificial neural networks or computer models have been used to explore visual illusions (Bertulis & Bulatov, 2001; Ogawa *et al.*, 1999; Corney & Lotto, 2007). In some cases, these artificial neural networks were not built in order to mimic neural computation, but rather to demonstrate statistical correlations in input data (Corney & Lotto, 2007). The model used in Corney & Lotto (2007) consisted of only one hidden layer with four homogenous neurons, which few would consider to be even a crude representation of visual cortex. The work presented in Ogawa *et al.* (1999) used a network with three hidden layers of ‘orientational neurons’, ‘rotation neurons’ and ‘line unifying neurons’. This network could roughly correspond to one layer of simple cells that provide orientation filters and one layer of complex cells that provide convolution. However, this

study did not present any quantitative data and did not clearly state details of their method, such as the size or connectivity of their network. Bertulis and Bulatov (2001) created a computer model to replicate the spatial filtering properties of simple cells and convolution of complex cells in visual cortex. They compared human and model data for the Brentano (single shaft) form of the Müller-Lyer Illusion with 45 degree fins, which combines contraction and elongation effects, so it is not straightforward to make a direct comparison of results with HMAX performance. The Bertulis and Bulatov (2001) model centred only on filtering properties of neurons. In contrast, our study employs machine learning techniques to train the model on multiple images before running a classification task and comparing the task of interest to a control. Our study allows us to separate out the inner workings of a model from the input fed into it, in the form of training images. So although studies exist that model visual illusions within artificial neural networks, we believe that the current study represents a significant advance, being the first to model a visual illusion in a ‘biologically plausible’ artificial neural network.

That HMAX is capable of object classification, the task for which it was originally developed, may be considered impressive, given the relative simplicity of the model, which includes no feedback. However, in the current study we have presented evidence that the model is able to predict human-like performance in a completely unrelated task: that involving the discrimination of line length. Further, the correspondence of performance between man and machine represents not just degrees of classification accuracy, but also captures the pattern of errors that are made as a function of difference in line length and fin angle, and produces evidence of an illusion. These were emergent properties, rather than the model being deliberately constructed to produce these features. This might raise questions as to other visual phenomena that HMAX may be capable of accounting for, and also raises the possibility that HMAX may be capable of predicting other yet to be observed phenomena. We

look forward to such research being carried out in the near future.

Acknowledgements

We thank Jim Mutch for making FHLib publicly available via GNU General Public License and Thorsten Joachims for creating the SVMLight package used within FHLib. We thank Max Coltheart from the Cognitive Science Department at Macquarie University for helpful discussions. We would also like to thank an anonymous reviewer for their helpful suggestions.

2.5 References

- Bertulis, A., & Bulatov, A. (2001). Distortions of length perception in human vision. *Biomedicine*, 1, 3–26.
- Brigell, M., & Uhlarik, J. (1979). The relational determination of length illusions and length aftereffects. *Perception*, 8, 187–197.
- Dewar, R. (1967). Stimulus determinants of the practice decrement of the Müller- Lye illusion. *Canadian Journal of Psychology*, 21, 504–520.
- Carrasco, M., Figuerola, J. G., Willen, J. D. (1986). A test of the spatial-frequency explanation of the Müller-Lyer illusion. *Perception*, 15, 553–562.
- Coren, S. (1970). Lateral inhibition and geometric illusions. *The Quarterly Journal of Experimental Psychology*, 22, 274–278.
- Coren, S., Porac, C. (1984) Structural and cognitive components in the Müller- Lye illusion assessed via cyclopean presentation. *Perception and Psychophysics*, 35, 313–318.
- Corney, D., Lotto, R. B. (2007). What are lightness illusions and why do we see them? *PLoS Computational Biology*, 3, e180.
- Fellows, B. J. (1967) Reversal of the Müller-Lyer illusion with changes in the length of the inter-fins line. *The Quarterly Journal of Experimental Psychology*, 19, 208–214.
- Fukushima, K. (1980). Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, 36, 193–202.
- Ginsburg, A. (1978). *Visual Information Processing Based on Spatial Filters Constrained by Biological Data*. Ph.D. thesis, Aerospace Medical Research Laboratory, Aerospace Medical Division, Air Force Systems Command.
- Goodale, M. A., & Milner, A. D. (1992) Separate visual pathways for perception and action. *Trends in Neurosciences*, 15, 20–25.
- Gregory, R. L. (1963) Distortion of visual space as inappropriate constancy scaling. *Nature*, 199, 678–680.
- Hubel, D. H., & Wiesel, T. N. (1959) Receptive fields of single neurones in the cat's striate

cortex. *Journal of Physiology*, 148, 574–591.

Howe, C. Q., & Purves, D. (2005). The Müller-Lyer illusion explained by the statistics of image–source relationships. *Proceedings of the National Academy of Sciences*, 102, 1234–1239.

LeCun, Y., Jackel, L. D., Boser, B., Denker, J. S., Graf, H. P., Guyon, I., Henderson, D., Howard, R. E., & Hubbard, W. (1989) Handwritten digit recognition: Applications of neural network chips and automatic learning. *IEEE Communications Magazine*, 27, 41–46.

LeCun, Y., & Bengio, Y., (1995) Convolutional networks for images, speech and time- series. In: Arbib MA, editor, *The handbook of brain theory and neural networks*. MIT Press. 255–258.

Mutch, J., & Lowe, D. G. (2008) Object class recognition and localization using sparse features with limited receptive fields. *International Journal of Computer Vision*, 80, 45–57.

Müller-Lyer, F. C. (1889) Optische Urteilstauschungen. *Archiv für Anatomie und Physiologie*, 2, 263–270.

Müller-Lyer, F. C. (1896a) Zur Lehre von den optischen Tauschungen über Kontrast und Konfluxion. *Zeitschrift für Psychologie und Physiologie der Sinnesorgane*, 9, 1–16.

Müller-Lyer, F. C. (1896a) Über Kontrast und Konfluxion. (Zweiter Artikel). *Zeitschrift für Psychologie und Physiologie der Sinnesorgane*, 10, 421–431.

Ogawa, T., Minohara, T., Kanaka, I., Kosugi, Y. (1999). A neural network model for realizing geometric illusions based on acute-angled expansion. In: *6th International Conference on Neural Information Processing (ICONIP '99) Proceedings*. IEEE, volume 2, 550–555.

Predebon, J. (1998). Decrement of the Brentano Müller-Lyer illusion as a function of inspection time. *Perception*, 27(2), 183–192.

Restle, F., & Decker, J. (1977) Size of the Mueller-Lyer illusion as a function of its dimensions: Theory and data. *Perception and Psychophysics*, 21, 489–503.

Serre, T., & Riesenhuber, M. (2004). Realistic modeling of simple and complex cell tuning in the HMAX model, and implications for invariant object recognition in cortex. *AI Memo 2004–017*, Massachusetts Institute of Technology, Computer Science and Artificial

Intelligence Laboratory (CSAIL).

Serre, T., Wolf, L., Poggio, T. (2005). Object recognition with features inspired by visual cortex. *In: Proceedings of 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005)*. San Diego: IEEE Computer Society Press, 886–893.

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

Serre, T., & Poggio, T. (2010). A neuromorphic approach to computer vision. *Communications of the ACM*, 53, 54–61.

Weidner, R., & Fink G. R. (2007) The neural mechanisms underlying the Müller-Lyer illusion and its interaction with visuospatial judgments. *Cerebral Cortex*, 17, 878–884.

Weidner, R., Boers, F., Mathiak, K., Dammers, J., Y & Fink, G. R. (2010) The temporal dynamics of the Müller-Lyer illusion. *Cerebral Cortex*, 20, 1586–1595.

2.6 Appendix: Determining whether low spatial frequency information may be influencing the SVM

To determine whether the illusion, or part of it, could be caused by higher weighting being placed on low spatial frequency information we took a closer look into a trained network, first extracting how information is weighted within the SVM layer of the model and then mapping these weights to spatial frequencies. Below is the two-stage process we adopted in order to extract this information and see whether certain spatial frequency information was favoured in making a classification decision.

2.6.1 Stage I: Extracting the highest weights entering the SVM layer

Two learning mechanisms play a role in the HMAX architecture making a decision:

- i) the classification uses a supervised learning procedure. In this a linear support vector machine is trained, with input from the top layer of the HMAX network that consists of 1000 features. The output of this module is the classification.
- ii) the neural network (layers S1,C1,S2,C2) is trained in an unsupervised way. The features (and spatial scales) in C2 (and also in C1) are the ones with maximum response in the layer below, i.e., for different types of input used during training, the features in C2 may correspond to different spatial scales.

To identify which spatial scales are the ones actually used for a classification, we needed to consider the effects of both mechanisms. Beginning at the classification module, we have our support vector machine which in its primal form aims to satisfy the equations:

- 1) $w \cdot x - b = 1$ for inputs x that belong to the first class, and
- 2) $w \cdot x - b = -1$ for inputs x that belong to the second class. (where w and x are vectors and b is a bias term).

The absolute value of a component of the weight vector w describes how strong the influence of its respective input feature to the classification is (weights can be positive or negative). The support vector machine is trained by solving using an optimisation problem in which a number of support vectors are found (points close to the decision plane that lies between both classes), and corresponding Lagrange multipliers α . We reconstruct w using the support vectors and α , and use them to identify the strongest contributing features from C2 to our classification.

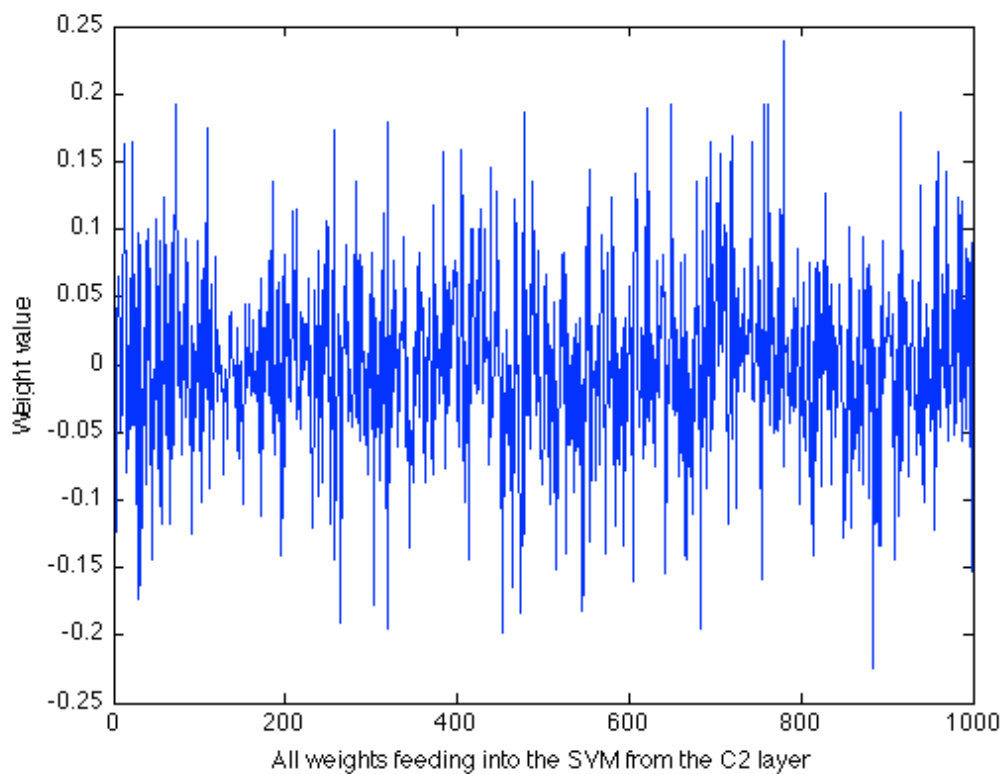


Figure 2-9: All 1000 weight values feeding into the SVM

Here we can see the majority of weights are around zero and the absolute value of the maximum weight does not go above 0.25. To extract the most influential weights we sorted by absolute value and had a look at the top 20 and top 100 weightings going into the SVM.

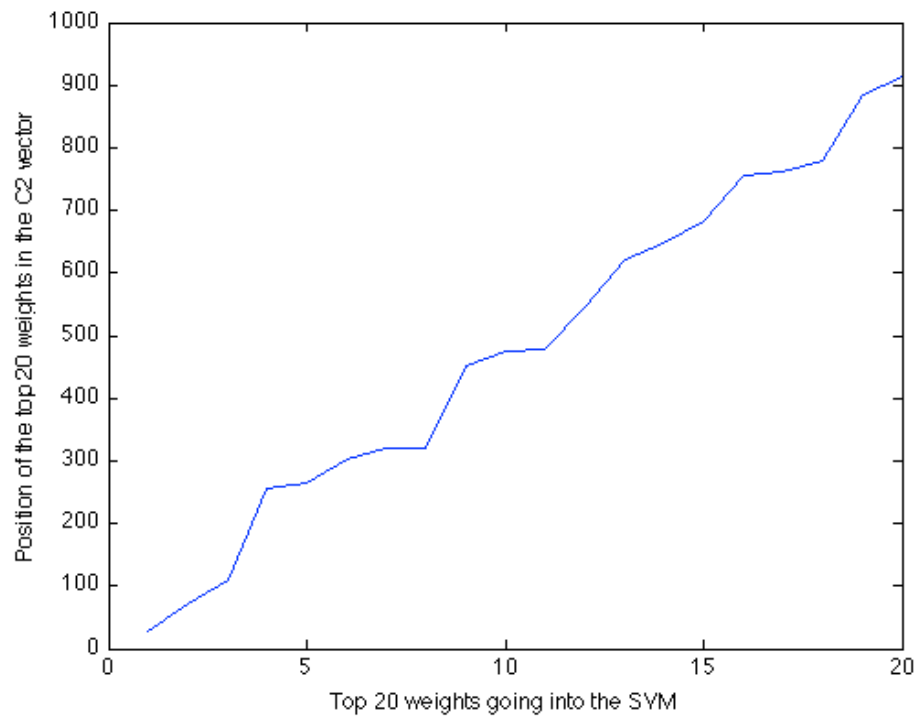


Figure 2-10 Positions of the top 20 weights within the SVM

2.6.2 Stage II: Identify the spatial scale of the top contributing features in C2.

Information is propagated in the model from S1->C1->S2->C2->SVM. So it is only the C2 layer that directly affects the SVM decision. In order to see how information in C2 is organised, we first look at how information is structured in the lower layers.

S1 is arranged into 12 orientations with patches of increasing size. The size of S1 is $12 \times (246 \times 246 + 205 \times 205 + 171 \times 171 + 142 \times 142 + 118 \times 118 + 98 \times 98 + 81 \times 81 + 66 \times 66 + 54 \times 54 + 44 \times 44)$. We can see that there are many more high spatial filters applied (246×246) than low spatial filters (44×44). Smaller Gabor patch filters (or receptive fields) correspond to high spatial filters. According to [1], spatial frequency is directly correlated with the receptive field size. This is because Serre and Riesenhuber (2004) fix the wavelength, aspect ratio and effective width of the Gabor filters in order "to account for general cortical cell properties,

that is: (i) Cortical cells' peak frequency selectivities are negatively correlated with the receptive field sizes. (ii) Cortical cells' spatial frequency selectivity bandwidths are positively correlated with their receptive field sizes".

A similar structure of filter sizes are preserved throughout the C1 and S2 layers, but the C2 layer takes the maximum of responses learned using unsupervised learning stage. Therefore connections need to be traced back to the previous layer in order to extract the bounding box sizes. We then map the area of bounding boxes to the C2 units that are most influential.

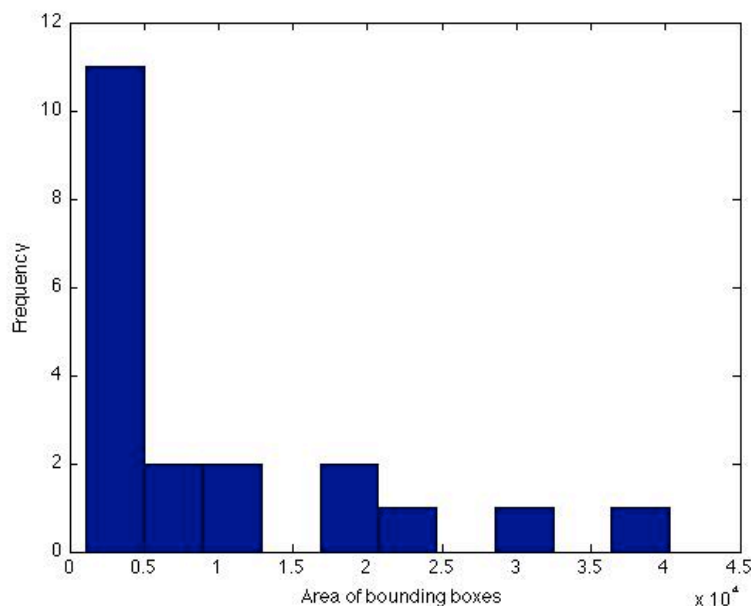


Figure 2-11 Top 20 most influential features ordered by size

From this graph above we can see the majority of the top 20 most influential features on the SVM are features with small bounding boxes or higher frequency spatial information. We can also plot the top 100 most influential features that contribute towards the SVM classification sorted by bounding box area. This is illustrated below a blue histogram. The 1000 available features that the SVM uses to make a classification decision are shown as a red histogram.

We can see that the majority of features that influence the SVM have small bounding boxes/contain high spatial frequency information.

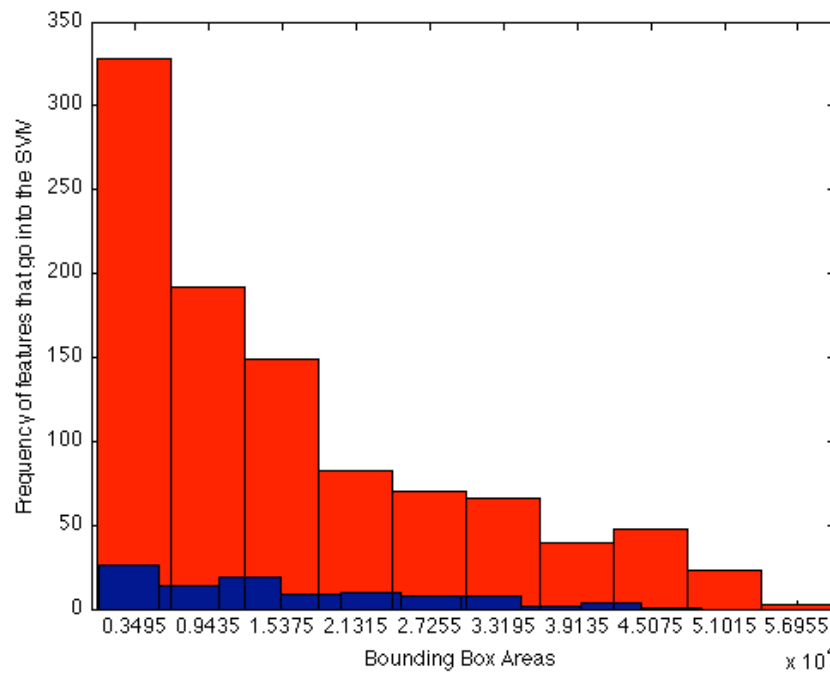


Figure 2-12 The 100 most influential features versus the 1000 available sorted by bounding box area

The previous graph illustrates that the SVM decision does not heavily rely on low spatial frequency information. This is in contrast to the prediction that the decision would be strongly dependent on the low spatial frequency information.

3 Study 2

Complex cells decrease errors for the Müller-Lyer illusion in a model of the visual ventral stream.

3.1 Abstract

To improve robustness in object recognition, many artificial visual systems imitate the way in which the human visual cortex encodes object information as a hierarchical set of features. These systems are usually evaluated in terms of their ability to accurately categorize well-defined, unambiguous objects and scenes. In the real world, however, not all objects and scenes are presented clearly, with well-defined labels and interpretations. Visual illusions demonstrate a disparity between perception and objective reality, allowing psychophysicists to methodically manipulate stimuli and study our interpretation of the environment. One prominent effect, the Müller-Lyer illusion, is demonstrated when the perceived length of a line is contracted (or expanded) by the addition of arrowheads (or arrow-tails) to its ends. HMAX, a benchmark object recognition system, consistently produces a bias when classifying Müller-Lyer images. HMAX is a hierarchical, artificial neural network that imitates the “simple” and “complex” cell layers found in the visual ventral stream. In this study, we perform two experiments to explore the Müller-Lyer illusion in HMAX, asking: (1) How do simple vs. complex cell operations within HMAX affect illusory bias and precision? (2) How does varying the position of the figures in the input image affect classification using HMAX? In our first experiment, we assessed classification after traversing each layer of HMAX and found that in general, kernel operations performed by simple cells increase bias and uncertainty while max-pooling operations executed by complex cells decrease bias and uncertainty. In our second experiment, we increased variation in the positions of figures in the input images that reduced bias and uncertainty in HMAX. Our findings suggest that the Müller-Lyer illusion is exacerbated by the vulnerability of simple cell operations to positional fluctuations, but ameliorated by the robustness of complex cell responses to such variance.

3.2 Introduction

Much of what is known today about our visual perception has been discovered through visual illusions. Visual illusions allow us to study the difference between objective reality and our interpretation of the visual information that we receive. Recently it has been shown that computational vision models that imitate neural mechanisms found in the ventral visual stream can exhibit human-like illusory biases (Zeman *et al.*, 2013). To the extent that the models are accurate reflections of human physiology, these results can be used to further elucidate some of the neural mechanisms behind particular illusions.

In this paper, we focus on the Müller-Lyer Illusion (MLI), which is a geometrical size illusion where a line with arrowheads appears contracted and a line with arrow-tails appears elongated (Müller-Lyer, 1889) (see Figure 3-1). The strength of the illusion can be affected by the fin angle (Dewar, 1967), shaft length (Fellows, 1967; Brigell and Uhlarik, 1979), inspection time (Coren and Porac, 1984; Predebon, 1997), observer age (Restle and Decker, 1977), the distance between the fins and the shaft (Fellows, 1967) and many other factors. The illusion classically appears in a four-wing form but can also manifest with other shapes, such as circles or squares, replacing the fins at the shaft ends. Even with the shafts completely removed, the MLI is still evident.

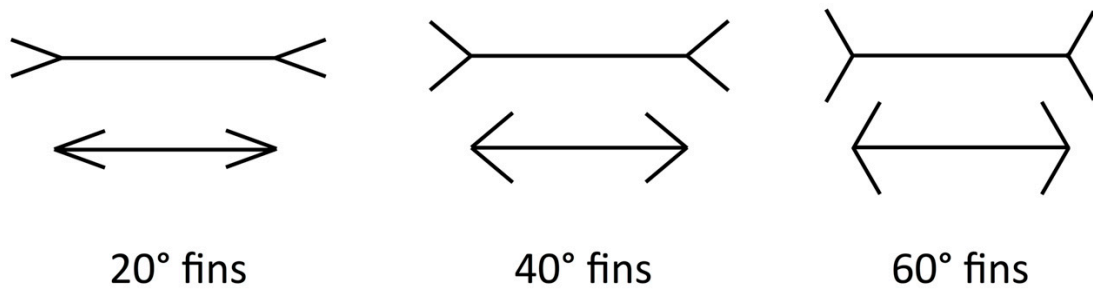


Figure 3-1 The ML illusion in classical four-wing form. Horizontal lines are the same length in all cases. The ML effect is stronger for more acute angles (Left) and weaker for more obtuse angles (Right).

Here, we employ an underused method to explore the Müller-Lyer illusion and its potential causes using an Artificial Neural Network (ANN). To date, few studies have used ANNs to explore visual illusions (Ogawa *et al.*, 1999; Bertulis and Bulatov, 2001; Corney and Lotto, 2007). In some cases, these artificial neural networks were not built to emulate their biological counterparts, but rather to demonstrate statistical correlations in the input. One such example is the model used by Corney and Lotto (2007), consisting of only one hidden layer with four homogenous neurons, which few would consider to be even a crude representation of visual cortex. The work presented by Ogawa *et al.* (1999) used a network with three hidden layers of “orientational neurons,” “rotational neurons” and “line unifying neurons.” This network could roughly correspond to one layer of simple cells that provide orientation filters and one layer of complex cells that combine their output. However, this study presented no quantitative data and lacked a detailed description of the model, such as the size or connectivity of their network. Bertulis and Bulatov (2001) created a computer model to replicate the spatial filtering properties of simple cells and the combination of these units’ outputs by complex cells in visual cortical area V1. Although they compared human and model data for the Müller-Lyer Illusion, their model centered only on the filtering properties

of early visual neurons. These models do not adequately represent the multi-layered system that would best describe the relevant neural structures. Neuroimaging studies have shown areas V1, V2, V4, and IT are recruited when viewing the MLI (Weidner and Fink, 2007; Weidner *et al.*, 2010) and hence the inclusion of operations from such visual ventral stream subdivisions is desirable. Therefore, studying the MLI in a computational model known to mimic these areas would provide a more biologically representative result.

In a previous report, we studied the MLI in a benchmark model of the ventral visual stream that imitates these cortical areas (Zeman *et al.*, 2013). Following from our hypothesis that the MLI could occur in a model that imitates the structure and function of visual ventral areas, we demonstrated its manifestation in a biologically plausible artificial neural network. Although the models listed above are capable of reproducing the MLI, we believe our work provides a significant advance, being one of the first studies to model a visual illusion in a simulated replica of the ventral visual stream. In addition, our study contrasts with those above by employing techniques to train the model on multiple images before running a classification task and comparing the task of interest to a control. This allows us to separate the inner workings of the model from the input in the form of training images.

The model we recruit, HMAX (Serre *et al.*, 2005), is a feed-forward, multi-layer, artificial neural network with layers corresponding to simple and complex cells found in visual cortex. Like visual cortex, the layers of HMAX alternate between simple and complex cells, creating a hierarchy of representations that correspond to increasing levels of abstraction as you traverse each layer. The simple and complex cells in the model are designed to match their physiological counterparts, as established by single cell recordings in visual cortex (Hubel and Wiesel, 1959). Here, we briefly describe single and complex cell functions and provide further detail on these later in Section 2.2.1. In short, simple cells extract low-level features, such as edges, an example of which would be Gabor filters that are often used to model V1

operations. The outputs of simple cells are pooled together by complex cells that extract combined or high-level features, such as lines of one particular orientation that cover a variety of positions within a visual field. Within HMAX, the max pooling function is used to imitate complex cell operations, giving the model its trademark name. In general, low-level features extracted by simple cells are shared across a variety of input images. High-level features are less common across image categories. The high-level features output by complex cells are more stable, invariant and robust to slight changes in the input.

HMAX has been extensively studied in its ability to match and predict physiological and psychological data (Serre and Poggio, 2010). Like many object recognition models, HMAX has been frequently tested using well-defined, unambiguous objects and scenes but has not been thoroughly assessed in its ability to handle visual illusions. Our previous demonstration of the MLI within HMAX showed not only a general illusory bias, but also a greater effect with more acute fin angles, corresponding to the pattern of errors shown by humans. Our replication of the MLI in this model allowed us to rule out some of the necessary causes for the illusion. There are a number of theories that attempt to explain the MLI (Gregory, 1963; Segall et al., 1966; Ginsburg, 1978; Coren and Porac, 1984; Müller-Lyer, 1896a, 1896b; Bertulis and Bulatov, 2001; Howe and Purves, 2005; Brown and Friston, 2012) and here we discuss two. One common hypothesis is the “carpentered world” theory - that images in our environment influence our perception of the MLI (Gregory, 1963; Segall et al., 1966). To interpret and manoeuvre within our visual environment, we apply a size-constancy scaling rule that allows us to infer the actual size of objects from the image that falls on our retina. While arrowhead images usually correspond to the near, exterior corners of cuboids, arrow-tail configurations are associated with more distant features, such as the right-angled corners of a room. If the expected distance of the features is used to scale our perception of size, when a line with arrowheads is compared to a line with arrow-tails that is physically equal in length,

the more proximal arrowhead line is perceived as being smaller. Another common theory is based upon visual filtering mechanisms (Ginsburg, 1978). By applying a low spatial frequency filter to a Müller-Lyer image, the overall object (shaft plus fins) will appear elongated or contracted. Therefore, it could simply be a reliance on low spatial frequency information that causes the MLI. In our previous study, we were able to replicate the MLI in HMAX, allowing us to establish that exposure to 3-dimensional “carpentered world” scenes (Gregory, 1963) is not necessary to explain the MLI, as the model had no representation of distance and hence involved no size constancy scaling for depth. We also demonstrated that the illusion was not a result of reliance upon low spatial frequency filters, as information from a broad range of spatial frequency filters was used for classification.

In the current study, we set out to investigate the conditions under which the Müller-Lyer illusion manifests in HMAX and what factors influence the magnitude and precision of the effect. In particular, we address the following questions: (1) How do simple vs. complex cell operations within HMAX affect illusory bias and precision? (2) How would increasing the positional variance of the input affect classification in HMAX? Our principal motivation is to discover how HMAX processes Müller-Lyer images and transforms them layer to layer. Following from this, we aim to find ways to reduce errors associated with classifying Müller-Lyer images, leading to improvements in biologically inspired computational models. We are particularly interested in how hierarchical feature representation could potentially lead to improvements in the fidelity of visual perception both in terms of accuracy (bias) and precision (discrimination thresholds).

3.3 Materials and Methods

3.3.1 Computational model: HMAX

To explore where and how the illusion manifests, we first examined the architecture of HMAX: a multi-layer, feed-forward, artificial neural network (Serre et al., 2005; Mutch and Lowe, 2008; Mutch et al., 2010). Input is fed into an image layer that forms a multi-scale representation of the original image. Processing then flows sequentially through four more stages, where alternate layers perform either template matching or max pooling (defined below). HMAX operations approximate the processing of neurons in cat striate cortex, as established by single cell recordings (Hubel and Wiesel, 1959). Simple cells are modeled using template matching, responding with higher intensity to specific stimuli, while complex cell properties are simulated using max pooling, where the maximum response is taken from a pool of cells that share common features, such as size or shape.

Image information travels unidirectionally through four layers of alternating simple (“S”) and complex (“C”) layers of HMAX that are labeled S1, C1, S2, and C2. When the final C2 level is reached, output is compressed into a 1D vector representation that is sent to a linear classifier for final categorization. While previous versions of HMAX employed a support vector machine (SVM), in this paper we used the GPU-based version of HMAX (Mutch et al., 2010) that uses a linear classifier to perform final classification. The task for the classifier was to distinguish Long (i.e., top shaft longer) from Short (top shaft shorter) stimulus categories under a range of conditions, where the top or bottom line length varied by a known positive or negative extent. Figure 3-2 summarizes the layers and operations in the model. Precise details are included in the original papers (Serre et al., 2005; Mutch and Lowe, 2008; Mutch et al., 2010).

3.3.2 Stimuli: Training and test sets (Control and Müller-Lyer)

To carry out our procedure, we generated three separate image sets: a training (cross fin) set, a control test set (CTL) and an illusion test set (ML). All images were 256×256 pixels in size, with black 2×2 pixel lines drawn onto a white background (see Figure 3-3). Each image contained two horizontal lines (“shafts”) with various fins appended. Each different image set was defined by the type of fins appended to the ends of the shafts. The fin type determines whether an illusory bias will be induced or not. Unlike the ML set, the cross fin and control test sets do not induce any illusions of line length in humans (Glazebrook *et al.*, 2005; Zeman *et al.*, 2013).

Within each two-line stimulus, the length of the top line was either “long” (L), or “short” (S), compared to the bottom line. The horizontal shaft length of the longest line was independently randomized between 120 and 240 pixels. The shorter line was varied by a negative extent randomly between 2 and 62 pixels for the training set, or by a known negative extent between 10 and 60 pixels for the test sets. The positions of each unified figure (shaft plus fins) were independently randomly jittered in the vertical direction between 0 and 30 pixels and in the horizontal direction between -30 and 30 pixels from center. The vertical position of the top line was randomized between 58 and 88 pixels from the top of the image while the bottom line’s vertical position was randomized between 168 and 198 pixels. Top and bottom fin lengths randomized independently between 15 and 40 pixels. Fin lengths, line lengths and line positions remained consistent across all image sets. The parameters that varied between sets were fin angle, the direction of fins and the set size. If an image was generated that had any overlapping lines, for example, arrowheads touching or intersecting, these images were excluded from the sets.

Training images contained two horizontal lines with cross fins appended to the ends of the shafts (see Row 1, Figure 3-3). Fin angles were randomized independently for the top and bottom lines between 10 and 90°. Five hundred images per category (long and short) were used for training.

Two sets of test images were used, one as a control test set (CTL) and one as an illusion test set (ML). The CTL set used for parameterization contained left facing arrows for the top line and right facing arrows for the bottom line (see Row 2, Figure 3-3). CTL fin angles were randomized between 10 and 80° (the angles between top and bottom lines was the same). For parameterization, we used 200 images per category (totalling 400 images for both long and short) to test for overall accuracy levels with a randomized line length difference between 2 and 62 pixels. To establish performance levels for the control set, we tested 200 images per pixel condition for each category i.e., 200 images at 10, 20, 30, 40, 50, and 60 pixel increment differences for both short and long.

The ML set was used to infer performance levels for images known to induce an illusory bias in humans. In this ML set, all top lines contained arrow-tails and all bottom lines contained arrowheads (see Row 3, Figure 3-3). Fin angles for ML images were fixed at 20 and at 40° in two separate conditions. Compared to our previous study, we removed the 60° condition because there was no bias effect present in the model. In this study, we were primarily interested in investigating the ML bias. At the C2 layer, we tested 200 images for each pixel condition within each category (totalling 1200 images for the short category at 10, 20, 30, 40, 50, and 60 pixel length increments and 1200 for the long category). For all other layers (Input, S1, C1, and S2), we tested 100 images per pixel condition within each category. In each case we took the average of 10 runs, randomizing the order of training images. Classification results for the input, S1 and C1 levels are based on deterministic operations, without

dependence on the weights developed during training. In these cases, randomizing the order of training images has no effect on classification results. To produce variation for these conditions, we generated additional test images that were randomized within the parameters specified above (with identical position ranges, fin angles, fin lengths, etc).

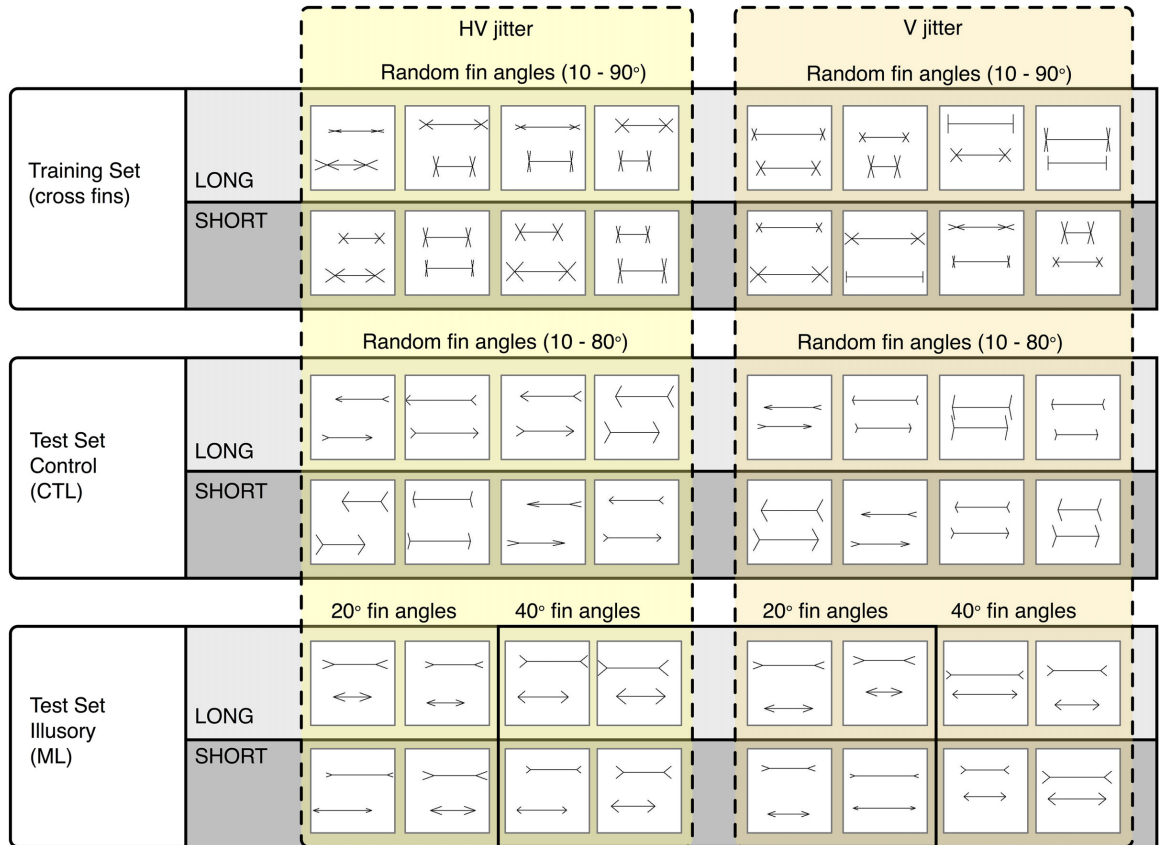


Figure 3-3 Representative sample of images categorized as LONG or SHORT. The Cross fin set (Top row) was used for training. The Control CTL set (Middle row) and Illusory ML set were both used for testing. Images are grouped into those that were jittered both horizontally and vertically (Left group) and those that were jittered only vertically (Right group).

3.3.3 Procedure: Learning, parameterization, illusion classification

Our method, established in Zeman *et al.* (2013), was carried out in three stages:

1. **Training.** Given a set of training images, a fixed-size network adjusted its internal weights during the learning process. In HMAX, both supervised learning and unsupervised learning are used. Unsupervised learning is first used to extract the most

informative features at the S2 layer (features at the S1 and C1 levels are fixed). Labeled data is then presented to the network to carry out the supervised learning phase, which adjusts weights within the network.

2. **Test Phase 1:** Parameterization. Using the CTL set, we ensured that the classifier was able to distinguish long from short images at an acceptable level of classification performance (above 85% correct), before testing with illusory stimuli. If performance fell below this level, we increased the size of the network and retrained (step 1). As shown in Figure 2-4, taken from Zeman et al. (2013), performance converges around 85-90% with a relatively small training image set (500 images for both V and HV conditions).
3. **Test Phase 2:** Illusion classification. Using the ML set, we established the discrimination thresholds and the magnitude of the illusion that manifested in the model.

3.4 Results

3.4.1 Experiment I: Classification of ML images after each level of HMAX

The aim of this experiment was to assess how simple and complex cell operations contribute toward bringing about the MLI. To this end, we examined the inner workings of HMAX, looking at classification performance for illusory images at each level of the architecture. We used a linear classifier to perform classification after each subsequent layer of HMAX, (which included processing of all previous layers required to reach that stage). Therefore, we ran classification on the Input only, on S1 (after information arrived from Input), on C1 (after information traversed through Input and S1 layers) and so on.

We first tested classification performance on our control images, which exceeded 85% when the size of the S2 layer was 1000 nodes. We acknowledge the HV condition would represent less learning than the V condition, given the same number of training images. We aimed to

keep the number of S2 nodes as well as the number of training images consistent across all conditions in order to make comparisons across these conditions. Using this network configuration, we tested classification on 20 and 40° ML images at the C2 level. We then tested classification at each layer of HMAX using the same illusory set.

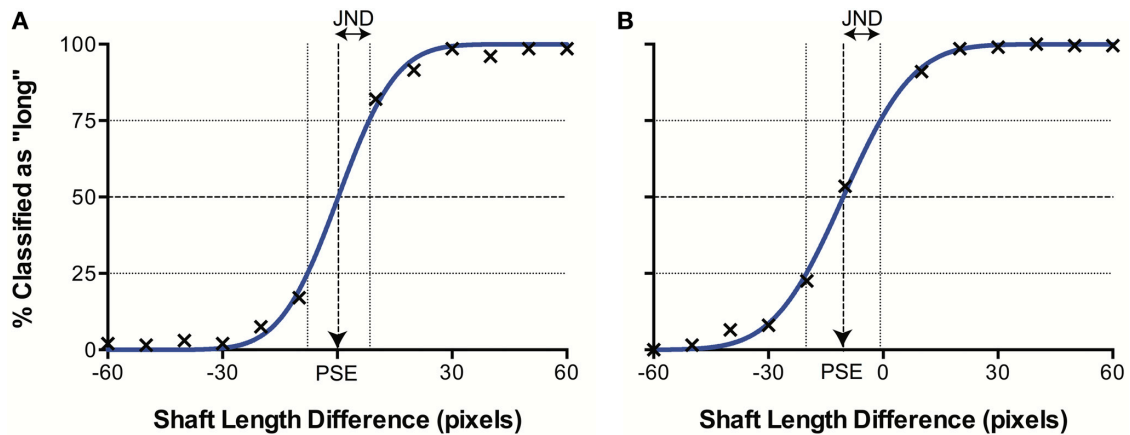


Figure 3-4 Example data sets from (A) CTL and (B) ML (40°) conditions. The best fitting curve (blue) allows derivation of the point of subjective equality (PSE) where classification is at 50%, and the just noticeable difference (JND), corresponding to the semi-interquartile difference.

When plotted in terms of the percentage of stimuli classified as “long” as a function of the difference in line length (top - bottom) for each separate data set (i.e., control images, illusory images with 20° fins and with 40° fins), we observed a sigmoidal psychometric function, characteristic of human performance in equivalent psychophysical tasks. The data were characterized by a cumulative Gaussian, with the parameters of the best fitting function determined using a least-squares procedure. Figure 3-4 illustrates an example data set. When Gaussian curves did not fit significantly better than a horizontal line at 50% (chance responding) in an extra sum of squares *F*-test, the results were discarded (2 runs out of a total of 52). This allowed us to determine the Point of Subjective Equality (PSE) the line length difference for which stimuli were equally likely to be classified as long or short (50%),

represented by the mean of the cumulative Gaussian. Here, PSEs are taken as a measure of accuracy, representing the magnitude of the Müller-Lyer Illusion manifested in the model. We also established the Just Noticeable Difference (JND) for each data set. The JND represents perceptual precision—the level of certainty of judgments for a stimulus type, and is indicated by the semi-interquartile difference of the Gaussian curve (the standard deviation multiplied by 0.6745). A higher JND represents greater uncertainty, and hence lower precision.

As can be seen in our results (see Figure 3-5A), the model produces a pattern of PSEs for illusory images consistent with human bias. We see a larger bias for more acute angles (20°) vs. less acute angles (40°), a pattern that is also consistent with human perception. This constitutes a replication of our previous findings (Zeman *et al.*, 2013) using a linear classifier, as opposed to a support vector machine (SVM), confirming that these findings are robust to the specific method of classification. These two trends are observable not only at the final C2 layer but at all levels of the architecture.

We observe that the illusion is present at the input level, suggesting that underlying statistical information may be present in our training images, despite careful design to remove bounding box cues and low spatial frequency information. The influence of image-source statistics on the Müller-Lyer illusion has already been studied using real world environmental images and an input layer bias is to be expected (Howe and Purves, 2005). Because the aim of our study is to explore the Müller-Lyer within a biologically plausible model of the visual ventral stream, we are more interested in how the network would process the input. Our novel contribution, therefore, is to focus on how such information is transformed in terms of changes in accuracy and precision layer to layer as we traverse the cortical hierarchy within the HMAX network.

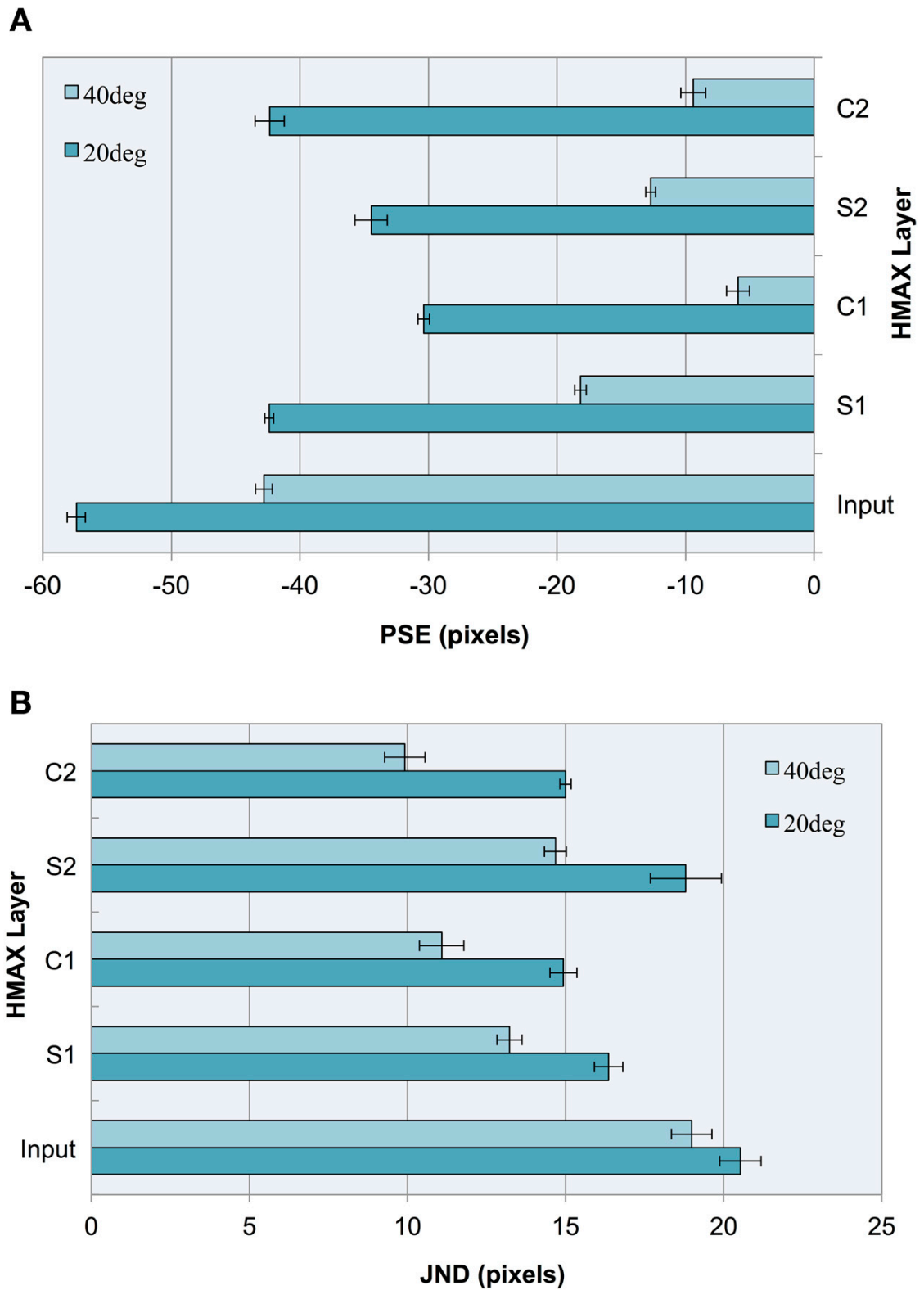


Figure 3-5 Experiment I results as a function of the HMAX layer for images with 20° and 40° fins. Error bars represent ± 1 s.e.m. across multiple runs. (A) Accuracy (PSEs). (B) Precision (JNDs).

Observing the PSE for each HMAX layer after a linear classifier is applied, this experiment demonstrates three key findings:

1. Running a linear classifier on the raw images revealed a bias at the input level that would represent statistical influences such as those proposed by Howe and Purves (2005). However, each layer of the HMAX architecture counteracts this bias producing a reduction in PSE magnitude after every S and C layer is traversed, when compared to the input layer.
2. In the majority of cases (87.5% of the time), illusory bias and uncertainty is reduced after complex cell operations have been applied. A reduction in uncertainty and bias can be seen when comparing the PSE and JND for S1 vs. C1 layers, for both 20 and 40° fin angles in the illusion set. Going from S2 to C2, PSE is reduced for 40° angles but not for 20° angles in the ML set, whereas JND is reduced for all cases.
3. When simple cell operations follow complex, illusory bias and uncertainty is increased. At the S2 layer, we see an increase in PSE and in JND for both 20° and 40° ML images.

The observations concerning accuracy data are echoed for precision. In Figure 3-5B, we see a higher JND (lower precision) for images with more acute fin angles at all levels of HMAX architecture. Looking at each layer of the architecture, we see lower JNDs (higher precision) at each level of HMAX compared to the input alone. We also observe higher precision (smaller JNDs) following processing by complex cells, but lower precision when the output from these layers is passed through a simple cell layer. In the case of results for precision, these observations held without exception.

Comparing these results directly to human data, Restle & Decker (1977) found that 20 degree angle fins would create an illusory bias of 26% and 40 degree angle fins would create a bias

of 23%. For our lines of 120 to 240 pixels, this would create an average PSE of 46.8 pixels for 20 degree fins and a PSE of 41.4 pixels for 40 degree fins. JND results were not reported by Restle & Decker (1977) and so a direct comparison between human results and HMAX cannot be made on this measure. However, it is possible to directly compare PSE at different layers of the HMAX model with human results. In this case, the S1 layer provides the best approximation of PSE for both 20 and 40 degree fins.

Considering the PSEs show in Figure 3-5A, an interesting observation is that ML figures with 40 degree angles present a "U" shape in traversing from the input layer to the C2 layer. This differs from the overall pattern presented by 20 degree angles that show an exponential shape when going from layer to layer. The main discriminating factor between the shapes of these 20 and 40 degree results is at the final C2 layer, where the PSE becomes larger in magnitude when going from S2 to C2 for 20 degree fins, yet decreases in magnitude for 40 degree fins. This demonstrates the influence of complex cell operations in reducing bias for 40 degree ML figures but not for 20 degree figures.

The contrast between results following processing by simple cell and complex cell layers encourages examination of the principal differences between the operations performed by these cells. The major distinction between S-layer and C-layer operations concerns the response to variance in the image. Unlike simple cells, whose outputs are susceptible to image variations such as fluctuations in the locations of features, complex cells' filtering properties allow them to respond similarly to stimuli despite considerable positional variance. When initially designing the training stimuli for HMAX, we wanted the system to build higher-level representations of short and long independent of line position, exact line length and of features appended to the shaft ends. This would require an engagement of complex cell functionality and less reliance on simple cell properties. To this end, we varied these parameters randomly in a controlled fashion to reduce reliance on trivial image details. If one

of our training parameters were to be restricted, the architecture would be less able to build such robust concepts of short and long. Given that complex cells are designed to pool information across simple cells with similar response properties and fire regardless of small changes in the afferent information, decreasing the variance in one of our training parameters would underutilize C cell properties and the short and long concepts within HMAX would become less flexible. This is likely to reduce the overall categorization performance of the computational model. More specifically, we hypothesize that restricting positional jitter to only one dimension would decrease accuracy and precision with which HMAX categorizes Müller-Lyer images. If this hypothesis holds true, we would demonstrate that greater positional variance reduces illusory bias and uncertainty. To seek further support for this proposition, we remove horizontal positional jitter from all stimuli in our second experiment.

3.4.2 Experiment II: HMAX classification of ML images with reduced variance

In our previous experiment, we observed a reduction in the level of bias after complex cell operations and hypothesized that introducing greater variance in the input would further reduce bias levels. To test this, we measured classification performance for HMAX layer C2 under two conditions: (1) Using our default horizontal and vertical jitter (HV) and (2) Under conditions of decreased positional jitter (V). We reduced the positional jitter in our training and test images from two- dimensional jitter in both the horizontal and vertical dimensions to one-dimensional, vertical jitter. While the top and bottom lines and their attached fins in our training and test sets remained independently jittered vertically (between 0 and 60 pixels), we removed all horizontal jitter, instead centering each stimulus. The vertical position of the top line was randomized between 48 and 108 pixels from the top of the image while the bottom line's vertical position was randomized between 148 and 208 pixels. We thus maintained a maximal 60 pixel jitter difference per line while limiting jitter to only one dimension.

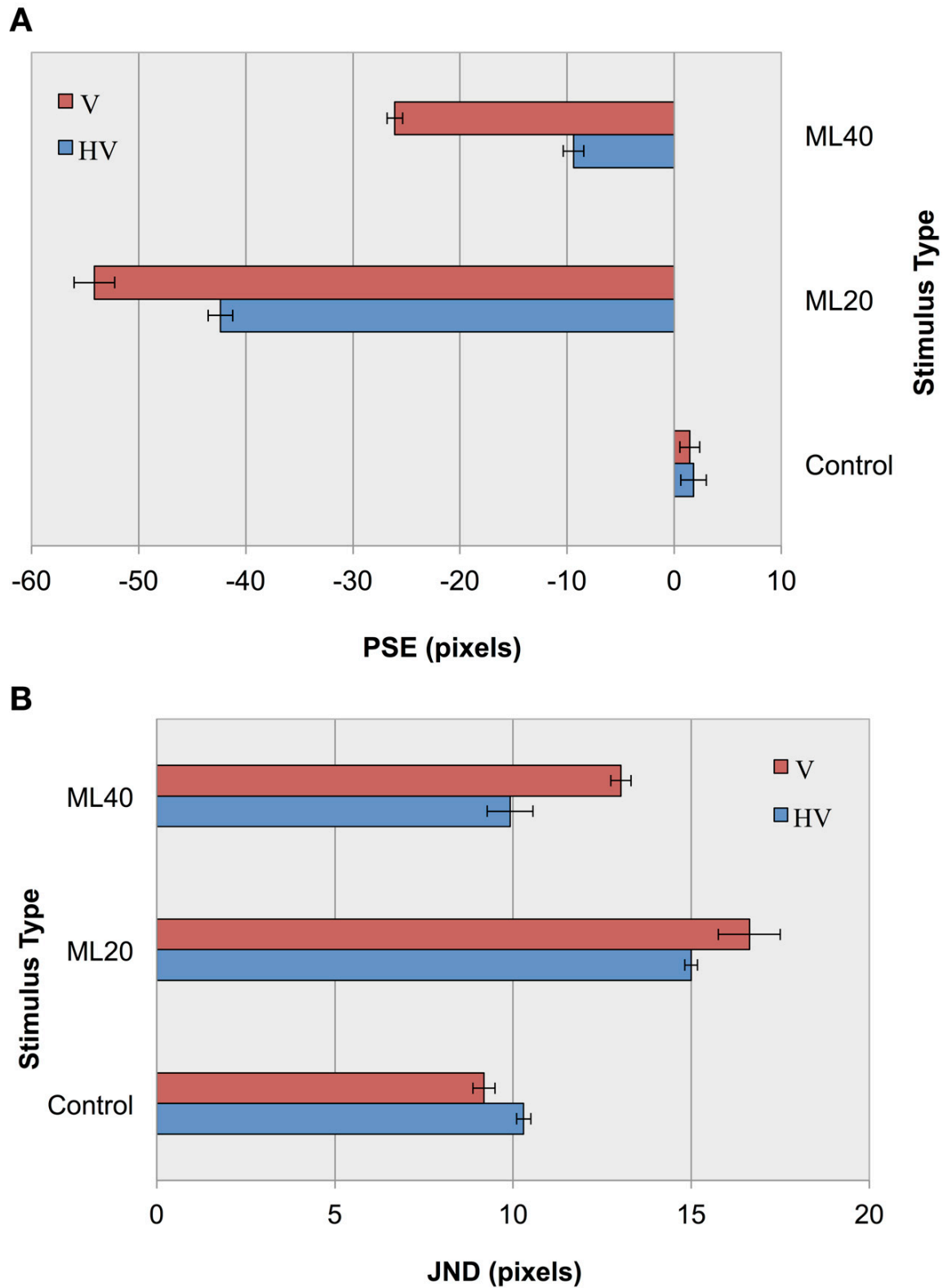


Figure 3-6 Experiment II results as a function of jitter type for control images, and Müller-Lyer images with 20° and 40° fins. (A) Accuracy (PSEs). (B) Precision (JNDs).

In an initial parameterization stage, we first tested performance using the CTL set, and found an overall classification score of 91.5% with an S2 size of 1000 nodes. The results of control and illusion image classification for our default jitter condition and for reduced positional jitter is shown in Figure 3-6. In terms of accuracy measurements (Figure 3-6A), it can be seen that for ML images PSEs are more extreme for V jitter only, compared to HV jitter. These results provide support for our hypothesis, demonstrating an increase in the magnitude of the Müller-Lyer effect for both 20 and 40° illusory conditions when reducing positional jitter, and hence image variance. The pattern of results for accuracy is echoed in terms of precision measurements (Figure 3-6B). Following the trend from our previous experiment, we see lower JND values for more obtuse angles compared to more acute angles. Comparing JND results for HV jitter with those for V jitter, we see that the classifier has higher precision when distinguishing short from long lines in the HV condition. In summary, decreasing the amount of positional variance in our stimuli increases bias and reduces the level of certainty in making decisions.

3.5 Discussion

Our aim for this study was to investigate the conditions under which the Müller-Lyer illusion manifests in HMAX and the factors that could influence the magnitude of the effect. Our primary motivation was to explore how hierarchical feature representation within HMAX affects classification performance. We ran two experiments performing binary image classification using HMAX. Images contained two horizontal lines that were jittered independently. Various configurations of fins were appended to the line shafts to create separate training and test images. Our first experiment compared the effects of operations performed by simple vs. complex cells by applying a linear classifier after each layer of HMAX when distinguishing long from short MLI images. Our second experiment examined

HMAX classification of MLI images with decreased positional jitter.

The main finding from our first experiment is that the addition of any simple or complex cell layers reduces bias, compared to classification directly made on the input images. Illusory bias changes from layer to layer within a simple-complex cell architecture, with increases in MLI magnitude as information passes through simple layers. In most cases, the effect decreases as information passes through complex layers. The pattern of results for accuracy is replicated when measurements of precision are considered. All levels of HMAX show improved precision compared to classified input images, with further JND reductions caused by complex cell layers, and increases caused by simple cell layers. Proposing that the C layers' property of invariant responding may underlie their ability to increase accuracy and precision, we hypothesized that decreasing variance in the input images and retraining the network would increase the MLI. We chose to decrease the positional variance by removing horizontal jitter and including only vertical jitter for the stimuli in our second experiment. Consistent with our hypothesis, experiment 2 showed an increase in illusion magnitude for both 20 and 40 degree angles.

In this paper and in our previous study, we focused solely on the ML illusion in its classical four-wing form. It would also be possible to study other variants of the Müller-Lyer and other illusory figures to test more generally for the susceptibility of hierarchical artificial neural networks. Some variants of the Müller-Lyer to be tested could include changing the fins to circles (the “dumbbell” version) or ovals (the “spectacle” version) (Parker and Newbigging, 1963). Other monocular line length or distance judgment illusions occurring within the visual ventral stream may also manifest in similar hierarchical architectures, for example, the Oppel-Kundt illusion (Oppel, 1854/1855; Kundt, 1863).

Some illusions are moderated by the angle at which the stimulus is presented (de Lafuente

and Ruiz, 2004). This raises the question whether illusory bias and uncertainty changes in classifying Müller-Lyer images that are presented diagonally, rotated by a number degrees to the left or to the right. Simple cells in HMAX consist of linear oriented filters, and are present in multiple orientations. The max pooling operations combine input from these and provide an output that is invariant to rotation. As a result, we would predict no difference in results when processing versions of the Müller-Lyer illusion in HMAX rotated at any arbitrary angle. This prediction is also consistent with human studies. While a number of illusions demonstrate an increase in magnitude when presented in a tilted condition, there is no difference in magnitude for the MLI (Prinzmetal and Beck, 2001).

In our last study, we recruited a previous version of HMAX known as FHLlib, a Multi-scale Feature Hierarchy Library (Mutch and Lowe, 2008). In the current study, a more recent, GPU-based version of HMAX, known as CNS: Cortical Network Simulator (Mutch *et al.*, 2010) was used. The main difference between these architectures was a linear classifier replacing the SVM in the final layer of the more recent code. The network setup between architectures was identical: one image layer followed by four layers of alternating S and C layers. Both had the same levels of inhibition (50% of cells in S1 and C1). The image layer contained 10 scales, each level $2^{1/4}$ smaller than the previous. Compared to our previous study, we were able to replicate similar levels of bias despite a change in the classifier, demonstrating that our result is robust and dependent upon properties of the HMAX hierarchical architecture, rather than the small differences between the implementation of these two related models.

Reflecting upon the implication of our results for other models, we would predict that those that have a similar hierarchical architecture would exhibit similar trends. That is, comparable networks would demonstrate increased bias with decreased precision when categorizing MLI images with less variance. Considering models that only contain filtering operations (akin to

layers of simple cells) we would observe an illusory effect that may also be exacerbated compared to those with more complex operations, with low accuracy and precision. Examples of would include the model of Bertulis and Bulatov (2001).

The reduction of bias in computer vision systems has significant ramifications for applications such as automated driving, flight control and landing, target detection and camera surveillance. Correct judgment of distances and object dimensions in these systems could affect target accuracy and reduce the potential for crashes and errors. Our hypothesis that increasing positional variance in the stimuli would reduce the magnitude of illusory bias could be extended to include other forms of variance, such as image rotation, articulation or deformation, hence examining the generality of this proposal. Furthermore, it would be informative to test the generality of the results presented in this study in other computational models. If a general effect could be confirmed, then we would advise the implementation of many forms of input variance during training to improve their judgment capabilities, providing more accurate and precise information.

Our work not only has implications for the field of computer science, but also for psychology. Computational models allow manipulations of parameters that are impossible or impracticable to perform in human subjects, such as isolating the contributions of different neural structures to the effect. Artificial architectures allow us to make predictions about overall human performance as well as how performance changes from layer to layer within the visual system. Considering that this model not only provides an overall system performance (C2 output), but also supplies information at multiple levels of the architecture that correspond approximately to identifiable neural substrates, it may be possible to test the model's predictions with neuroimaging data. Using functional magnetic resonance imaging (fMRI), we could obtain blood-oxygen-level dependent (BOLD) signals at different levels of the visual cortices of observers viewing the MLI compared to a control condition (using a similar

method to that described by Weidner and Fink, 2007). Then by applying a classifier to these signals, we could map this information to changes in model bias and quantify how well the model matches human brain data. This forms a possible direction for future research.

Funding

Astrid Zeman is supported by a CSIRO Top-Up Scholarship and the Australian Postgraduate Award (APA) provided by the Australian Federal Government. Astrid Zeman is also supported by the Australian Research Council Centre of Excellence for Cognition and its Disorders (CE110001021) <http://www.ccd.edu.au>.

Acknowledgements

We thank Dr. Kiley Seymour and Dr. Jason Friedman from the Cognitive Science Department at Macquarie University (Dr. Friedman now at Tel Aviv University, Israel) for helpful discussions. We thank Jim Mutch for making HMAX publicly available via GNU General Public License. We would like to thank Assistant Professor Sennay Ghebreab from the Intelligent Systems Lab at the University of Amsterdam for valuable input and suggestions. Lastly, we thank the reviewers for their helpful feedback, particularly with suggestions that we have incorporated into our discussion.

3.6 References

- Bertulis, A., and Bulatov, A. (2001). Distortions of length perception in human vision. *Biomedicine* 1, 3–26.
- Brigell, M., and Uhlarik, J. (1979). The relational determination of length illusions and length aftereffects. *Perception* 8, 187–197. doi: 10.1068/p080187
- Brown, H., and Friston, K. J. (2012). Free-energy and illusions: the cornsweet effect. *Front. Psychol.* 3:43. doi: 10.3389/fpsyg.2012.00043
- Coren, S., and Porac, C. (1984). Structural and cognitive components in the Müller-Lyer illusion assessed via cyclopean presentation. *Percept. Psychophys.* 35, 313–318. doi: 10.3758/BF03206334
- Corney, D., and Lotto, R. B. (2007). What are lightness illusions and why do we see them? *PLoS Comput. Biol.* 3:e180. doi: 10.1371/journal.pcbi.0030180
- de Lafuente, V., and Ruiz, O. (2004). The orientation dependence of the Hermann grid illusion. *Exp. Brain Res.* 154, 255–260. doi: 10.1007/s00221-003-1700-5
- Dewar, R. (1967). Stimulus determinants of the practice decrement of the Müller-Lyer illusion. *Can. J. Psychol.* 21, 504–520. doi: 10.1037/h0083002
- Fellows, B. J. (1967). Reversal of the Müller-Lyer illusion with changes in the length of the inter-fins line. *Q. J. Exp. Psychol.* 19, 208–214. doi: 10.1080/14640746708400094
- Ginsburg, A. (1978). *Visual Information Processing Based on Spatial Filters Constrained by Biological Data*. Ph.D. thesis, Aerospace Medical Research Laboratory, Aerospace Medical Division, Air Force Systems Command, Cambridge.
- Glazebrook, C. M., Dhillon, V. P., Keetch, K. M., Lyons, J., Amazeen, E., Weeks, D. J., et al. (2005). Perception–action and the Müller-Lyer illusion: amplitude or endpoint bias? *Exp. Brain Res.* 160, 71–

78. doi: 10.1007/s00221-004- 1986-y

Gregory, R. L. (1963). Distortion of visual space as inappropriate constancy scaling. *Nature* 199, 678–680. doi: 10.1038/199678a0

Howe, C. Q., and Purves, D. (2005). The Müller-Lyer illusion explained by the statistics of image–source relationships. *Proc. Natl. Acad. Sci. U.S.A.* 102, 1234–1239. doi: 10.1073/pnas.0409314102

Hubel, D. H., and Wiesel, T. N. (1959). Receptive fields of single neurones in the cat's striate cortex. *J. Physiol.* 148, 574–591.

Kundt, A. (1863). Untersuchungen über Augenmaß und optische Täuschungen. *Poggendorffs Annalen der Physik und Chemie* 120, 118–158. doi: 10.1002/andp.18631960909

Müller-Lyer, F. C. (1889). Optische Urteilstäuschungen. *Archiv für Anatomie und Physiologie* 2, 263–270.

Müller-Lyer, F. C. (1896a). Über Kontrast und Konfluxion. (Zweiter Artikel). *Zeitschrift für Psychologie und Physiologie der Sinnesorgane* 10, 421–431.

Müller-Lyer, F. C. (1896b). Zur Lehre von den optischen Täuschungen über Kontrast und Konfluxion. *Zeitschrift für Psychologie und Physiologie der Sinnesorgane* 9, 1–16.

Mutch, J., Knoblich, U., and Poggio, T. (2010). *Cns: A GPU-Based Framework for Simulating Cortically-Organized Networks*. Cambridge, MA: Technical Report MIT-CSAIL-TR-2010-013 / CBCL-286, Massachusetts Institute of Technology.

Mutch, J., and Lowe, D. G. (2008). Object class recognition and localization using sparse features with limited receptive fields. *Int. J. Comp. Vis.* 80, 45–57. doi: 10.1007/s11263-007-0118-0

Ogawa, T., Minohara, T., Kanaka, I., and Kosugi, Y. (1999). “A neural network model for realizing geometric illusions based on acute-angled expansion,” in *6th International Conference on Neural Information Processing (ICONIP '99) Proceedings*, Vol. 2 (Perth, WA: IEEE), 550–555.

- Oppel, J. (1854/1855). Über geometrischoptische Täuschungen (Zweite Nachlese). *Jahres-Bericht des Physikalischen Vereins zu Frankfurt am Main* 37–47.
- Parker, N. I., and Newbigging, P. L. (1963). Magnitude and decrement of the Muller-Lyer illusion as a function of pre-training. *Can. J. Psychol* 17, 134–140. doi: 10.1037/h0083262
- Predebon, J. (1997). Decrement of the Brentano Müller-Lyer illusion as a function of inspection time. *Perception* 27, 183–192. doi: 10.1068/p270183
- Prinzmetal, W., and Beck, D. M. (2001). The tilt-constancy theory of visual illusions. *J. Exp. Psychol. Hum. Percept. Perform.* 27, 206–217. doi: 10.1037/0096-1523.27.1.206
- Restle, F., and Decker, J. (1977). Size of the Mueller-Lyer illusion as a function of its dimensions: theory and data. *Percept. Psychophys.* 21, 489–503. doi: 10.3758/BF03198729
- Segall, M. H., Campbell, D. T., and Herskovits, M. J. (1966). *The Influence of Culture on Visual Perception*. New York, NY: The Bobbs-Merrill Company, Inc.
- Serre, T., and Poggio, T. (2010). A neuromorphic approach to computer vision. *Commun. ACM* 53, 54–61. doi: 10.1145/1831407.1831425
- Serre, T., Wolf, L., and Poggio, T. (2005). “Object recognition with features inspired by visual cortex,” in *Proceedings of 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005)* (San Diego, CA: IEEE Computer Society Press), 886–893. doi: 10.1109/CVPR.2005.254
- Weidner, R., Boers, F., Mathiak, K., Dammers, J., and Fink, G. R. (2010). The temporal dynamics of the Müller-Lyer illusion. *Cereb. Cortex* 20, 1586–1595. doi: 10.1093/cercor/bhp217
- Weidner, R., and Fink, G. R. (2007). The neural mechanisms underlying the Müller-Lyer illusion and its interaction with visuospatial judgments. *Cereb. Cortex* 17, 878–884. doi: 10.1093/cercor/bhk042
- Zeman, A., Obst, O., Brooks, K. R., and Rich, A. N. (2013). The Müller-Lyer illusion in a

computational model of biological object recognition. *PLoS ONE* 8:e56126. doi: 10.1371/journal.pone.0056126

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

4 Study 3

An exponential filter model predicts lightness illusions

Abstract

Lightness, or perceived reflectance of a surface, is influenced by surrounding context. This is demonstrated by the Simultaneous Contrast Illusion (SCI), where a grey patch is perceived lighter against a black background and vice versa. Conversely, assimilation is where the lightness of the target patch moves toward that of the bounding areas and can be demonstrated in Whites Effect. Blakeslee and McCourt (1999) introduced an oriented difference-of-Gaussian (ODOG) model that is able to account for both contrast and assimilation in a number of lightness illusions and that has been subsequently improved using localized normalization techniques. We introduce a model inspired by image statistics that is based on a family of exponential filters, with kernels spanning across multiple sizes and shapes. We include an optional second stage of normalization based on contrast gain control. Our model was tested on a well-known set of lightness illusions that have previously been used to evaluate ODOG and its variants, and model lightness values were compared with typical human data. We investigate whether predictive success depends on filters of a particular size or shape and whether pooling information across filters can improve performance. The best single filter correctly predicted the direction of lightness effects for 21 out of 27 illusions. Combining two filters together increased the best performance to 23, with asymptotic performance at 24 for an arbitrarily large combination of filter outputs. While normalization improved prediction magnitudes, it only slightly improved overall scores in direction predictions. The prediction performance of 24 out of 27 illusions equals that of the best performing ODOG variant, with greater parsimony. Our model shows that V1-style orientation-selectivity is not necessary to account for lightness illusions and that a low-level model based on image statistics is able to account for a wide range of both contrast and assimilation effects.

4.1 Introduction

Lightness is the perceived reflectance of a surface, which can vary greatly according to surrounding context, as demonstrated by lightness illusions (see Kingdom (2011) for a recent review). One clear and well-known example is the Simultaneous Contrast Illusion (SCI), where a grey target patch is perceived as lighter when surrounded by a black background and darker when surrounded by a white background (Chevreul, 1839) (Figure 4-1 left). The SCI demonstrates the contrast phenomenon, where lightness shifts away from surrounding luminance values, luminance being the amount of light that reaches the eye. Under other circumstances, lightness can shift towards the luminance values of bordering areas – a phenomenon known as assimilation¹. This is effectively demonstrated by a version of White's Illusion (White, 1979), where the test patches are not as wide as they are tall (Figure 4-1 right).

Theories that aim to explain lightness illusions can be broadly categorized into low-level and higher-level accounts. Higher-level theories argue that scene interpretation is necessary to account for lightness illusions, where cortical processing of surface curvature, depth and transparency are known to influence perceived reflectance (Knill and Kersten, 1991). For instance, Schirillo et al. (1990) demonstrated that lightness perception is dependent upon depth cues. Given that depth perception is thought to be a cortical function, higher-level areas must be recruited when perceiving reflectance. In 1999, Gilchrist et al. (1999) established the Anchoring Theory of lightness, where perceived reflectance of a patch is “anchored” to the highest luminance value within the retinal image (global information) and is also “anchored” to luminance values in surface groups that share commonalities such as being situated within the same depth plane (local information). Another notable high-level theory is Anderson

¹ In some cases, target patches have equal bordering white and black areas, making it difficult to distinguish whether a contrast or assimilation effect is predominantly present.

(1997)'s Scission Theory, based upon the principle that a visual scene is split into different causal layers of reflectance, transparency and illumination (the amount of light incident on a surface), to determine the surface properties of a homogenous area. While these high-level theories are able to offer consistent explanations for a variety of complex lightness phenomena, our aim in this paper is to quantify the performance of low-level models whose computations do not require higher-level scene interpretation. In the interests of providing a succinct quantitative account of a range of lightness phenomena, we apply Occam's Razor, emphasizing the capability of low-level theories to deliver improved modeling precision with greater parsimony.

Low-level theories concentrate on filtering operations and statistical image properties as the key explanation behind many lightness illusions. The main principle underlying low-level theories is that of image reconstruction: that lightness is inferred by reconstructing the most probable source image using filtering operations (Blakeslee and McCourt, 1999; Dakin and Bex, 2003). The filters concerned are considered to reside in early stages of the visual hierarchy such as the retina, LGN and/or V1. Blakeslee and McCourt (1997) designed a low-level model using a multi-scale array of two-dimensional Difference of Gaussian filters (DOG) with responses that were pooled and normalized. The isotropic filters in this model approximated retinal ganglion or LGN single cell function. The DOG model was able to account for the contrast effect shown in the SCI but not the assimilation observed in White's Effect. To account for assimilation, Blakeslee and McCourt (1999) extended this model to include anisotropic filters (oriented difference of Gaussians, or ODOG filters) that were pooled non-linearly. These orientation selective filters best approximate V1 functions, shifting the focus of the model from pre-cortical to cortical operations to account for a larger set of lightness illusions. Shortly after this, Dakin and Bex (2003) introduced an isotropic filter model that reweighted filter outputs using spatial frequency (SF) properties found in image

statistics. Using a series of center-surround, Laplacian of Gaussian filters, they demonstrated that low SF structure is an essential ingredient of two well-known lightness illusions: White's Effect and the Craik–Cornsweet–O'Brien Effect (O'Brien, 1958; Craik, 1966; Cornsweet, 1970). Dakin and Bex (2003) demonstrated that orientation selective filters were not required to successfully model assimilation effects, and highlighted the importance of weighting or normalization schemes within these low-level models.

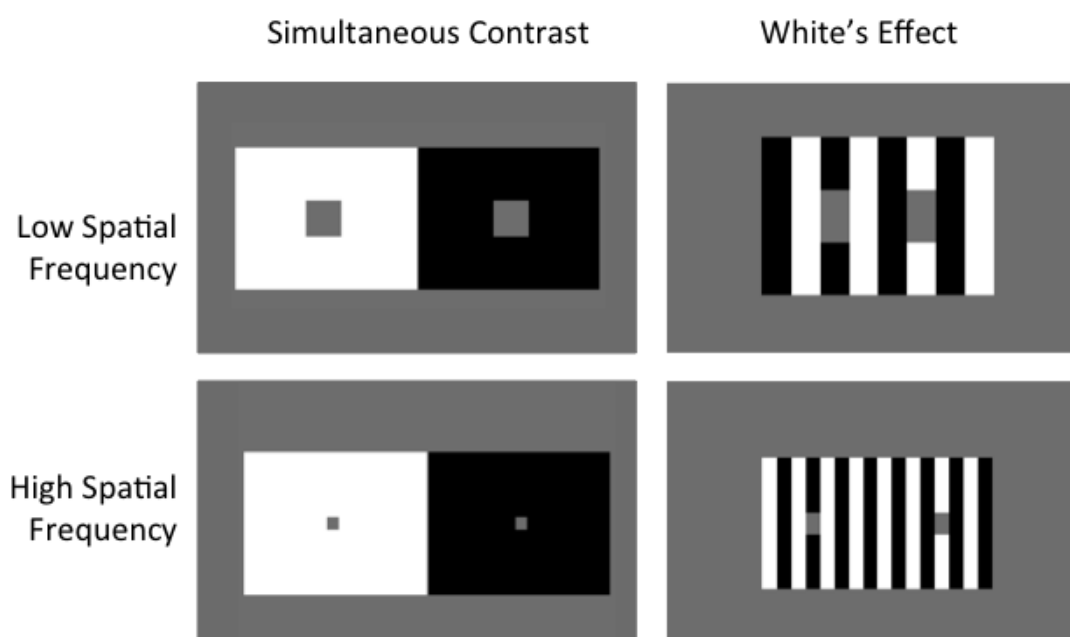


Figure 4-1 Simultaneous Contrast versus White's Effect. Albedo of grey target patches in Simultaneous Contrast shift away from background, demonstrating contrast. Targets in White's Effect shift towards surrounding context, demonstrating assimilation. Increasing spatial frequency increases the effect in both cases.

Since Dakin and Bex's paper, focus on statistical image properties (Corney and Lotto, 2007) and on post-filtering operations that weight the relative filter outputs (Robinson et al., 2007) has intensified in the context of low-level lightness models. Corney and Lotto (2007) demonstrated contrast and assimilation effects using an approach inspired by image statistics, training an artificial neural network with virtual scenes that possess naturalistic structure. In

contrast to Dakin and Bex (2003) who made statistical relationships explicit through weighting operations, Corney and Lotto (2007) trained an artificial neural network to implicitly learn the relationships between images and their underlying statistics. In the same year, Robinson et al. (2007) focused on applying different normalization schemes to improve predictions using the ODOG model. Normalization is commonly used as a weighting scheme to smooth distributions and scale all values to a baseline magnitude (usually 1). Robinson et al. (2007) focused on applying two different normalization schemes to the ODOG model: local-normalization of filter outputs (LODOG) and spatial frequency-specific local normalization (FLODOG). In LODOG and FLODOG, parameters of the normalization function (such as normalization window size) were adjusted to produce different model predictions. Robinson et al. (2007) systematically tested ODOG, LODOG and FLODOG on a catalog of 28 stimuli, 27 of which are known to induce illusions of contrast or assimilation in human observers. While ODOG was able to predict only 13 illusions in the correct direction, the best performing LODOG model was able to predict 18. FLODOG proved the most effective, correctly predicting 24 lightness illusions with an optimal parameter set.

Here we extend the literature using an approach inspired by natural image statistics. As established by Dakin and Bex (2003), the underlying distribution of structural properties present in natural images can greatly influence lightness judgments. Natural images share common underlying statistics, regardless of their origin (Zhu and Mumford, 1997a,b). For example, contrast histograms for natural images are skewed towards lower contrasts and have an exponential tail (Field, 1987; Ruderman and Bialek, 1994). Basu and Su (2001) investigated filters that encode the distribution of contrasts over different spatial frequencies. They concluded that exponential distributions provide a better fit than the Gaussian kernels that have been used in the models described above. By employing exponential filters of different sizes and shapes within a computational model, we represent the profile of contrast

statistics present in natural images and observe how these may influence the direction and magnitude of a set of lightness illusions. These filters have x and y-axis symmetry, ranging from ridged, “peaky” distributions to flatter, more rounded distributions (illustrated in Figure 4-2).

The exponential filters we explore in this study are offered as another kind of inhibitory mechanism, since the image filtered by the exponential function is subtracted from the original image. As such, this model shares much in common with other filtering approaches, such as ODOG (Blakeslee and McCourt, 1999). Indeed, this filtering approach bears similarity to the extra classical surround model of Ghosh et al. (2006) and is most similar to the filtering approach of Shapiro and Lu (2011), with the exception of the shape of the surround.

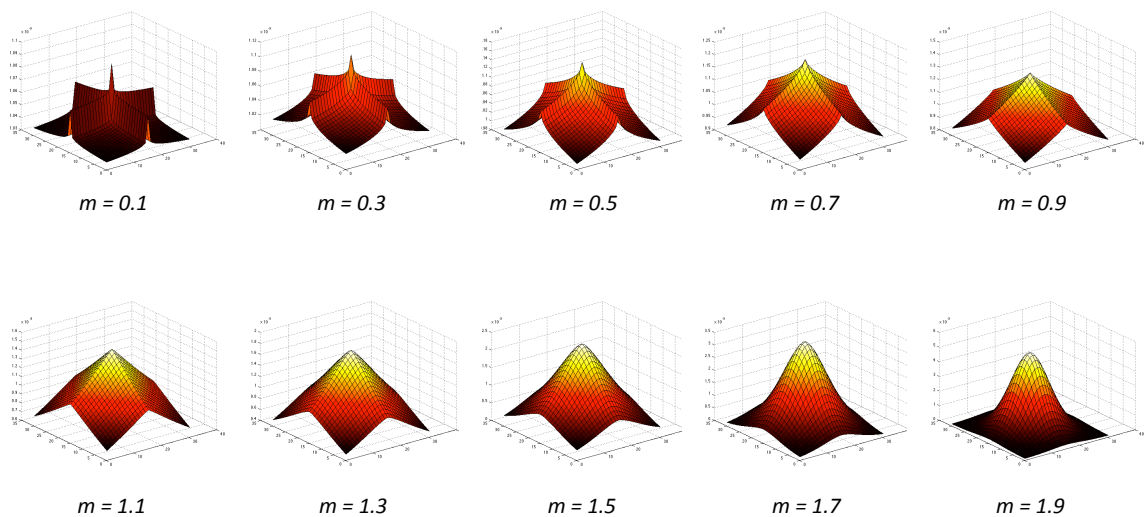


Figure 4-2 Exponential function family, taken from Basu & Su (2001) with increasing values of the m exponent.

While the filters in ODOG (and variants) approximate the functioning of orientation-selective V1 cells, and while Difference or Laplacian of Gaussian filters approximate the operations of isotropic LGN or retinal ganglion cells, exponential filters, not unlike those forming the basis of our model, have been identified in H1 horizontal retinal cells (Packer and Dacey, 2002, 2005). Our model is predominantly motivated by the computer vision literature, where exponential filters have been shown to be excellent edge detectors as well as resilient to noise (Zhu and Mumford, 1997a). The level of biological plausibility in our model is not strongly emphasized, but we do identify possible neurobiological equivalents to the filters that we apply. Geisler (2008) illustrates responses to natural images of a sensor that has a receptive field profile similar to V1, where an exponential function shows a better fit over a Gaussian distribution. While there are parallels here in demonstrating that an exponential fit is better than Gaussian in terms of filter responses, the filters that we apply are not oriented V1-style filters. Therefore, we would not suggest any relationship between our model results and the involvement of cortical neurons.

Our study differs from that of Corney and Lotto (2007) in that we make statistical relationships explicit through filtering and normalization operations, instead of training an artificial neural network to implicitly learn the relationships between images and their underlying statistics. Our method is similar to that of Dakin and Bex (2003), in that we both capitalize on the properties of image statistics to reconstruct the final image. In our method, we employ exponential shape filters that are based on image statistics. In Dakin and Bex (2003), the authors split an image into different spatial frequencies (SFs) using band-pass filters. The distribution of SFs was then reweighted to match that which occurs in natural scenes. In our model and in that of Dakin and Bex (2003), the filters are designed to extract the most salient features while being robust to noise (Basu and Su, 2001). In this way, both of our studies align with the predictive coding principle by Srinivasan *et al.* (1982) - that by

exploiting the spatial correlations of natural scenes, early visual systems are much better able to handle noise in the environment.

In the current study, we set out to investigate how well an exponential model is able to predict human data in response to a large battery of 28 lightness illusions previously used to test ODOG and its derivatives (Blakeslee and McCourt, 1999, 2001, 2004; Blakeslee et al., 2005; Robinson *et al.*, 2007). We apply exponential filters with a range of different shapes and sizes to an input image, with and without normalization of varying spatial extent. The outputs of this model are taken as predictions of perceived lightness both for single filters and for multiple-filter combinations.

4.2 Material & Methods

4.2.1 Stimuli

A standard battery of 28 figures known to produce particular lightness effects was used as a stimulus set in this study (see Robinson et al., 2007). Each stimulus (with the exception of the Benary Cross) involves a pair of uniform, mean luminance target patches, each surrounded by details with the opposite contrast polarity. Stimuli are illustrated in Figure 4-3 (reproduced from Robinson et al., 2007). All stimuli are 512 x 512 pixels in size. Each stimulus is listed below in Table 1 with original sources and comparative results reported for human responses where available. Table 1 also includes the reported illusion direction by humans as the patch perceived as the lightest within the image and the corresponding classification of the predominant effect as contrast or assimilation.

Table 1 Stimuli with original sources, reproduced results (for strength comparison) and illusion direction reported by humans

Figure	Original Source	Reproduced Results	Human Direction	Contrast (C) or Assimilation (A)?
a	White (1979)	Blakeslee and McCourt (1999)	Left	A
b	White (1979)	Blakeslee and McCourt (1999)	Left	A
c	Robinson et al. (2007)		Top	A
d	Anderson (2001)	Blakeslee et al. (2005)	Right	A
e	Howe (2001)	Blakeslee et al. (2005)	No illusion	N/A
f	Clifford and Spehar (2003)		Left	A
g	Anstis (2003)		Bottom	A
h	Anstis (2003)		Bottom	A
i	Anstis (2003)		Bottom	A
j	Anstis (2003)		Bottom	A
k	Howe (2005)		Right	A
l	Howe (2005)		Right	A
m	Howe (2005)		Right	A
n	McCourt (1982)	Blakeslee and McCourt (1999)	Area between black	C
o	Chevreul (1839)	Blakeslee and McCourt (1999)	Right	C
p	Chevreul (1839)	Blakeslee and McCourt (1999)	Right	C
q	Pessoa et al. (1998)	Blakeslee and McCourt (1999)	Left (Right in original)	C
r	Todorovic (1997)	Blakeslee and McCourt (1999)	Right	A
s	Todorovic (1997)	Blakeslee and McCourt (1999)	Right	N/A
t	Pessoa et al. (1998)	Blakeslee and McCourt (1999)	Right	A
u	De Valois and De Valois (1988)	Blakeslee and McCourt (2004)	Right	A
v	De Valois and De Valois (1988)	Blakeslee and McCourt (2004)	Right	A
w	De Valois and De Valois (1988)	Blakeslee and McCourt (2004)	Left	C
x	Adelson (1993)	Blakeslee and McCourt (2001)	Bottom	C
y	Benary (1924)	Blakeslee and McCourt (2001)	Left	N/A
z	Todorovic (1997)	Blakeslee and McCourt (2001)	Second in 1-2 Fourth in 3-4	N/A N/A
aa	Bindman and Chubb (2004)		Left	A
bb	Bindman and Chubb (2004)		Left	A

The majority of images exhibit assimilation effects, with contrast effects demonstrated by figures n, o, p, q, w, and x. In some cases, target patches have equal bordering white and black areas, making it difficult to establish whether a lightness effect should be defined as a contrast or assimilation effect (as in stimulus s). Stimuli y and z demonstrate opposing illusion directions for patches with identical bordering surrounds, presenting both contrast and assimilation effects simultaneously. In most cases, illusory effect directions reported in the original articles have been replicated in follow-up studies by Blakeslee and McCourt (used here and in Robinson *et al.* (2007) for direct strength comparisons). However, due to slight differences in methodology, stimuli (e) and (q) demonstrate discrepancies between the two

sets of human data. In these cases, we follow the convention of Robinson *et al.* (2007) to allow for easy comparison between their models and those described here.

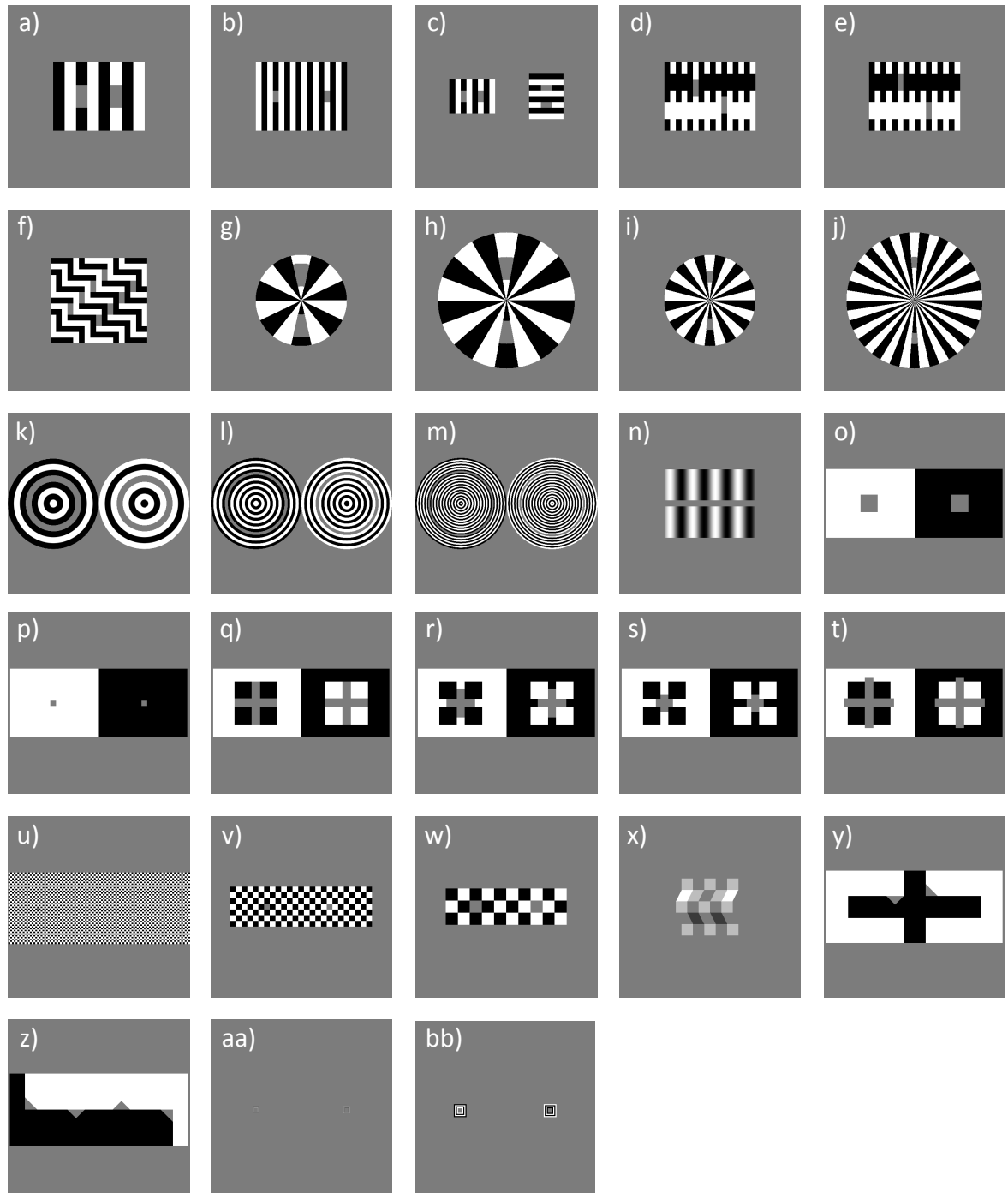


Figure 4-3 Illusions tested, replicated from Robinson *et al.* (2007)

As each stimulus involves 2 (or more) uniform, mean luminance target patches, each surrounded by details with the opposite contrast polarity, the lightness effects observed on these patches are expected to be equal and opposite. Our model's predictions regarding the presence of contrast or assimilation effects are made by taking mean lightness values from the largest rectangular patch inside the bounds of the target areas (matched for size) and subtracting the values for the patch that appears darker from those for the lighter. For stimulus n, ("grating induction"), we select rectangular areas that are 26 pixels wide to the left and right of center for our prediction comparison (0.4 of the spatial period of the grating), while maintaining the same patch height as Robinson *et al.* (2007).

4.2.2 Model

Our model consists of two-stages: 1) linear filtering using exponential functions 2) non-linear divisive normalization by coefficient of variation. Although the details of each stage may vary, this linear-nonlinear modeling method is commonly used to model physiology (Nykamp and Ringach, 2002; Schwartz and Simoncelli, 2001). Once the two stages of the model have produced lightness values at each pixel location of each target patch, we produce a prediction by calculating the mean difference over the target patches and applying linear scaling. Details of each step in the model and on calculating the comparison metric are described below.

4.2.2.1 Filtering

The set of exponential filters we apply are taken from Basu and Su (2001). These exponential filters are two dimensional in shape and possess x-symmetry, y-symmetry and symmetry with respect to the origin. They take the form:

$$g(x) = \frac{1}{K_1} \exp^{-K_2|x|^m} \quad (1)$$

where K_1 , K_2 and m are all positive constants. The m exponent corresponds to the shape of the filter. The normalization or scaling factor K_1 is calculated using K_2 and m as follows:

$$K_1 = (1/K_2^{1/m})(1/m)\Gamma(1/m) \quad (2)$$

where constant K_2 is a function of the variance of $g(x)$, which denotes the size of the filter.

$\Gamma(x)$ is the Gamma function defined as:

$$\Gamma(x) = \int_0^{\infty} t^{(x-1)} \exp(-t) dt \quad (3)$$

Figure 4-2 illustrates the variety of exponential filter shapes. When m is small, the exponential filter is described as having “high kurtosis”, showing a sharper peak with more prominent ridges. When m is large the exponential filter has “low kurtosis”, being flatter and rounder with smoother ridges. A special case is formed when $m = \frac{1}{2\sigma^2}$, where the function becomes a Gaussian with added rotational symmetry.

Each filter of a specific size and shape is applied to every pixel within the image. The size of the filter affects the information that is gleaned from an image. Smaller filters (high spatial frequencies or SFs) show better responsiveness but are less resilient to noise. Larger filters (low SFs) blur a lot of information, essentially losing information present in the images, but cope better with noise. There is a trade-off between selecting precise information and having greater resilience to noise, which is where scale selection comes in. The most appropriate filter selection finds the right compromise between these two factors, taking the smallest scale with the most reliable response.

A small amount of Gaussian noise is added to the image (0.1%) before filtering. Adding noise to the image is to avoid divide-by-zero errors when implementing divisive normalization. We

are aware of other approaches to avoid divide-by-zero errors, such as adding a constant to the denominator term (Cope et al., 2013).

Responses are then convolved to create a filtered image of the same dimensions as the original input. The filtered convolved image is subtracted from the original image as the final step in processing. We explore a range of different filter shapes and sizes and produce a set of filtered images for every size and shape of filter. We use 10 filter sizes ranging from 5 pixels to 95 in increments of ten. The filter shapes range from 0.1 to 1.9 in increments of 0.2. Figure 4-4 illustrates the result of applying three example filters with different shape parameters to White's Illusion. . The predictive success of this particular filter size is well-demonstrated for this particular image, regardless of filter shape. The bottom row in Figure 4-4 demonstrates a close approximation to the Gaussian filter, which in this case is able to predict the direction and magnitude of White's Effect. This filter differs from the DOG filters used by Blakeslee and McCourt (1997)'s model in two key ways. Firstly, Blakeslee and McCourt use a Difference- of- Gaussian (DOG) filter, rather than an approximate Gaussian pictured here. Secondly, Figure 4-4 demonstrates a single filter operation, rather than a bank of filters used by Blakeslee and McCourt (1997).

4.2.2.2 Normalization (optional)

After applying a specific contrast filter with shape m and size K_2 to each pixel location in the image, we optionally normalize the filter outputs. Normalization is not only useful in its primary function of constraining the dynamic response range of image filters, but is also beneficial for generating a faithful representation of image contrast. Following Bonin et al. (2005), at each image location we divisively normalize the linear filter output by the output of a suppressive field, which computes the statistics of filter outputs surrounding the image location of interest. Bonin *et al.* (2005)'s normalization method, referred to as contrast gain

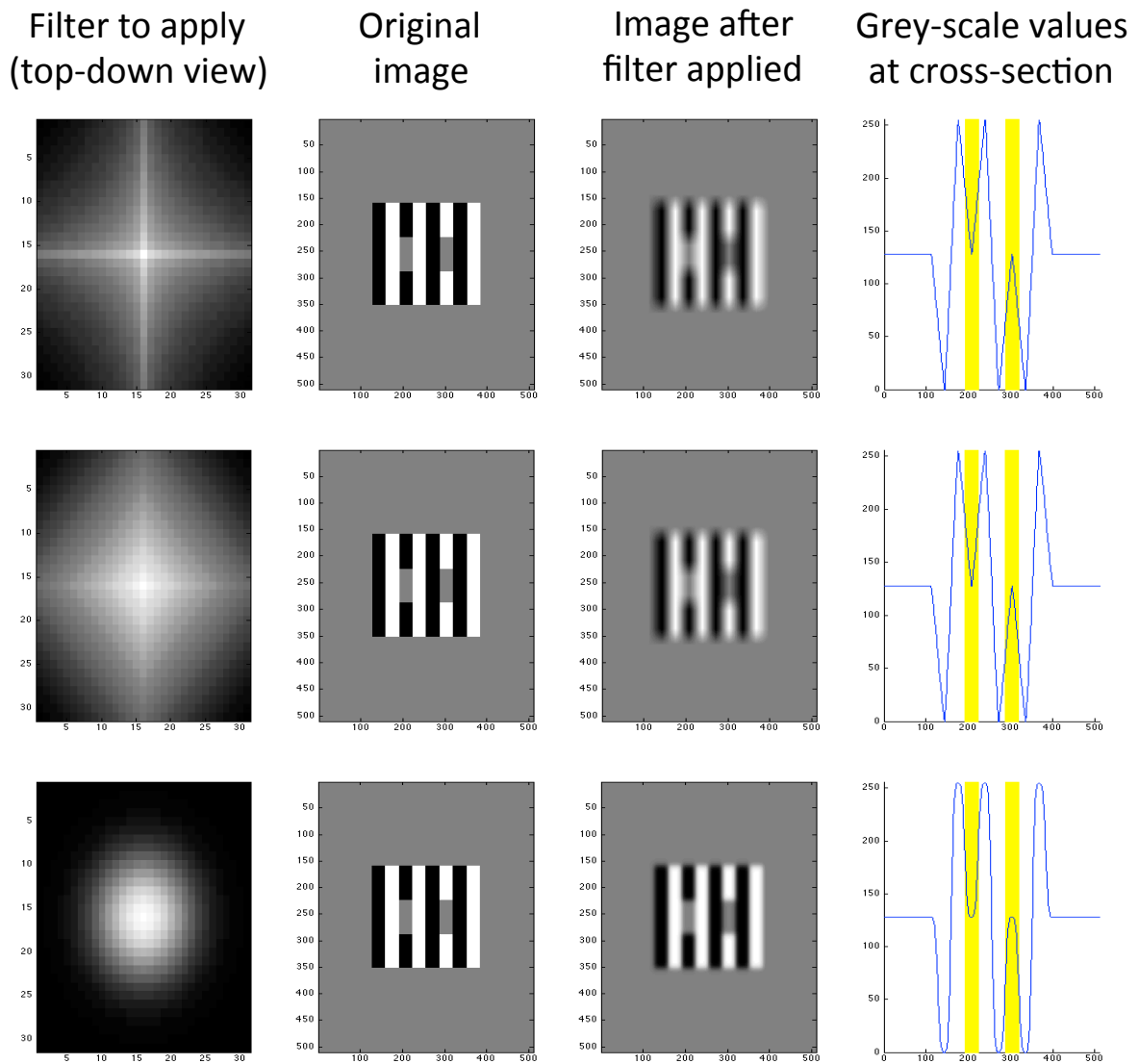


Figure 4-4 Exponential filters applied to White's illusion, all with size $K_2 = 5$. The top row shows a filter with high kurtosis ($m = 0.5$), the middle row shows a medium kurtosis filter ($m = 1.0$) and the bottom row shows a low kurtosis filter ($m = 2.0$). From left to right, column 1 is a top-down view of the filter shape, column 2 is the original image (of size 512 x 512 pixels), column 3 is the same image filtered and column 4 is a cross section of greyscale values through row $y=250$ pixels (where 0 represents black and 255 represents white). The locations of target patches are highlighted yellow in the final column.

control, is closely related to that found in the LGN and so we apply it here as a biologically plausible method for normalization in pre-cortical areas. In contrast to Bonin *et al.* (2005), who take the local root-mean-square contrast as the suppressive field, we divide filter

responses by the local coefficient of variation. The local coefficient of variation is inversely related to local Weibull statistics and as such is diagnostic of local image structure. Divisive normalization by the local coefficient of variation amplifies local image contrast. Similarly to Bonin *et al.* (2005), we compute normalized filter outputs using the following formula:

$$V = V_{max} \frac{g(x)}{c_{50} + c_{local}} \quad (4)$$

where c_{50} determines the strength of the suppressive field, V_{max} is the maximum response of the filter to the image, and $g(x)$ is the filtered response defined above. Finally, c_{local} is the local coefficient of variation:

$$c_{local} = \frac{\sigma}{\mu} \quad (5)$$

c_{local} is calculated based on the mean (μ) and the size of the suppressive field (σ) that is used as one of the parameters in our normalization step. The σ parameter specifies the size of the suppressive field compared to the size of the receptive field. When $\sigma = 1$, the size of the suppressive field is equal to that of the receptive field. When $\sigma = 2$, the size of the suppressive field is twice that of the receptive field.

4.2.2.3 Analysis Metrics

For each stimulus that we analyze, we take the resultant values (denoted as R) from the filter-only output (step 1) or from normalized output (step 2) with either $\sigma = 1$ or $\sigma = 2$ (as described above, σ represents the size of the suppressive field, as a proportion of the receptive field). We refer to $\sigma = 1$ as short-range normalization, where the suppressive field is the same area as the receptive field. $\sigma = 2$ is referred to as long-range normalization, where the suppressive field is twice the size of the receptive field. Within each image, we compare

values over the two areas that have been assigned to be target patches (see section 4.2.1). The lighter patch (as established in human experiments) is assigned to be patch A and the darker patch is assigned to be patch B. Mean values are obtained for both target patches before the mean of patch B is subtracted from the mean of patch A. Because patch A is assigned to be the lighter patch, a prediction in the correct direction is indicated by a positive value, whereas an incorrect prediction is negative. A value of zero indicates no difference in patch lightness values and therefore no illusion.

To compare resultant values, we scale the difference between target patches to the strength of White's Illusion for ease of comparison. The magnitude of White's illusion is denoted as R_a . This means that all resultant values are scaled to the strength (or magnitude) of White's illusion. A resultant value of 1 is then interpreted as having identical illusory strength to White's illusion. A value greater than 1 indicates the illusion is stronger than White's, and a value less than 1 (and above 0) indicates the illusion is weaker than White's. Although any stimulus could have been selected for comparative purposes, we follow Robinson *et al.* (2007)'s convention by selecting stimulus a as our comparative figure.

$$R = (\bar{A} - \bar{B}) ./ |R_a| \quad (6)$$

We also calculate the difference between model predictions and human results (where available) to quantify how well different model configurations match human data. We do this by subtracting the human result R_{human} from the model result R_{model} for stimuli from a to bb for which human results are available, and calculating the root mean square error ($\text{RMS}_{\text{error}}$). The smaller the $\text{RMS}_{\text{error}}$ value, the better the model matches human data, and the greater the predictive accuracy of the model in terms of illusion magnitude or strength.

$$RMSE_{error} = \sqrt{\frac{1}{n} \sum_a^{bb} (R_{model} - R_{human})^2} \quad (7)$$

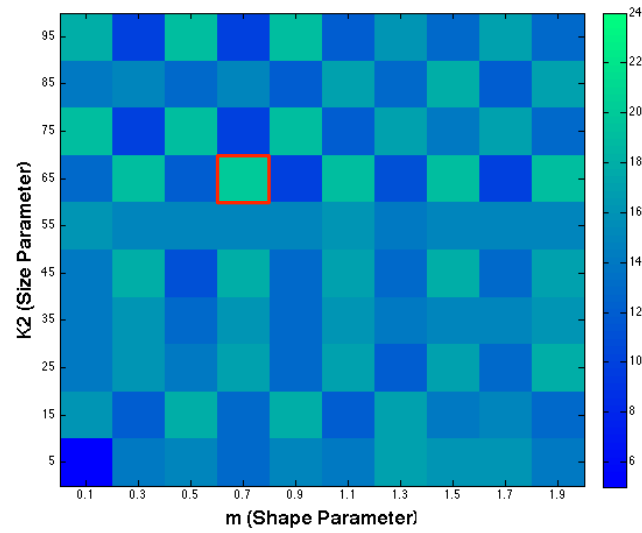
When combining the outputs of two filters α and β of different sizes or shapes, we simply sum the difference in mean responses to the light and dark patches separately for each filter (removing scaling to figure a):

$$R^{\alpha\beta} = R^\alpha + R^\beta \quad (8)$$

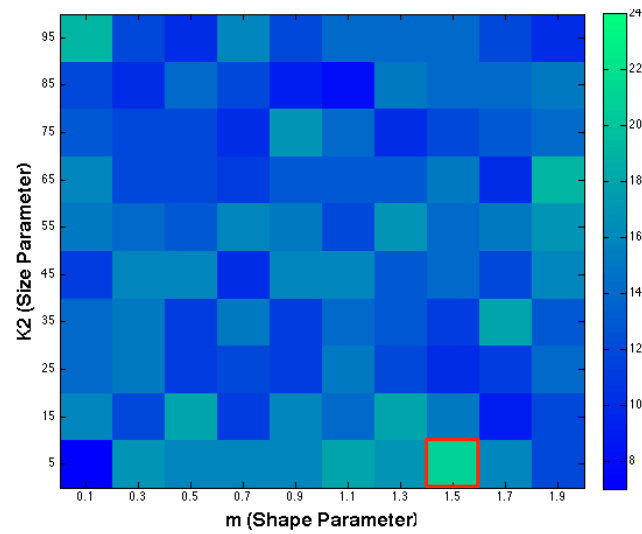
4.3 Results

We assess the performance of our model in two ways: the number of predictions in the correct direction, and also how closely the predicted values match the scaled human data on illusion magnitude. We exclude figure e from our analysis, given that no illusion direction is reported for humans. Figure 4-5 illustrates the number of illusion directions correctly predicted (out of a maximum possible of 27) using a single filter over a range of 10 filter shapes and 10 filter sizes. For figure z, there are two predictions, annotated as z_{2-1} and z_{4-3} , for comparing the two left patches and the two right patches in the image respectively. We take a correct result to be when $(z_{2-1} + z_{4-3}) / 2 > 0$.

Not normalized
 Best Result: 20
 $K_2 = 65, m = 0.7$



Normalized, $\sigma = 1$
 Best Result: 21
 $K_2 = 5, m = 1.5$



Normalized, $\sigma = 2$
 Best Result: 19
 $K_2 = 85, m = 0.7$

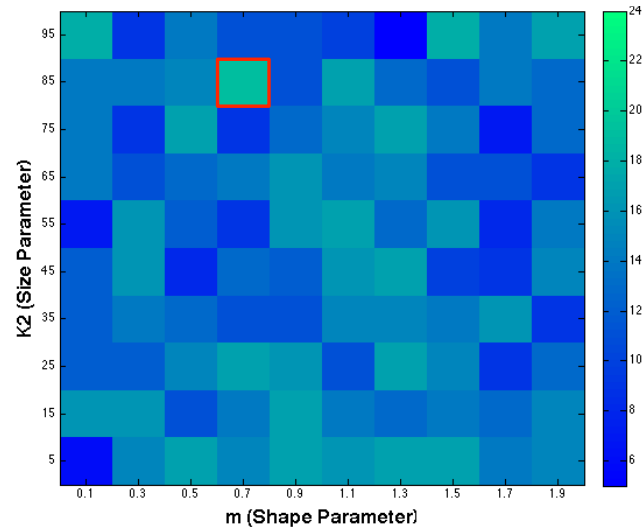


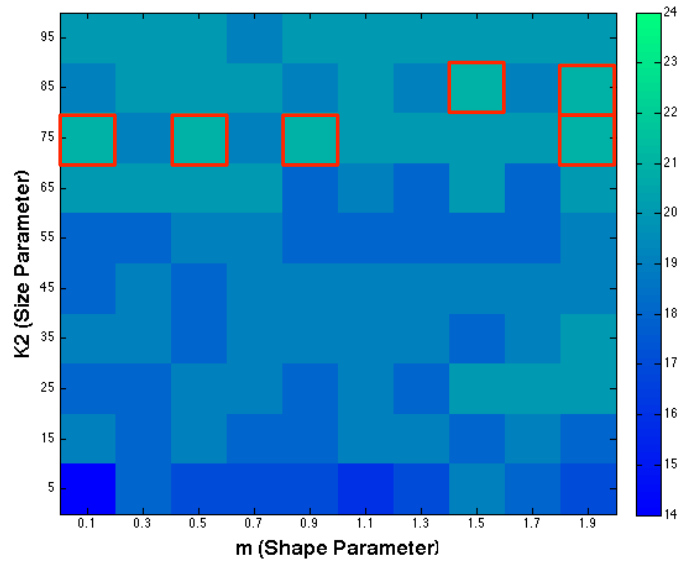
Figure 4-5 Single filter predictions over 10 different shapes and 10 different sizes. The number of correct illusions predicted (out of 28 possible) for different model configurations using a single filter.

$\text{RMS}_{\text{error}}$ is also calculated using the average over these two comparisons. We show predicted results for various model configurations: with no normalization, and with 2 ranges of local normalization ($\sigma = 1$ and $\sigma = 2$). With no normalization, the highest number of correct direction predictions made by a single filter was 20 illusions using a large-sized filter with medium kurtosis. With short-range normalization ($\sigma = 1$), the highest number of correct direction predictions made by a single filter was 21 illusions (present in a small-sized filter with high kurtosis). With an increased normalization range ($\sigma = 2$), the best prediction result was slightly lower at 19 out of 28.

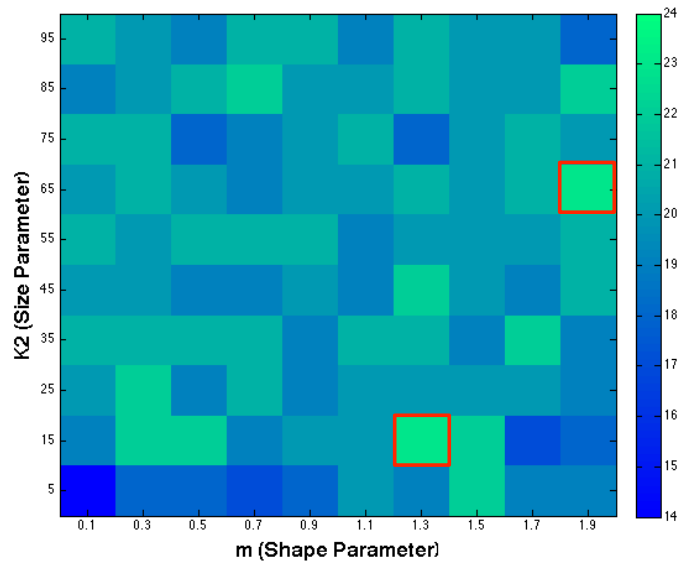
Table 2 lists the results for the best performing size and shape filter in terms of the difference of the mean values over the target patches. As mentioned above, we exclude figure e from our results because no illusion is reported in human results. We report values for z_{2-1} and z_{4-3} (in gray) and take the average of these two as our prediction for z , maintaining a single value prediction per illusion. In table 2 we also reproduce results from Robinson et al., (2007) for the ODOG, best LODOG and best FLODOG model alongside human scaled results for direct comparison. Predictions in the correct direction are shown in bold and tallies of the number of these correct predictions are presented at the bottom. For each model, we also list the $\text{RMS}_{\text{error}}$ that represents how well the model's predictions match the magnitude of human results.

Table 2 shows that performance was maintained (in terms of number of correct direction predictions) when going from raw filter output to short-range normalized results for single filter predictions. Normalized results provided predictions with much smaller magnitudes of lightness illusions, as we would expect. Across predictions of both direction and magnitude, normalized results with $\sigma = 1$ provided the best predictions for single filters, showcasing the highest number of correct direction predictions (21) and reasonable magnitudes for these

Not normalized
Best Result: 21



Normalized, $\sigma = 1$
Best Result: 23



Normalized, $\sigma = 2$
Best Result: 21

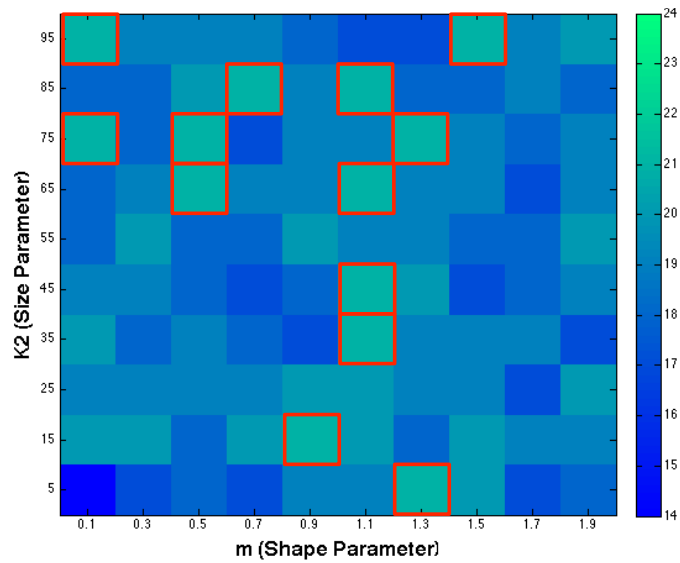


Figure 4-6 Dual filter predictions. Highest predictive success when combining a filter of specified size and shape with any other size and shape filter

predictions (indicated by a substantially reduced $\text{RMS}_{\text{error}}$ compared to filter-only output). Indeed, in this case $\text{RMS}_{\text{error}}$ shows an accuracy of prediction that is matched only by the small values of the ODOG model, which fares considerably less well in terms of number of correct direction predictions (13). The $\text{RMS}_{\text{error}}$ increased when the normalization range was extended to $\sigma = 2$, where only 19 correct direction predictions were made.

Table 2. Model results for the best single filter with and without normalization alongside ODOG and unscaled human results. The result reported for z is the average of z1 and z2 listed in gray.

Figure	Shorthand Name	Human Scaled	ODOG	LODOG $n = 2$	FLODOG $n = 2s$ $m = 0.5$	Exp Model Single filter No norm	Exp Model Single filter Norm $\sigma = 1$	Exp Model Single filter Norm $\sigma = 2$
a	WE-thick	1	1.00	1.00	1.00	1.00	1.00	-1.00
b	WE-thin-wide	1.1	2.08	2.08	2.52	19.36	0.73	1.18
c	WE-dual		-0.30	1.36	1.93	-8.57	0.28	-0.49
d	WE-Anderson	1.54	-0.15	-0.30	-0.43	-1.68	-0.37	0.95
f	WE-zigzag		-0.51	-0.76	1.26	55.52	0.42	-1.69
g	WE-radial-thick-small		-0.67	-0.39	0.46	0.52	0.18	0.16
h	WE-radial-thick		-0.41	0.01	0.18	0.16	-0.16	1.09
i	WE-radial-thin-small		-0.34	0.21	2.74	2.32	0.28	-1.00
j	WE-radial-thin		-0.22	0.83	3.24	0.52	0.58	1.91
k	WE-circular1		-0.82	-1.04	0.28	1.24	0.22	0.52
l	WE-circular0.5		-0.53	-0.67	1.84	-2.84	0.67	2.40
m	WE-circular0.25		-0.38	-0.49	3.64	-2.15	0.55	-1.30
n	Grating induction	1.49	2.03	1.69	0.66	0.20	0.18	-0.30
o	SBC-large	2.72	4.75	7.56	3.96	4.01	3.09	0.75
p	SBC-small	4.73	6.22	14.94	5.96	8.79	4.52	7.02
q	Todorovic-equal	0.53	-0.36	-0.26	0.08	0.19	0.02	1.33
r	Todorovic-in-large	0.57	0.49	0.55	0.39	0.03	0.20	-1.00
s	Todorovic-in-small	1.05	0.80	0.95	1.08	0.32	0.19	0.80
t	Todorovic-out	0.37	0.35	0.38	0.03	0.34	-0.07	1.86
u	Checkerboard-0.16	1.78	1.10	0.94	8.03	-0.34	0.33	2.69
v	Checkerboard-0.94	0.68	0.40	0.35	-4.89	20.26	-0.19	3.93
w	Checkerboard-2.1	1.36	0.69	0.60	-1.48	0.61	0.05	0.77
x	Corrugated Mondrian	2.6	0.95	0.91	0.12	12.92	-0.02	2.32
y	Benary cross	2.2	0.09	0.06	0.05	-559.23	0.23	-1.94
z1	Todorovic Benary 1-2	2.86	-0.12	0.55	0.11	-1408.10	0.23	1.77
z2	Todorovic Benary 3-4	2.28	-0.12	0.58	0.14	1383.70	0.14	7.16
z avg	Todorovic Benary Average	2.57	-0.12	0.57	0.13	-12.20	0.18	4.47
aa	Bullseye-thin		-0.74	-0.35	0.54	0.18	0.02	3.31
bb	Bullseye-thick		-0.77	-0.38	0.07	1.16	-1.49	1.59
<hr/>								
Total Correct			13	18	24	20	21	19
<i>RMS_{Error}</i>			1.29	1.80	2.56	140.59	1.32	1.85

The results presented so far have demonstrated the capability of single filter predictions. We also combined multiple filters to observe the possibility of improving predictive success. Figure 4-6 shows the result of combining pairs of filters together, taking a particular size and shape filter and combining it with the best possible match to maximize the number of correct directions predicted. The best result across all environments (normalized and filter-only), for dual filter combinations was 23 correct directions. The best resultant combinations in terms of maximizing the number of correct prediction directions occurred for a number of filter pairings within different environments. In the filter-only environment, the best filter pairs occurred across a combination of 6 different large sized filters ranging from high to low kurtosis. For normalized filters with $\sigma = 1$, the best filter pair was with a small sized filter with medium kurtosis and a medium sized filter with low kurtosis. For normalized filters with a larger range of normalization ($\sigma = 2$), the best pairings occurred across a range of filters with medium kurtosis over various sizes, or were large in size and had low to medium kurtosis.

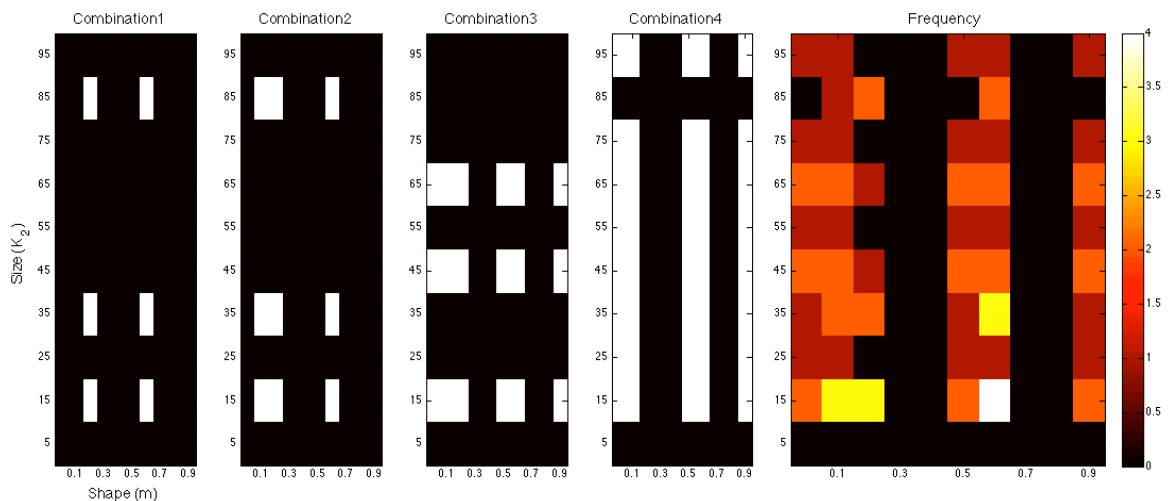


Figure 4-7 The four filter combinations that achieve the maximum of 24 correct illusion direction predictions for the exponential filter model. These combinations were found for short-range, normalized filters. The filters across all four combinations were tallied and the frequency of these is presented on the right.

We extended our multi-filter analysis to allow for the combination of any number of size and shape filters to determine whether an optimal combination of multiple filters exists. Using an ordered search sequence over the space of all possible shape and size filter combinations, we found that the maximum predictive success (in terms of illusion direction) that the model was able to achieve was 24 out of 27. This value represents the upper bound of performance of this exponential filter model and was found for the set of short-range normalized filters. Figure 4-7 illustrates the four filter combinations that achieve the maximum of 24 correct illusion direction predictions for the exponential filter model. This was found for the set of normalized ($\sigma = 1$) filters. The filters across all four combinations were tallied and the frequency of these is presented on the right. A minimum of 6 filters was required to reach the best prediction as shown in combination 1. These were filters of size $K_2 = \{15, 35, 85\}$ and shape $m = \{0.5, 1.3\}$. Combinations 2 -4 in Figure 4-7 show the other filter combinations for which 24 illusions were correctly predicted. We see that a spread of different size and shape filter combinations is required to produce the best predictive performance. Certain filters are found to be informative whereas others are found to be consistently uninformative. Looking at the frequency of specific size and shape filters across all five most successful combinations, we see that filter ($K_2 = 15, m = 1.3$) is common across all filter arrangements. It is also evident that the organization of multiple filters is distributed across the parameter space.

4.4 Discussion

In this study, we applied a series of exponential filters differing in scale and shape to a set of lightness illusions that have previously been tested with Oriented Difference-of-Gaussian (ODOG) filters and associated models. The exponential model far outperforms the early ODOG models, and demonstrates predictive capabilities that match the successes of more recent elaborations of these models – LODOG and FLODOG – that incorporate local

normalization post filtering. Using a single filter, the direction of 21 (out of a possible 27) illusions can be predicted successfully. Using a two-filter combination, the predictive success of the model increases to 23. Extending the model to include any number of combined shape or size filters allows us to define the maximum capability of this model as 24 correct illusion direction predictions. Our results show that a low-level filtering model based on exponential filters can account for a large number of lightness illusions without requiring orientation-selective filters.

Comparing our work to the current literature, we highlight that existing models are restricted to filters of a specific shape (either DOG or LoG). We wanted to explore the effect of variation in the shape of the filters, which remains fixed in existing models. Our aim was not to emphasize stronger prediction performance, but to investigate whether filters inspired by image statistics can provide predictions on par with current state-of-the-art models. We have shown that this is indeed the case, where Gaussian-shaped filters do not provide the best predictability for the illusion set under all circumstances.

While the 28 stimuli used in this study feature substantial differences, one pertinent respect in which they vary is the induction of contrast or assimilation. Six of our illusions can be classified as predominantly contrast effects, whereas 18 primarily produce assimilation, with 4 illusions unclassifiable (see 2.1). Our best single-filter model was able to achieve 5/6 and 13/18 accuracy for contrast and assimilation effects respectively, showing its ability to deal effectively with both classes of effect.

Among our catalog of illusions there are several sets of images that vary principally in terms of SF. These not only include low and high SF versions of White's Effect (a and b) and the SCI (o and p) as highlighted in Figure 4-1. Variations in SF are also seen for radial White's Effect (figures g through to j), circular configurations of White's Illusion (figures k, l and m),

the Checkerboard illusion (u, v and w) and Bullseye figures (aa and bb). In table 2 (column 3), we list values of illusion magnitudes where human data is directly comparable with various SF configurations of the same illusion (reproduced from Robinson et al., 2007). Such comparisons are available for White's illusion (a and b), the SCI (o and p) and the Checkerboard illusion (u, v and w). We draw direct conclusions for the performance of our best single-filter model to these figures. For the remaining figures with no directly comparable human data, we make observations based on the general rule that higher spatial frequencies yield greater effects. Our best single-filter model (normalized with $\sigma = 1$) predicts the correct direction of illusion for both high and low SF versions of White's illusions (stimuli a and b) and of the SCI (figures o and p). In the case of the SCI the model can also account for the change in the size of the illusion as a function of SF, successfully predicting a larger effect at higher SF. However, in conflict with the human data, a reduction of the effect at higher SF is predicted for White's illusion. The Checkerboard illusion is an interesting case where the direction of the effect flips from assimilation to contrast for human observers when the visual angle of checkerboard squares is greater than approximately 1 degree of visual angle. Our best single-filter model is able to successfully account for two out of three illusion directions, with an appropriate increase in magnitude when comparing the lowest (w) and highest (u) SF versions. Despite an incorrect direction being predicted for figure v, the model correctly predicts a reduction in magnitude compared with u. Comparing the performance of our model to the best ODOG variants, we see that only ODOG and LODOG are able to account for all variations of correct illusory magnitudes where human data is available, performing with 5/5 correct relative magnitudes (for comparisons $b > a$, $p > o$, $u > v$, $w > v$ and $u > w$). The best performing model in terms of illusion direction, FLODOG, is able to successfully account for 3 out of a possible 5 illusory magnitudes consistent with SF. We conclude that our model is able to surpass that of FLODOG, with 4/5 illusion magnitudes that are commensurate with human data for both high and low spatial frequencies.

Reflecting on the best performance of the exponential model using a single filter, we note that two particular illusions that were predicted incorrectly – *t* (Pessoa et al., 1998); and *x* (Adelson, 1993) – warrant closer inspection. Stimulus *t* can be said to belong to the family of modified Simultaneous Contrast Illusion figures from *q* to *t*. Figure *s* is a modified version of figure *o* (conventional SCI), where squares with opposite contrast polarity to the background are overlaid onto the target patch, creating equal boundaries of light and dark. Figures *r*, *q* and *t* are modified versions of *s* with increasing crossbar lengths. The spectrum of figure arrangements from *q* to *t* demonstrate changes to figure-ground relationships in terms of object assignment, depth placement and scene segmentation. In figures *q*, *r*, and *s*, the target patch appears to be contiguous with the surrounding white or black regions (as in the SCI: see stimuli *o* and *p*), and is positioned behind black or white square occluders. However, in stimulus *t* – the figure that posed a problem for our most successful single filter model – a quite different depth arrangement is evident, as the target patch now forms a cross that appears to be the most proximal object, and no longer shares the same depth plane as the surround. The exponential model we adopt does not include higher-level information such as depth cues of occlusion. Depth information is also evident in the corrugated Mondrian (figure *x*), providing shadow cues that could be processed by higher cortical levels for lightness judgments. These results may be taken to support suggestions that some illusions may escape successful prediction by low-level mechanisms if their lightness depends on depth relationships (Schirillo et al., 1990).

While the ODOG model and its variants closely approximate the orientation selective operations in V1, exponential filters based on image statistics represent an efficient coding scheme that could be present in pre-cortical areas as early as the retina. The prevailing view in early work with lightness illusions was that they arose from retinal interactions, rather than cortical processing (Cornsweet, 1970; Todorovic, 1997). However, more recent research

highlights the influence of higher-level mechanisms on our lightness perception (Adelson, 2000; Anderson and Winawer, 2005; Gilchrist, 2006). Using our model, we do not prescribe that filtering mechanisms alone can explain all lightness illusions. Instead, we set out to quantify the gap between what filtering operations can and cannot demonstrate. We propose that our exponential filtering model represents the first stage in a process of operations to estimate lightness. Later operations, such as those responsible for the scission of a scene into its component causal layers (Anderson, 1997) would occur post filtering and normalization. The anchoring of lightness values to local and global context (Gilchrist, 2006) could occur within normalization operations or post normalization. In our model's normalization step, the filtered image is first scaled to local responses (using local coefficient of variance) and then to the global maximum response within the image. This provides one of many approximations for the anchoring of lightness values.

The filtering approach we use reshapes contrast distributions towards those that best describe natural images using the exponential filter family. Similarly to Dakin and Bex (2003), we essentially reconstruct an image that represents the most probable naturally occurring source. By redistributing lightness values to more closely reflect the underlying statistical relationships of images within our environment, we can form predictions of perceptual lightness estimates that align with a large array of lightness illusions. **Figure 4-8** illustrates the power spectra for a set of images that are unfiltered (left column) and filtered (right column) using different shape filters that are all of size 5 pixels. The top row illustrates power spectra for 28 natural images. From these graphs we can see that the power spectra for filtered natural images is quite similar to the power spectra for unfiltered natural images. The bottom row shows the power spectra for illusory images. The unfiltered images in the bottom left graph show a flatter power spectrum in the lower SFs than the filtered images in the bottom right graph. By applying these exponential filters, we see that they not only push the power

spectra of illusory images toward that of natural images, reflecting the properties of image statistics. Applying these filters also boosts low SF information, hypothesized to be a driving factor in the perception of lightness illusions (Dakin and Bex, 2003). Dakin and Bex (2003) find that low spatial frequencies are primarily responsible for the Craik, Cornsweet, and O'Brien (CCOB) illusion that they study. The LoG filters that they apply boost this information when it is not present. From their results, Dakin and Bex (2003) hypothesize that low SF information may drive many illusions.

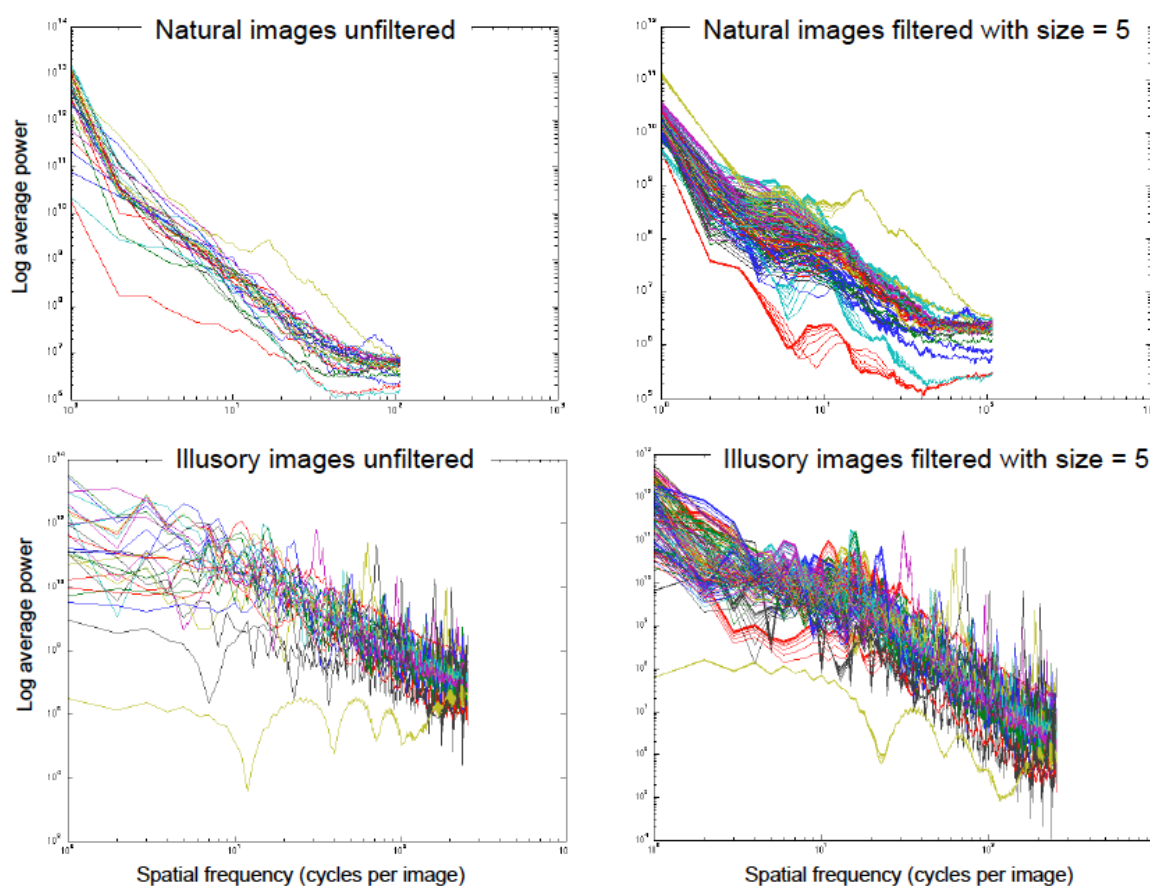


Figure 4-8 Power spectra for images that are unfiltered (left column) and filtered with size=5 pixels (right column). Top row: 28 natural images. Bottom row: 28 illusory images.

In a post-hoc analysis, we analyse whether filters of a particular shape aid in boosting low SF information, which is postulated by Dakin and Bex (2003) as a driving factor for many illusions. Figure 4-9 illustrates the effect of different shape filters on the power distribution of a filtered White's Illusion image. Looking at the left side of the graph, we see that different shape filters have an effect on the low spatial frequency distributions. Filters with high kurtosis (those that have a low exponent and a sharper distribution) boost low SFs more than filters with low kurtosis (those that have a high exponent and a flatter distribution). The exponential filters therefore provide a mechanism to boost lower spatial frequency information more than Gaussian filters.

We emphasize that this study was conducted to investigate filters that are best able to push the power spectra of images toward that of natural images as well as preserve image structure while being resilient to noise. In earlier work, we showed that a filter size selection model helps in extracting and amplifying local image structure (Ghebreab et al., 2009). This model locally selects the smallest filter (extracting high-frequency information) with a response above a noise threshold (ensuring resilience to noise). In a similar fashion, local selection of filter shape may further enhance local image structure. Instead of performing local scale and shape selection in this paper, we study how different types of filters, varying in size and shape, may explain illusions.

The two-stage process of our model uses exponential filters that allow for efficient coding, followed by divisive normalization to boost shallow edges, promoting faithful representation of salient image features. In this way, the filtering stage of our model relies on the Efficient Coding Hypothesis, a theoretical model of sensory coding in the brain (Barlow, 1961). The Efficient Coding Hypothesis states that sensory information is represented in the most efficient way possible, such that it is closely representative of an organism's natural environment. The Efficient Coding Hypothesis is closely related to the Predictive Coding

approach (Srinivasan *et al.*, 1982), which states that the representation of sensory information in a statistically efficient way allows sensory systems to reduce redundancies and also provides greater resilience to noise (Barlow, 1961, 2001). In the specific case of our model, there is ample evidence from Basu and Su (2001) that exponential filters are resilient to many types and intensities of noise. From Dakin and Bex (2003) we see that statistical image representation and noise handling complement one another in understanding and predicting lightness illusions. Alongside Dakin and Bex (2003), by successfully modeling illusions using properties of image statistics, we support the predictive coding approach proposed by Srinivasan *et al.* (1982).

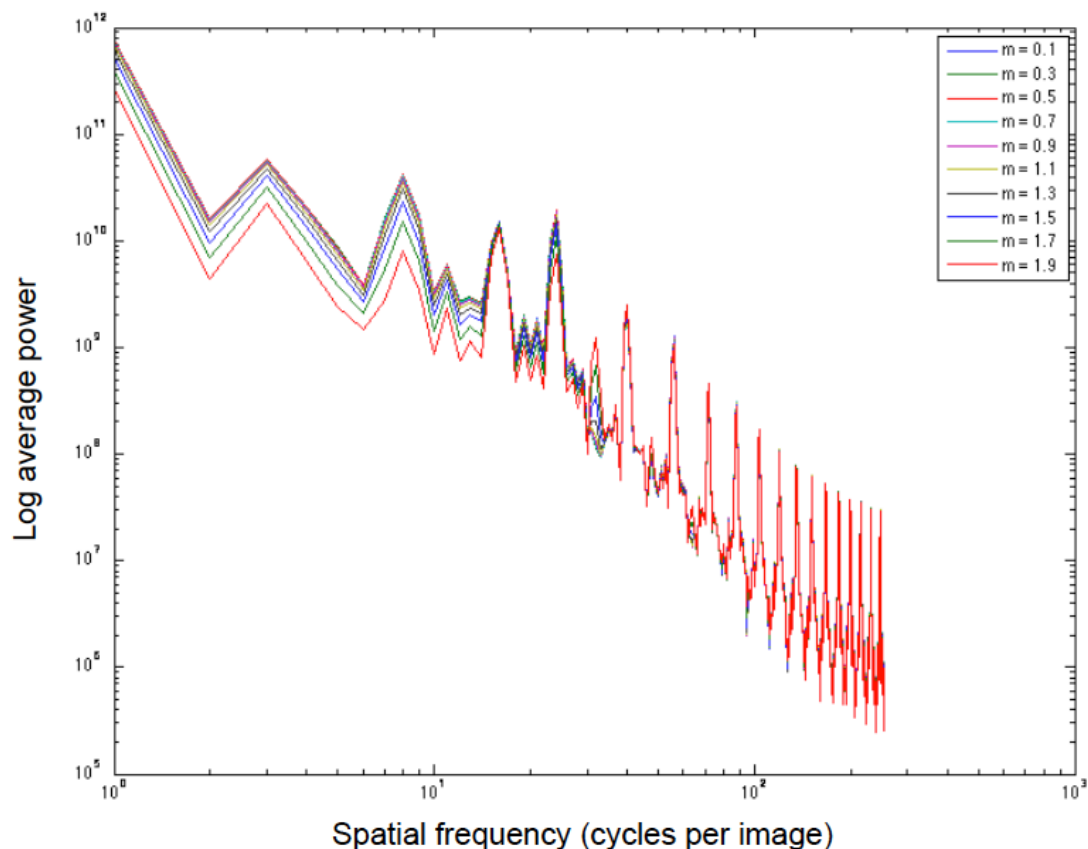


Figure 4-9 Average power over spatial frequency of different shape filters applied to White's Illusion (Figure a). All filters are of size 5 pixels. m refers to the exponent.

In earlier work we showed that globally processing images with filters of different sizes results in scale space image representations that account for different visual phenomena (Ghebreab et al., 2009). We also showed that collapsing scale space representations into a single image representation via local scale selection accounts for even further visual phenomena. This model locally selects the smallest filter (extracting high-frequency information) with a response above a noise threshold (ensuring resilience to noise). In a similar fashion, local selection of filter shape may further enhance local image structure. Instead of performing local scale and shape selection, in this work we first studied if and how different types of filters, varying in size and shape, may explain illusions. We found this is indeed the case. We also tested whether combining different image representations, obtained by globally applying different filters, adds to explaining illusions. The next step in our work would be to determine whether local selection of filter size and shape, based on a model similar to Ghebreab et al. (2009), is able to further explain illusions.

An interesting future direction of study would be to explore additional versions of White's effect, particularly those that have been found to produce an inverted effect (Spehar et al., 1995; Ripamonti and Gerbino, 2001; Spehar et al., 2002). It is well-known that White's effect holds only when the luminance of the two target patches lies between the luminance values of the surrounding gratings (Spehar et al., 1995). Modifying the luminance values of the test patches to double-increments or double-decrements, relative to the gratings, not only drastically reduces the magnitude of illusion, but can also reverse the direction of the illusion from assimilation to contrast (Spehar et al., 1995; Ripamonti and Gerbino, 2001; Spehar et al., 2002). Inverted versions of White's effect have not been successfully accounted for using Blakeslee and McCourt (1999)'s ODOG model, according to Spehar et al. (2002). Testing double-increment and double-decrement versions of White's effect in the exponential filter model may further demonstrate its robustness in accounting for an even larger range of

lightness illusions.

Another direction for follow-up work would be to investigate the effects of different types and intensities of noise on human perception of lightness illusions and observe how closely these results are matched by our exponential filter model. Dakin and Bex (2003) show that when introducing different levels of noise into their stimuli, their model maintains a close approximation to human performance. However, ODOG has shown discrepancies in matching human response magnitudes for noisy stimuli (Betz *et al.*, 2014). If the exponential filter model demonstrates results similar to human observers in classifying illusory images with noise manipulations, this would provide further support for predictive coding (Srinivasan *et al.*, 1982).

In summary, our study demonstrates that a filter model based on contrast distribution statistics of natural images is able to account for the direction of 21 out of 27 lightness illusions using a single filter. When two filter combinations are considered, the number rises to 23, with asymptotic performance at 24 for an arbitrarily large combination of filter outputs. We observe the effect of incorporating non-linear divisive normalization, providing a better understanding of the role that contrast gain control provides in the perception of these illusions. While short-range normalization only slightly improves the number of correct direction predictions, it considerably reduces the error in predicting illusion magnitude, measured as $\text{RMS}_{\text{error}}$. The exponential filters we employ are not orientation selective, demonstrating that V1-style operations are not required to account for a large number of lightness illusions. Given that these exponential filters could be found as early as the retina, it is possible that the majority of these lightness effects result from pre-cortical operations, leaving only a few to be explained by higher level mechanisms.

Acknowledgement

Funding: This work was partially conducted under the Research Priority Program “Brain & Cognition” at the University of Amsterdam and supported by the Dutch national public-private research program COMMIT to S.G. AZ is supported by the Australian Research Council Centre of Excellence for Cognition and its Disorders (CE110001021) <http://www.ccd.edu.au>.

4.5 References

- Adelson, E. H. (1993). Perceptual organization and the judgment of brightness, *Science*, 262(5142), 2042–2044
- Adelson, E. H. (2000). Lightness Perception and Lightness Illusions, *In: The New Cognitive Neurosciences*, 2nd edition, MIT Press, Cambridge, MA, 339–351.
- Anderson, B. L. (1997). A theory of illusory lightness and transparency in monocular and binocular images: The role of contour junctions, *Perception*, 26, 419–454.
- Anderson, B. L. (2001). Contrasting theories of White’s illusion. *Perception*, 30(12), 1499–1501.
- Anstis, S. (2003). White’s effect radial, <http://www.cogsci.ucsd.edu/stanonik/illusions/wer0.html> Online Demonstration.
- Barlow, H. (1961). Possible principles underlying the transformation of sensory messages, *In: Sensory Communication*, MIT Press, 217–234.
- Barlow, H. (2001). Redundancy reduction revisited, *Network: Computation in Neural Systems*, 12, 241–253.
- Basu, M. and Su, M. (2001). Image smoothing with exponential functions, *International Journal of Pattern Recognition and Artificial Intelligence*, 14(4), 735–752.
- Benary, W. (1924). Beobachtungen zu einem Experiment über Helligkeitskontrast, *Psychologische Forschung*, 5, 131–142.
- Betz, T., Wichmann, F. A., Shapley, R. and Maertens M. (2014). Testing the ODOG brightness model with narrowband noise stimuli, *Perception*, 43, ECVF Abstract Supplement, 164.
- Bindman, D. and Chubb, C. (2004). Brightness assimilation in bullseye displays, *Vision Research*, 44, 309–319.
- Blakeslee, B. and McCourt, M. E. (1997). Similar mechanisms underlie simultaneous brightness contrast and grating induction, *Vision Research*, 37 (20), 2849–2869.
- Blakeslee, B. and McCourt, M. E. (1999). A multiscale spatial filtering account of the White effect, simultaneous brightness contrast and grating induction, *Vision Research*, 39, 4361–4377.

- Blakeslee, B. and McCourt, M. E. (2001). A multiscale spatial filtering account of the Wertheimer-Benary effect and the corrugated Mondrian, *Vision Research*, 41(19), 2487–2502
- Blakeslee, B. and McCourt, M. E. (2004). A unified theory of brightness contrast and assimilation incorporating oriented multi-scale spatial filtering and contrast normalization, *Vision Research*, 44(21), 2483–2503.
- Blakeslee, B. and McCourt, M. E. (2008). Nearly instantaneous brightness induction, *Journal of Vision*, 8(2), 1–8.
- Blakeslee, B., Pasioka, W., and McCourt, M. E. (2005). Oriented multiscale spatial filtering and contrast normalization: a parsimonious model of brightness induction in a continuum of stimuli including White, Howe and simultaneous brightness contrast, *Vision Research*, 45(5), 607–615
- Blakeslee, B., Reetz, D., and McCourt, M. E. (2008). Coming to terms with lightness and brightness: Effects of stimulus configuration and instructions on brightness and lightness judgments, *Journal of Vision*, 8(11), 1–14.
- Bonin, V., Mante, V., and Carandini, M. (2005). The suppressive field of neurons in lateral geniculate nucleus, *The Journal of Neuroscience*, 25(47), 10844–10856.
- Carandini, M. and Heeger, D. J. (2012). Normalization as a canonical neural computation, *Nature Reviews Neuroscience*, 13, 51–62.
- Chevreul, M. E. (1839). *De la loi du contraste simultané des couleurs et de l'assortiment des objets colorés*. Translated into English by C. Martel as *The principles of harmony and contrast of colours*. (English Second Edition: Longman, Brown, Green and Longmans 1855).
- Clifford, C. W. G. and Spehar, B. (2003). Using colour to disambiguate contrast and assimilation in White's effect, *Journal of Vision*, 3(9), 294–294.
- Cope, D., Blakeslee, B., and McCourt, M. E. (2013). Modeling lateral geniculate nucleus response with contrast gain control, *Journal of the Optical Society of America A*, 30(11), 2401–2408.
- Corney, D. and Lotto, R. B. (2007). What are lightness illusions and why do we see them?, *PLoS Comput Biol*, 3(9), e180, doi:10.1371/journal.pcbi.0030180
- Cornsweet, T. N. (1970). *Visual Perception*, Academic Press.
- Craik, K. J. W. (1966). *The nature of psychology: a selection of papers, essays and other writings by the late K. J. W. Craik*. Cambridge University Press.
- Dakin, S. C. and Bex, P. J. (2003). Natural image statistics mediate brightness 'filling in', *Proceedings of the Royal Society of London, Biological Sciences*, 270(1531), 2341–2348.
- Daugman, J. G. (1989). Entropy reduction and decorrelation in visual coding by oriented neural receptive fields, *IEEE Transactions on Biomedical Engineering*, 36(1), 107–114.

- De Valois, R. L. and De Valois, K. K. (1988). *Spatial Vision*. Oxford University Press.
- Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells, *Journal of the Optical Society of America A, Optics and Image Science*, 4(12), 2379–2394.
- Geisler, W. S. (2008) Visual perception and the statistical properties of natural scenes. *Annual Review of Psychology*, 59, 167-192.
- Ghebreab, S., Scholte, H. S., Lamme, V. A. F., and Smeulders, A. W. M. (2009), A biologically plausible model for rapid natural scene identification, in *Proceedings of Neural Information Processing Systems 2009*, 629 – 637.
- Ghosh, K., Sarkar, S., and Bhaumik, K. (2006), A possible explanation of the low-level brightness contrast illusions in the light of an extended classical receptive field model of retinal ganglion cells, *Biological Cybernetics*, 94(2), 89–96.
- Gilchrist, A. (2006). *Seeing black and white*. Oxford University Press.
- Gilchrist, A., Kossyfidis, C., Bonato, F., Agostini, T., Cataliotti, J., Li, X., et al. (1999). An anchoring theory of lightness perception, *Psychological Review*, 106(4), 795–834
- Gilchrist, A. L. (1977), Perceived lightness depends on perceived spatial arrangement, *Science*, 195, 185–187.
- Howe, P. D. L. (2001). A comment on the Anderson (1997), the Todorovic (1997) and the Ross and Pessoa (2000) explanations of white's effect, *Perception*, 30, 1023–1026.
- Howe, P. D. L. (2005), White's effect: Removing the junctions but preserving the strength of the illusion, *Perception*, 34(5), 557 – 564.
- Kingdom, F. A. (2011). Lightness, brightness and transparency: A quarter century of new ideas, captivating demonstrations and unrelenting controversy, *Vision Research*, 51, 652–673.
- Knill, D. C. and Kersten D. (1991). Apparent surface curvature affects lightness perception, *Nature*, 351, 228 – 230
- McCourt, M. E. (1982). A spatial frequency dependent grating-induction effect, *Vision Research*, 22(1), 119–123, 125–134.
- Nykamp, D. Q. and Ringach, D. L. (2002). Full identification of a linear-nonlinear system via cross-correlation analysis, *Journal of Vision*, 2, 1–11.
- O'Brien, V. (1958), Contour perception, illusion and reality, *Journal of the Optical Society of America*, 48(2), 112–119
- Packer, O. S. and Dacey, D. M. (2002). Receptive field structure of H1 horizontal cells in macaque monkey retina, *Journal of Vision*, 2, 272–292

- Packer, O. S. and Dacey, D. M. (2005). Synergistic center-surround receptive field model of monkey H1 horizontal cells, *Journal of Vision*, 5, 1038–1054.
- Pessoa, L., Barattoff, G., Neumann, H., and Todorovic, D. (1998). Lightness and junctions: variations on White's display, *Investigative Ophthalmology and Visual Science (Supplement)*, 39, S159.
- Ripamonti, C. and Gerbino, W. (2001). Classical and inverted white's effect, *Perception*, 30(4), 467–488.
- Robinson, A. E., Hammon, P. S., and de Sa, V. R. (2007). Explaining brightness illusions using spatial filtering and local response normalization, *Vision Research*, 47, 1631–1644.
- Ruderman, D. L. and Bialek, W. (1994). Statistics of natural images: Scaling in the woods, *Physical Review Letters*, 73(6), 814–817.
- Schirillo, J., Reeves, A., and Arend, L. (1990). Perceived lightness, but not brightness, of achromatic surfaces depends on perceived depth information, *Perception and Psychophysics*, 48(1), 82–90.
- Schwartz, O. and Simoncelli, E. P. (2001). Natural signal statistics and sensory gain control, *Nature Neuroscience*, 4(8), 819–825.
- Shapiro, A. and Lu, Z.-L. (2011). Relative brightness in natural images can be accounted for by removing blurry content, *Psychological Science*, 22(11), 1452–1459.
- Simoncelli, E. P. (2003). Vision and the statistics of the visual environment, *Current Opinion in Neurobiology*, 13, 114–149.
- Spehar, B., Clifford, C. W. G., and Agostini, T. (2002). Induction in variants of white's effect: common or separate mechanisms? *Perception* 31(2), 189–196. doi: 10.1068/p10sp
- Spehar, B., Gilchrist, A., and Arend, L. (1995). White's illusion and brightness induction: the critical role of luminance relations. *Vis. Res.* 35(18), 2603–2614.
- Srinivasan, M. V., Laughlin, S. B., and Dubs, A. (1982). Predictive coding: a fresh view of inhibition in the retina, *Proceedings of the Royal Society of London B Biological Sciences*, 216(1205), 427–459.
- Todorovic, D. (1997). Lightness and junctions, *Perception*, 26(4), 379–394.
- White, M. (1979). A new effect of pattern on perceived lightness, *Perception*, 8(4), 413–416.
- Zhu, S. C. and Mumford, D. (1997a). Prior learning and Gibbs reaction-diffusion, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(11), 1236 – 1250.
- Zhu, S. C. and Mumford, D. B. (1997b). Learning generic prior models for visual computation, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, IEEE Computer Society, Los Alamitos, CA, 463–469.

5 Discussion & Conclusion

5.1 Chapter overview

This chapter provides a general discussion aimed at further elucidating some of the elements of the studies included within this thesis. This chapter is divided into five main sections. Section 5.2 summarises each of the studies and integrates these together. Section 5.3 discusses how this work fits in with the wider academic literature and how this work impacts on knowledge within the field. Potential improvements in the methods that we used and possible future directions for follow-up studies are highlighted in Section 5.4. Section 5.5 suggests some of the applications of this work for the fields of psychology and computer science fields. Finally, in Section 5.6, we provide some closing remarks.

5.2 Summary and integration of studies

5.2.1. Each of the studies in summary

In our first study (Chapter 2, Zeman *et al.*, 2013), we conducted a series of three experiments to show that the MLI manifested in a benchmark model of the visual ventral stream, HMAX, which is commonly associated with object recognition (Mutch & Lowe, 2008). HMAX adopts filters that emulate the simple and complex cell firing behaviour that was recorded in Hubel and Wiesel's landmark studies (Hubel, 1959; Hubel & Wiesel, 1959, 1962, 1965). We found that HMAX was not only susceptible to the MLI, but it also generated bias patterns that were consistent with human results, demonstrating greater bias for fins with more acute angles. We were able to rule out some of the necessary factors required to generate illusory bias by demonstrating that the effect in HMAX was present without exposure to 3D images and without the presence of feedback (Gregory, 1963). Using HMAX, we also found that there was no particular reliance on low-spatial frequency filters in generating the effect, ruling out mandatory reliance on low-spatial frequency information (Carrasco *et al.*, 1986).

Our second study (Chapter 3, Zeman *et al.*, 2014) built on the findings of the first, presenting two experiments looking at levels of accuracy and precision in the model's responses to MLI figures at each stage of the HMAX hierarchy. Surprisingly, we found that HMAX reduced the illusion compared to the input level. This suggests that image statistics are the main driver for the illusion in the model, in line with proposals from Howe and Purves (2002, 2005a, 2005b) and Corney and Lotto (2007). We found that at every stage of the model, bias and uncertainty were reduced when compared with input-level classification (without any processing by HMAX simple or complex cell layers). Within the model, we found that in the majority of cases (87.5%), complex cells reduced illusory bias or uncertainty. Following on from this finding, we hypothesised that increasing the positional variance at the input level would engage more complex cell functionality and would therefore reduce errors. We confirmed this result by introducing horizontal positional variance (in addition to vertical jitter) and demonstrating a reduction in bias. By comparing input-level classification to model classification at various layers, and by manipulating properties of the training image set, we concluded that image statistics were the main driving factor behind the illusion in this model. Our research focus then shifted to more statistics-driven approaches and to image representation at lower levels of the visual hierarchy.

For our final study (Chapter 4, Zeman *et al.*, *in submission*), we conducted a series of comprehensive simulations focused on modelling a set of 28 lightness illusions, using an approach inspired by image statistics. We looked at a set of illusions that would be likely to occur within early visual areas (Blakeslee & McCourt 1999, 2001, 2004; Robinson *et al.*, 2007). The most extensively researched model of illusions in early visual areas is ODOG (oriented difference of Gaussians) (Blakeslee & McCourt 1999, 2001, 2004, 2008; Blakeslee *et al.*, 2005, 2008), which supports a low-level account of lightness illusions. However, some

researchers argue for the involvement of higher-level cortical processes in influencing our lightness perception (Gilchrist, 1977; Knill and Kersten, 1991; Anderson, 1997; Gilchrist et al., 1999; Schmid & Anderson, 2014). Our aim was to demonstrate the extent to which lightness illusions can be accounted for by low-level filtering mechanisms. We used filters derived from contrast distributions found in natural images that have been shown to preserve image structure while removing noise. We found that by employing exponential filters of multiple sizes and shapes (Basu and Su, 2001), we were able to account for a large array of existing lightness illusions. The benefits of our approach over existing models is that the exponential filter model has greater resilience to noise; provides a normalisation scheme that has greater biological plausibility (Bonin et al., 2005); has greater flexibility in allowing for shape as well as size selection of filters, and has greater parsimony over existing models.

5.2.2. Common threads between studies

The three studies contained within this thesis cover the computational modelling of pre-cortical and cortical areas of vision, demonstrating illusory effects at multiple stages of visual processing. To tie these studies together, we highlight common threads between the models that we used and their relationship to studying illusions:

- both models that we used were feed-forward, allowing us to establish the influence of bottom-up only connections in bringing about illusory effects
- both models allowed us to assess the suitability of lower or higher level explanations for a range of illusions
- both models quantified the influence of different cell types or filters on the overall effect
- both models were influenced by image statistics in separate ways

- both models combined filtering techniques with image statistics

We now elaborate on each of these points in turn.

Both models that we implemented, HMAX and our exponential filter model, are feed-forward. One – the exponential filter model of perceived lightness – is fully deterministic (Zeman *et al.*, *in submission*), while the other – HMAX – has a degree of stochasticity influenced by feature learning from a training set of images (Serre et al., 2005a, 2005b, 2007; Mutch & Lowe, 2008; Serre & Poggio, 2010). For each of these models, information only flows one-way, allowing us to investigate the degree to which an illusory effect can be brought about by the feed-forward sweep within an artificial network. In all three studies, we were able to measure levels of bias for the illusions that we studied and additionally assess other quantifiable parameters. We measured effect magnitudes while manipulating the fin angle or line positioning for the MLI (Zeman *et al.*, 2013, 2014). For the studies using HMAX, we were also able to assess precision, or the level of certainty, for different manipulations of the MLI. In our third study, using our exponential filter model, we were able to quantify the number of correct predictions made for a range of illusions as well as how closely these predictions matched human performance (Zeman *et al.*, *in submission*). Therefore, in all of our studies, we were able to put forward a set of metrics that establish the success of these models in predicting a range of illusions and allow for direct comparison with human results. This set of quantifiable measurements provides a baseline for other competing models to be assessed and compared against, establishing a framework for further comparative studies to be made in the future.

The models that we implemented can be broadly mapped to different levels of the neural hierarchy. The analogies between biology and the models that we used allow us to separate out bottom-up versus top-down explanations of particular illusions. Taking an example from

our third study (Chapter 4, Zeman *et al.*, *in submission*), we tested our exponential filter model using a range of images, including the Benary Cross illusion (Benary, 1924), where two identical grey triangles that share the same bordering information appear to have different luminance that depends on contextual placement inside or outside an object (illustrated in Figure 5-1). Theories surrounding the Benary Cross range from lower-level to higher-level explanations (Salmela & Laurinen, 2009). The dependence on figure-ground assignment could involve the recruitment of higher areas such as IT (Baylis and Driver, 2001), although other studies suggest that the neural processes associated with border ownership could occur within earlier areas such as V1 and V2 (Zhou *et al.*, 2000). Given that we were able to show the illusion manifest within our exponential filter model that has no hierarchical architecture, we demonstrate that higher levels are not necessary for bringing about the Benary Cross illusion, in agreement with other studies such as Salmela & Laurinen (2009).

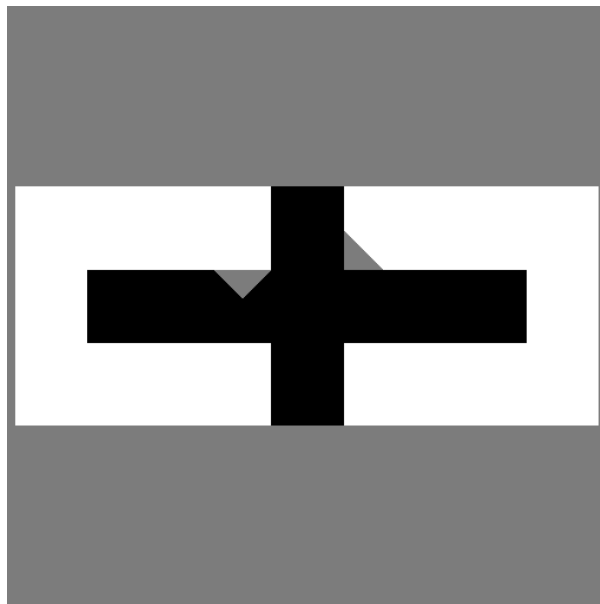


Figure 5-1 Benary Cross illusion, reproduced from Zeman *et al.*, *in submission*.

Although both of the models covered in this thesis, the exponential filter family model and HMAX, combine filtering with image statistics, the influence of image statistics on each is slightly different. In the exponential filter model, natural image statistic profiles are embedded

in the filter functions that are applied to the image (Basu and Su, 2001). The contrast distributions for natural images are shown to have a common underlying relationship (Zhu and Mumford, 1997). Zhu and Mumford propose that exponential filters provide a better way to extract high spatial frequency information such as edges and preserve the distribution of spatial frequencies after they are applied to images. By applying exponential filters, the distribution of contrasts over different spatial frequencies is more closely matched to that of natural images. HMAX, on the other hand, learns hidden relationships between input images and the labels assigned to them during the training phase, much like the model of Corney & Lotto (2007). In other words, the exponential filter model explicitly encodes relationships with image statistics, whereas HMAX implicitly learns this.

The models chosen for this thesis were purposefully selected for their simplicity. By observing effects reproduced in a scaled-back model, we aimed to provide better explanations for illusory effects by observing what components are first necessary in order to bring about bias. While it is tempting to opt for more sophisticated models that incorporate, for instance, feedback or temporal dynamics, these models would be excessive and unnecessary unless there are specific questions that require these features. For example, if we were interested in how bias for the Müller-Lyer illusion changes with inspection time (Coren and Porac 1984; Predebon et al. 1993; Predebon, 1998, 2006), then it would be useful to incorporate feedback in a model. Instead, we selected feed-forward models in the interests of quantifying bottom-up effects and eliminating top-down interactions from higher-levels.

Competing theories surround the illusions in humans that have been presented in this thesis. Many of these theories can be separated into low-level and higher-level accounts, corresponding to the brain areas that are believed to contribute towards biases. Computational modelling provides a useful tool to delineate the influence of these areas by providing

parallels with designated brain regions. Levels of HMAX map to areas of the visual ventral stream (Riesenhuber & Poggio, 1999; Serre *et al.*, 2007; Serre & Poggio, 2010), giving an opportunity to observe simulation results at each hierarchical level and, as a result, observe higher or lower-level influences on bias. The exponential filter model, while inspired in part by successful edge detection algorithms in computer vision, best maps to the receptive fields of retinal cells and LGN normalisation operations. This is in contrast to the highly successful ODOG model (Blakeslee & McCourt, 1999), which is most closely matched to filtering operations in V1. By matching theories and models to their appropriate neural correlates, it is possible to narrow down the level of contribution or reduction that each brain area may have towards a particular illusory effect (Chapter 3, Zeman *et al.*, 2014).

5.3 Our studies in the context of wider academic literature

5.3.1. Can illusions manifest in artificial brains?

Historically, artificial models existed that did not contain multiple layers but were still able to demonstrate illusory bias. These models were able to produce output similar to human behaviour when presented with illusory figures, either by emulating the filtering operations of cells (Bertulis & Bulatov, 2001, 2005) or by analysing statistics in the environment (Howe and Purves 2002, 2005a, 2005b; Corney and Lotto, 2007). However, these models were deterministic, non-hierarchical systems that did not involve any feature learning. It was not until Brown and Friston (2012) that hierarchical systems were first considered as candidates for modelling illusions, even though the authors omitted important details of the model's architecture, such as the number of layers they recruited.

In order to find a hierarchical model capable of learning, we turned to current state-of-the-art object recognition systems that were inspired by neurobiology. Such artificial systems have

been highly developed and are capable of accurate object classification under a variety of harsh conditions including changes to lighting and angle (Serre *et al.*, 2005a, 2005b, 2007; Mutch & Lowe, 2008; Serre & Poggio, 2010). HMAX provided an architecture that was hierarchical in structure and capable of learning (Riesenhuber & Poggio, 1999; Serre, 2014).

Initially, we considered that a network consisting only of feed-forward connections and lacking any feedback would be a weak contender for showing any manifestation of the Müller-Lyer illusion (Müller-Lyer, 1889). This was given that the prevalent explanation of the illusion centred on common feature associations (between arrowhead and arrow-tail placement and distance to objects) that were fed back into the network from higher to lower levels (Gregory, 1963, 1966, 1997). The plan for our first study was to show that if no illusion occurred in such a feed-forward, hierarchical model, then additional elements would be required to demonstrate a bias.

Surprisingly, the HMAX model did show a repeatable bias for the Müller-Lyer illusion with several different fin angle configurations (Chapter 2, Zeman *et al.*, 2013). The magnitude of bias was greater for more acute angles compared to obtuse angles, consistent with human data (Restle & Decker, 1977). So to summarise, illusions can manifest in artificial systems that are both hierarchical and capable of learning. Whether these networks rely on exposure to the same images that we see during training, or on filtering mechanisms that are based on similar neural operations, they produce a consistent and repeatable illusory bias. In terms of Marr's (1982) levels of description (see section 1.6), it appears that illusions can manifest at the hardware level (Howe & Purves, 2005a, 2005b) and at the algorithmic/representational level (Bertulis & Bulatov, 2001, 2005; Zeman *et al.*, 2013).

5.3.2. Alternatives to HMAX and associated illusion predictions

Our first two studies used HMAX as our model of choice for studying the MLI. There are, of course, other models of the visual ventral stream, with some containing only feed-forward connections and others that employ feedback connections within their architecture. We now take a brief look at some other models of the visual ventral stream and compare these with HMAX. Based on the similarities and differences between these models and HMAX, we then provide a set of predictions for how these models would perform when presented with a set of Müller-Lyer images.

Considering other well-known feed-forward models of the visual ventral stream, we turn our attention to SpikeNet: an eight-layer network that incorporates the coding of information across time as a sequence of spikes (VanRullen *et al.*, 1998; Thorpe *et al.*, 2001; VanRullen and Thorpe, 2001a, 2001b, 2002; VanRullen *et al.*, 2005). SpikeNET uses Rank Order Coding (ROC) as a temporal coding scheme, which is where a cell fires with a short delay after stimulus onset if the cell prefers that stimulus. Conversely, cells with a lower preference for the input will fire with a greater delay (Thorpe *et al.*, 2001). Using this coding scheme, the precise timing of spikes is less important than the order in which neurons fire (Gautrais and Thorpe, 1998). SpikeNET uses on-centre and off-centre cells as filters within its hierarchy (VanRullen and Thorpe, 2001). The spatial scales of on- and off- centre cells increase as the layers are traversed. The first layer of the system is an approximate model of retinal ganglion cells and the second layer is the rough equivalent of V1 (Thorpe *et al.*, 2001). Supervised learning is used to train neurons to fire in the appropriate order.

Comparing HMAX to SpikeNET, one obvious difference is that HMAX omits on- and off-centre cell functions. This raises the question whether processing by on- and off-centre filters

would also demonstrate system susceptibility to bias when viewing Müller-Lyer images. From our second study, we found that bias was evident at the input level and that filter processing by subsequent levels of HMAX reduced bias when compared to the input. We would hypothesise that a similar mechanism would occur within SpikeNET, which is that processing by filter layers would serve to reduce bias but not eliminate it altogether. Figure manipulations such as fin angle may have greater activation levels with orientation-selective cells (such as those present in HMAX), however on- and off-centre cells would still be activated, albeit to a lesser degree, within different regions of the image.

Other differences also exist between HMAX and SpikeNET. Masquelier and Thorpe (2007) provide an informative comparison between a modified version of SpikeNET and HMAX, the main difference between these two approaches being the size of their filter banks. HMAX has a large set of available filters that are activated by particular image features. Exposure to an image causes activations to cascade through the network, later activating nodes in the final layer of the model. The last layer of HMAX is used for classification that is explicitly relevant to the learning task. SpikeNET, on the other hand, consists of a smaller feature dictionary that automatically selects features that hold the greatest importance. The difference in network sizes between these two models should not affect the classification of input images. Within each model, the size of the network, which reflects the size of available filter banks, would affect the percent correct score. As shown in our first study, adjusting the size of one of the layers of HMAX affects the proportion of correct categorisations for our control condition. Therefore, once the network size for a particular model is adjusted so that it is able to reach high performance levels for the classification of control images, we would still expect a bias to be present for MLI images.

We now move away from focusing solely on feed-forward models and instead look at comparisons between feedback models and HMAX. Being a descriptive (i.e. feed-forward) model, HMAX differs from generative models of the visual ventral stream, such as those presented by Hinton and Zemel (1994), Lee and Mumford (2003), Kersten *et al.*, (2004), Friston (2005a, 2008, 2010, 2012) and Bastos *et al.* (2012), which generate predictions based on long and short histories of previous input. In contrast, HMAX contains only a conceptually long history of exposure to inputs, which is accumulated during training and remains static during run-time of the model. In regards to whether these alternative models of the visual ventral stream would predict the MLI, we hypothesise that all of these models would be susceptible to the illusion. From our second study, we concluded that the MLI, at least as manifested in HMAX, is a result of the statistics of the input training images. Given that all of the above examples are artificial neural networks that would be trained on these images, they would be exposed to mappings between object sizes and visual templates of arrowheads and arrow-tails (Howe and Purves, 2005b). In feedback models, it would be interesting to observe the level of bias over time, to see if, as with humans, the magnitude decreases with inspection time (Predebon, 1998). Given that we've shown how complex cells reduce errors associated with the Müller-Lyer (Chapter 3, Zeman *et al.*, 2014), it would be interesting to see whether bias and uncertainty also change between levels of a hierarchical feedback model. If errors decrease within each ascending layer of the network, propagating this information back down through the system would help to reduce errors at the lower levels and therefore demonstrate a reduction in bias over time.

5.3.3. Alternatives to the exponential model and associated illusion predictions

As outlined in the introduction (Chapter 1), a number of low-level models of vision exist that either emulate the filtering operations of cells in early visual areas (including the retina and

LGN) or that capitalise on the statistical regularities (such as contrast distributions) of images found within the natural environment. We now revisit some of the most successful of these models to provide a comparison with our exponential filter model and to provide illusion predictions in these alternate models. We consider the difference-of-Gaussians (DOG) model (Blakeslee and McCourt, 1997) and the oriented-difference-of-Gaussians (ODOG) (Blakeslee and McCourt, 1999, 2001, 2004, 2008; Blakeslee *et al.*, 2005, 2008; Robinson *et al.*, 2007), which have been used to model the same range of lightness illusions as our third study (Chapter 4). Our third study included the predictions made by these models for the same set of lightness illusions that we covered. In this subsection we discuss recent extensions of the DOG model (Cope *et al.*, 2013, 2014a, 2014b) for comparison with our model and propose predictions for other illusions in these DOG-variant models versus ours. We also provide more in-depth discussion of Dakin and Bex's (2003) model.

Our exponential filter model differs from existing models in two key ways – either in the filtering operations it employs or in the normalisation procedure. The original DOG model used a set of isotropic filters (Blakeslee and McCourt, 1997). An orientation component was added to account for White's illusion in Blakeslee and McCourt (1999). Subsequent studies using the ODOG model showed its versatility in being able to account for a large series of illusions (Blakeslee and McCourt, 1999, 2001, 2004, 2008; Blakeslee *et al.*, 2005, 2008; Robinson *et al.*, 2007). The recent modelling work of Blakeslee and McCourt has returned to non-oriented DOG filters (Cope *et al.*, 2013), where such isotropic filters would best describe retinal ganglion and LGN cell operations (Kuffler, 1953, 1973). Recent models presented by Blakeslee and McCourt have also adopted greater biological plausibility, incorporating contrast gain control (Cope *et al.*, 2014a, 2014b). New versions of the DOG model are therefore more similar to our exponential filter model than versions of ODOG, in that they employ gain control mechanisms and that they use filters that are not orientation-selective. The main difference

between these two systems lies in the kernel functions applied to the input image, on which we elaborate below.

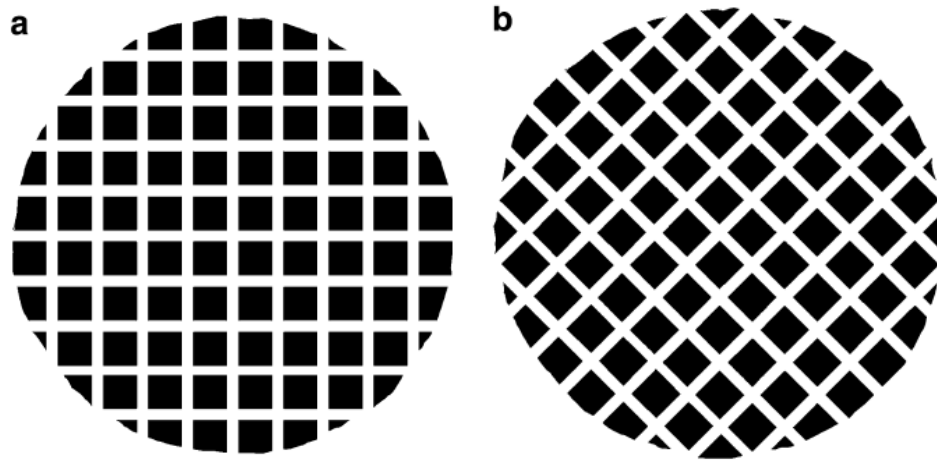


Figure 5-2: Hermann grid illusion, reproduced from Lafuente & Ruiz (2004).

a) Standard Hermann grid b) Tilted Hermann grid

One key feature that differentiates our model from existing models is our use of exponential filters, Dakin and Bex (2003) use Laplacian of Gaussian or LoG filters, which best approximate the Hodgkin and Huxley (1952) “Mexican hat” function. DOG filters provide an approximation for LoG filters (Klette, 2014, p.75), having circular symmetry that produces identical filtering magnitudes at any rotation angle. This property of isotropy is what differentiates our exponential filters from DOG or LoG filters. Turning our attention to the ODOG model, we see that ODOG uses oriented filters presented at 6 different orientations (Blakeslee and McCourt, 2001). The authors describe these filters as being similar to simple cells in the cat or monkey, being selective to orientation and spatial frequency. Orientations in ODOG are weighted after the response of each orientation filter is calculated, so as to produce equivalent activity levels across each orientation. Comparing the exponential filter model to ODOG, exponential functions require no orientation-specific weighting, having an in-built preference for increasing activation along the cardinal axes.

Our use of exponential filters, which differentiates our model from other existing approaches, could potentially impact on predictions for tilted illusions. Tilted lightness illusions are those where magnitude changes as a function of orientation, usually showcasing a decrease in magnitude is usually observed. Examples of tilted lightness illusions include the tilted Hermann Grid (de Lafuente and Ruiz, 2004, Figure 5-2), the scintillating grid illusion (Qian *et al.*, 2009) and the jaggy diamonds illusion (Kawabe *et al.*, 2010). Some illusions, however are unaffected by orientation (Hamburger and Shapiro, 2009) and these could be straightforwardly modelled using isotropic filters without loss of generality. In order to account for a reduction in magnitude for tilted lightness illusions, a model would need to incorporate anisotropic filters. DOG and LoG models produce an even distribution of activation levels in all directions, predicting the same magnitude of an effect in all directions and be unable to account for tilted phenomena. Exponential and ODOG filters could both potentially account for tilted lightness illusions considering that both types of filters are anisotropic. As pointed out above, many orientations within ODOG are equivalently weighted despite being individually represented, which would create difficulties in simulating changes in magnitude for oriented illusion forms.

Considering exponential filters, these show strong activation levels for stimuli presented along the cardinal axes and reduced activation at obliques. In relation to the biological analogy between horizontal retinal cell activation and the exponential function (Packer and Dacey, 2002, 2005), we emphasise that the exponential model shows increased activation along the cardinal axes and reduced activation at obliques - such that levels of activation are still present for oriented stimuli and simply produce a smaller magnitude compared to stimuli presented along cardinal axes. Having shown greater activation levels for cardinal axes, we hypothesise that exponential filters may predict larger magnitudes for illusions presented at

horizontal and vertical orientations than illusions presented at an angle. This provides predictions in line with tilted illusions such as the Hermann Grid (de Lafuente and Ruiz, 2004). Therefore, one possibility for differentiating between low-level visual models, including LoG, DOG, ODOG and the exponential filter model, may be in their prediction of tilted lightness illusions.

5.4 Possible improvements and future studies

5.4.1. Evaluating theories using computer models

As pointed out in the introduction (Chapter 1), the link between theories and models is not straightforward. Norris (2005) captures this succinctly in his statement that ‘there is rarely a straightforward one-to-one mapping between model and theory’. Starting from the perspective of a particular theory and looking at how this can be implemented in a model requires a number of stages. To explore how to interpret theories in computer models, we take the example of the misapplied size constancy scaling theory of the Müller-Lyer illusion (Thiéry, 1895; Woodworth, 1938; Gregory, 1963) and break it down how this can be evaluated from a number of studies, including our first study (Chapter 2, Zeman *et al.*, 2013).

Inappropriate size constancy scaling was highlighted in Gregory (1963) as the dominating factor that causes the Müller-Lyer illusion. Size constancy scaling refers to the way in which our visual system relates the size of retinal images to their real-life object size, based on the principle that the physical size of an object remains constant despite the size of its image changing on our retina. Using depth cues to judge the distance to an object, we implicitly scale the sizes of all objects that we see. So if depth is perceived erroneously, then the judged size of the object will also be incorrect. Observers commonly encounter two types of display: outside corners (Figure 5-3 left) and inside corners (Figure 5-3 right). If the inside corner is

perceived as further away than the outside corner, then the constancy scaling would be different for each figure. Given their identical retinal size, the display perceived as more distant (being the outside corner with arrow-tails) would be perceived as larger than the more proximal display (the inside corner with arrowheads). Gregory (1963) proposed that the presence of arrow-tails or arrowheads appended to the ends of lines present misleading depth cues when placed out of context. Figure 5-3 illustrates this misapplied size constancy scaling theory (reproduced from Gregory, 1966).

Unpacking Gregory's theory, there are a number of elements to consider. To provide a implementation of misapplied size constancy scaling theory in a model, one prerequisite may be for an artificial visual system to have some representation of depth. An artificial model may also need to view many examples of real world images, such as that presented in Figure 5-3, in order to extract depth information. The previous statement is simply one interpretation of Gregory's theory. Some researchers may assert that exposure to 3D natural scenes is not mandatory for Gregory's theory to hold true, but that it is only necessary to retain the underlying image statistics inherent in such images and present these to an artificial neural network. Taking this interpretation based on image statistics, which is abstracted away from exposure to 3D scenes, infers that training a model by exposing it to artificial environments (such as those presented by Corney and Lotto, 2007) would equate to an implementation of Gregory's theory. Howe and Purves (2005a) demonstrate that the MLI can be driven by the statistics of object features coupled with their distances to an observer. Although they concluded that the MLI arises from the statistical relationships between visual stimuli and their real-world sources, they also demonstrated that straight-edged corners, such as those shown in Figure 5-3, contribute minimally to the MLI effect. They conclude that while Gregory's intuition that linking retinal images to their sources is a main driver for the effect, straight-edged corners are not a dominating factor.

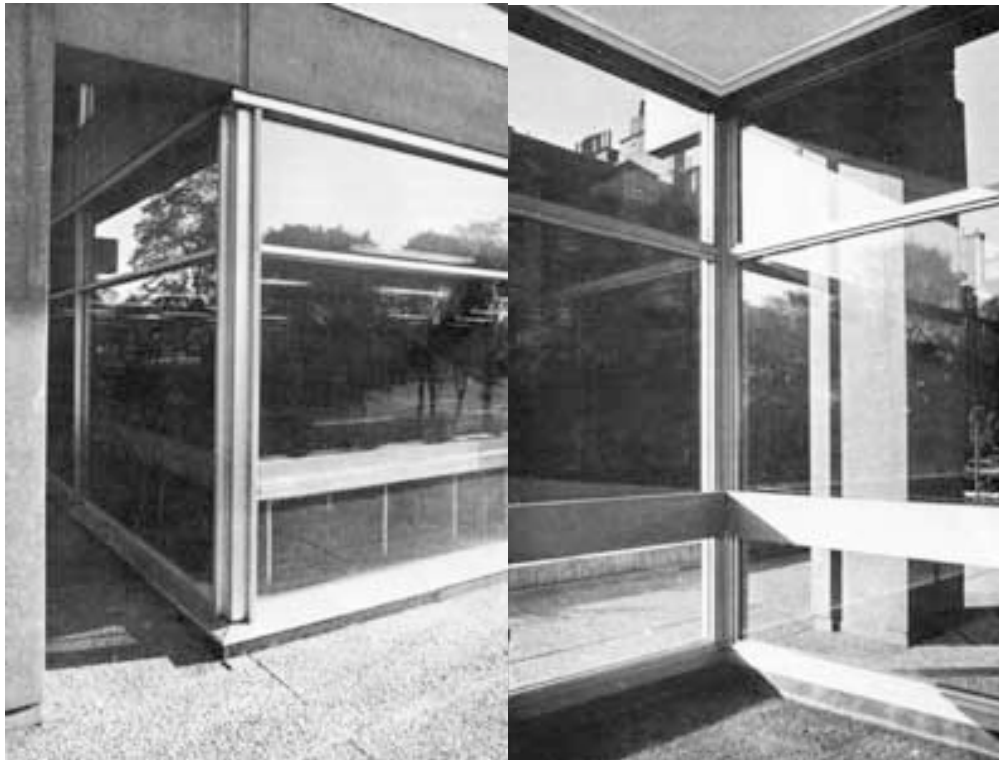


Figure 5-3: A visualisation of misapplied size constancy scaling theory, where “regions indicated by perspective as distant are expanded, while near regions are shrunk”.

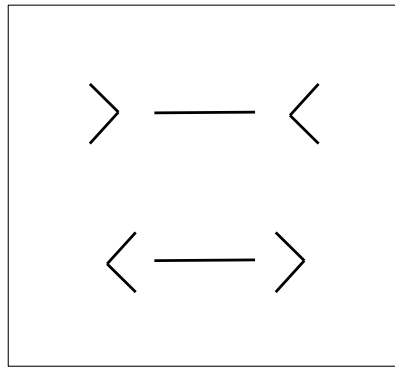
Reproduced from Gregory (1966).

In this example, it is difficult to say whether Howe and Purves' (2005a) study supports the theory of misapplied size constancy scaling or not. Interpretation about the involvement of architectural influences is just one instance of determining the link between computer model results and support for or against a theory. Whether or not Howe and Purves' (2005a) study demonstrates support for Gregory's theory, this example highlights one of the benefits of computational modelling in light of theories in general. This work shows how computer models can expose assumptions (the necessity of straight edged corners) and push for more accurate definitions of theories.

5.4.2. Extensions and limitations of the HMAX model

5.4.2.1. *Possible follow-up experiments on the Müller-Lyer illusion in HMAX*

There are a number of variations and manipulations of the Müller-Lyer (Brentano, 1892; Cooper and Runyon, 1970; Greist-Bousquet and Schiffman, 1981; Predebon, 1992, 1994) that could provide further comparisons between humans and artificial networks, or even between different models of explanation. For instance, Predebon (1992, 1994) demonstrated a reversed Müller-Lyer effect when the wings of the figure were amputated from the centre shaft and displaced horizontally away from the shaft (see Figure 5-4). By testing these images using HMAX, we would be able to determine the extent to which modelling the MLI in HMAX is able to generalise to other forms of the illusion. If bias is produced in the model for modified forms of the MLI, we would again be able to separate the influence of statistics (at the input level) from the influence of filtering operations (at subsequent levels within HMAX), much like what was shown in Chapter 3 (Zeman *et al.*, 2014). This would allow us to determine the extent to which the reversed form of the Müller-Lyer is driven by statistics versus determined by filter processing. If bias is observed within the model, we would also be able to test whether conjoined features, such as wings attached to the shaft, directly contribute to the MLI effect in the model, or whether the effect is more reliant on simpler features such as single or double lines. This second point emphasises the influence of higher-level features, and their corresponding neurological counterparts, on perception of the MLI. These experiments would elucidate more information about causes of the MLI. From our previous study, we found that the main driver of the MLI effect was the underlying correlations between feature statistics and line length estimation (Chapter 3, Zeman *et al.*, 2014). We would be able to determine whether this explanation can be extended to account for reversed forms of the MLI, or whether a different explanation needs to be sought.



**Figure 5-4: MLI configuration with wings displaced away from shaft
(Predebon, 1992, 1994)**

5.4.2.2. Experiments on other illusions in HMAX

We considered modelling other line length illusions in HMAX, such as the Ponzo illusion (Ponzo, 1911), which, like the Müller-Lyer illusion involves two parallel horizontal lines of equal length that are flanked on the outside by two converging oblique lines (see Figure 5-5 top row). To test the Ponzo illusion using HMAX, we would construct a binary classification task, much like that presented in Chapter 2 (Zeman *et al.*, 2013). We would present two cases to the machine learner: a long case, where the top horizontal line is longer than the bottom (Figure 5-5B), and a short case, where the top horizontal line is shorter than the bottom (Figure 5-5C). To measure the size of the illusion we would adjust the length of the inside horizontal lines (Fisher, 1968). Other manipulations of the outer oblique lines that have been shown to affect illusion strength in humans, such as their length, angle and distance to the horizontal lines (Jordan and Randall, 1987), could also be employed.

Before testing any illusory figures using HMAX, a training set of images would need to be defined as well as a control image set. see Figure 5-5D provides one example of a training image that we predict would not induce any illusion but would expose a machine learning

algorithm to all possible features that may be present in any one image. As with the illusory test images, a number of aspects of the training images could also be manipulated, including the length of the inside lines, the distance between the inside lines, the length of the outer lines, the angle of the outer lines, the distance between the inner and outer lines, or the global positioning of the unified figure. Figure 5-6 demonstrates these manipulations using a series of example images. Once HMAX had completed training using these images, it would then be presented with a set of control images (see Figure 5-5 E, F) to ensure that it is able to complete a baseline task of line length judgment with an acceptable level of performance. Demonstration of the illusion in HMAX would provide evidence that feedback is not necessary to bring about the effect.

The Ponzo illusion provides an interesting case in that it is classed within the same category as the Müller-Lyer according to Gregory (1963, 2005). Gregory (1963, 2005) assigns both of these illusions as being potentially caused by features that may indicate depth through an observer's perspective. The Ponzo illusion is often referred to as the "railroad tracks" illusion, which immediately makes apparent the link between environmental images and the illusion (see Figure 5-7, reproduced from Gregory, 1968). In the Ponzo illusion, the horizontal line that would appear further away when laid on railroad tracks is perceptually enlarged compared to the horizontal line that would appear nearer. This relates again to Gregory's misapplied size constancy scaling theory, where visual depth cues can cause the enlargement or contraction of an object in order for the size of that object to be coherent its placement in 3D space. However, just as with the Müller-Lyer illusion, there are conflicting accounts as to the source of the illusion. Prinzmetal *et al.* (2001) propose another theory to account for the Ponzo illusion, that they call tilt constancy theory. They propose that the Ponzo illusion is as a result of misperceiving orientation induced by local visual cues (see Figure 5-8, reproduced from Prinzmetal *et al.*, 2001). They suggest that the mechanisms that normally help us

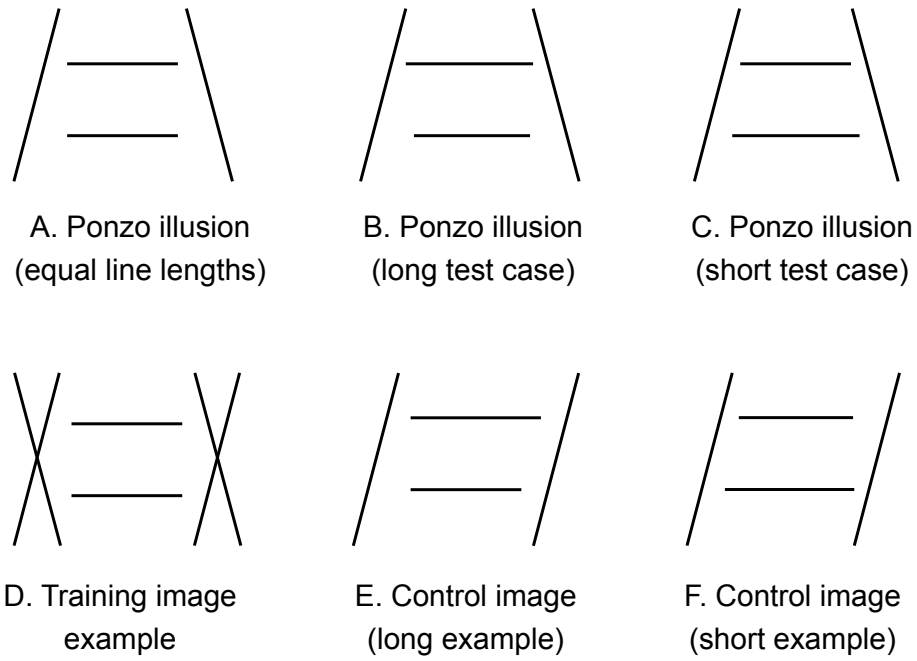


Figure 5-5: The Ponzo illusion (top row) with corresponding training and control images for testing in HMAX. The long case is where the top line is longer than the bottom line. The short case is where the top line is shorter than the bottom line.

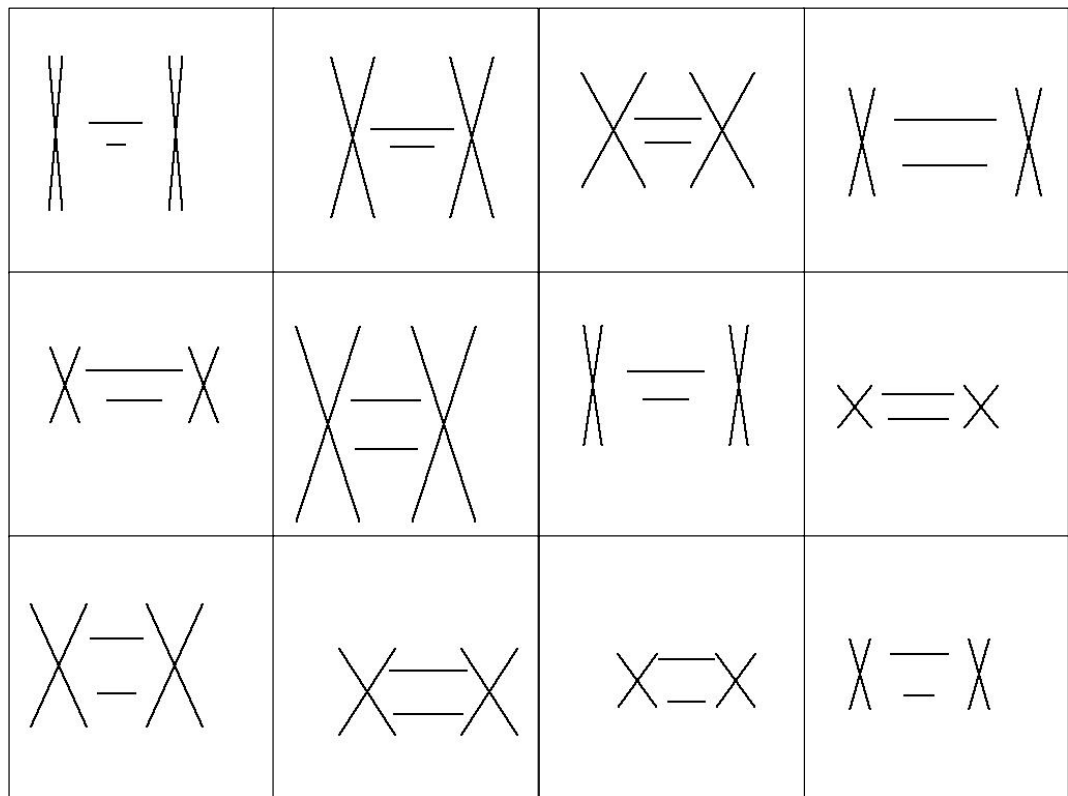


Figure 5-6: Some example training images for testing the Ponzo illusion in HMAX

perceive constant orientation despite changes in retinal or body orientation (tilt constancy) underlie the effect. Take Figure 5-8A, where the top dot appears to the left of the bottom dot due to the tilt induction effect. This effect also causes the top dot to appear slight to the right in Figure 5-8B. Combining these two effects causes the Ponzo effect in Figure 5-8C. Other explanations for Ponzo illusion are based solely on low-level mechanisms, such as low-pass filtering (Ginsburg, 1984). Ginsburg (1984) suggests that stronger weighting placed on low spatial frequencies over higher spatial frequencies is the main driver for the Ponzo effect. Ginsburg's low-pass filtering explanation extends to other illusions, including the Müller-Lyer (Ginsburg, 1978). Our study in Chapter 2 (Zeman *et al.*, 2013) demonstrated that heavier reliance on low-spatial frequency information (large sized filters) over high-spatial frequency (small sized filters) information was not responsible for the Müller-Lyer effect we found in HMAX.

By testing the Ponzo illusion in HMAX we can separate the influence of depth cues present at the input level from filtering mechanisms present within the network. If we are able to reproduce the Ponzo using HMAX, we can assert that misapplied depth cues are not necessary for bringing about the illusion, since the model does not account for any depth information. This allows us to rule out Gregory's (1963) theory of misapplied size constancy scaling as a cause of the Ponzo illusion in the model. If bias for the Ponzo is present in the network, we can determine which features are primarily responsible for the effect. We can separate out the influence of low-spatial frequency information from high-spatial frequency information in making the final classification decision, much like that presented in Zeman *et al.* (2013). If low-spatial frequency information (shown by the activation of smaller-sized filters) receives a higher weighting than higher spatial frequency information (shown by the activation of large-sized filters), then this provides support for Ginsburg's (1984) account of the Ponzo illusion

in the model. To show whether the Ponzo effect was due to tilt constancy scaling, we would look at the orientation filters used to make a decision in the final classification layer. If heavy



Figure 5-7: The Ponzo illusion overlaid on a set of railroad tracks.
Reproduced from Gregory (1968).

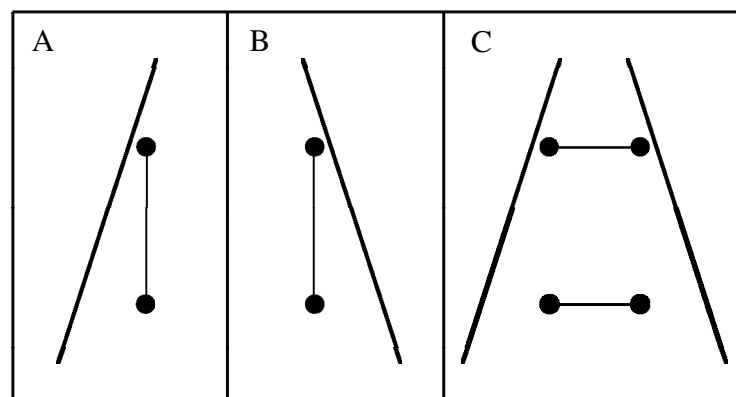


Figure 5-8: Prinzmetal et al.'s (2001) demonstration of the tilt induction effect as an explanation for the Ponzo illusion. A. The top dot appears slightly left to the bottom dot because of the tilt induction effect. B. The top dot appears slightly to the right of the bottom dot. C. Combined together, the misplacement of the endpoints cause the Ponzo illusion. Reproduced from Prinzmetal *et al.* (2001).

reliance is placed on the vertical orientation nodes in making a decision at the final layer, then this would indicate that Prinzmetal *et al.* (2001)'s tilt constancy scaling account is responsible for the illusion in HMAX.

5.4.2.3. *What illusions cannot be modelled in HMAX?*

One of the obvious restrictions of HMAX is that it is limited to feed-forward processing. While this constraint allows the model to be appropriate for the modelling of rapid visual classification (within the first 100-200ms) (Serre *et al.*, 2007), HMAX would be unable to model any visual phenomena that require feedback or contain some form of temporal dynamics. Looking at Gregory's (1997) taxonomy of illusions, we can generate hypotheses about which illusions can and cannot be simulated in HMAX. Gregory (1997) proposes four classes of illusion that are labelled as ambiguities, distortions, paradoxes and fictions (see section 1.X). We use each of these illusion categories as a rough guideline for investigating those that can be simulated in HMAX, identifying modelling requirements for emulating human performance within each category. If HMAX is not sufficient, we propose an alternative model that is capable of reproducing the illusion class.

Illusions of ambiguity involve images that give rise to multiple percepts that change over time, despite the stimulus remaining constant (Gregory, 1997). For example, the Necker cube (Figure 5-9A) is a line drawing of a transparent-faced cube, where the surfaces that are perceived as the front and back switch with viewing time (Necker, 1832). Other ambiguous illusions include figure-ground illusions such as the Rubin face-vase illusion (Figure 5-9B), where depending on your assignment of the main figure versus the background, an observer can perceive either a face or a vase (Rubin, 1915). We consider what features would be

advantageous to model these illusions computationally. It would be favourable to be able to model the activation of the network across time, so as to demonstrate an oscillation between representations. For example, a network may switch from the face description to the vase description as being more likely at any given time point, without settling permanently on any one stable representation.

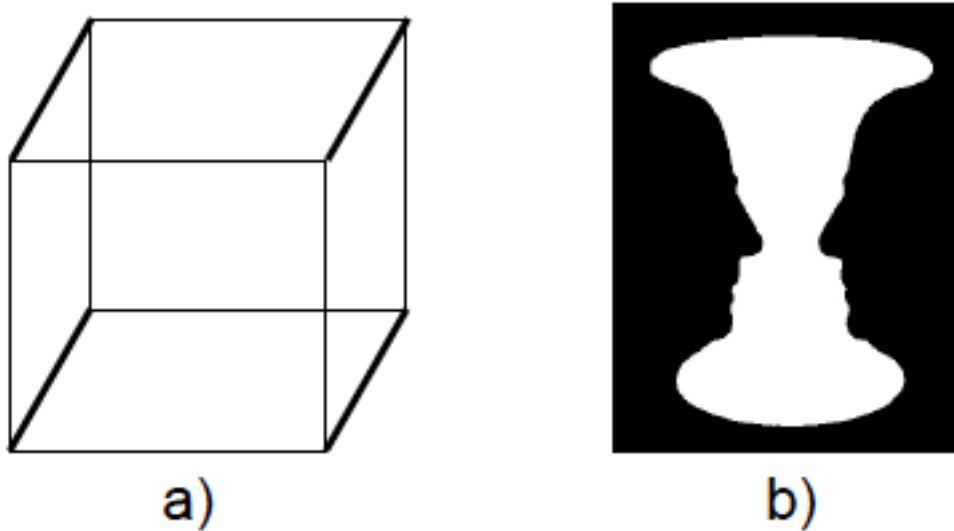


Figure 5-9 Ambiguous Figures. a) Necker cube b) Face-vase illusion

We turn our focus now to HMAX, to assess its suitability for modelling illusions of ambiguity. Within HMAX, the network may trigger multiple nodes that are activated layer by layer up to the C2 level. Then the final stage would select the most likely classification given the input, using a winner-take-all strategy. HMAX has no temporal modelling, so exposing the network to an ambiguous stimulus would produce a final percentage classification of the stimulus representing one percept versus another, e.g. the face versus the vase. This percentage split of classifications would be the metric used to assess the overall success of the model. A human experiment could be run that presents observers with discrete, brief-duration exposures of the face-vase that would be used to measure the percentage with which they

classify one percept versus the other. The overall trial results of human observers could be compared to the classification split obtained when the model is presented with the same stimuli.

Using HMAX to model ambiguous illusions, although possible, is far from an ideal. The limitation of HMAX in not being able to model fluctuations in network dynamics across time make it unsuitable as the best modelling approach for this class of illusions. Instead, models incorporating feedback would work best in this scenario. Probabilistic generative models, which we introduced in Chapter 1 Section 1.9.4, would optimally satisfy the prerequisites for modelling this class of illusions, being able to generate likelihoods for multiple representations based on the state of the network at each point in time. Examples of such probabilistic generative models include Lee and Mumford (2003), Kersten *et al.*, (2004), Friston (2005a, 2008, 2010, 2012), Brown and Friston (2012), Brown *et al.*, (2013).

Illusions of distortion are those that produce a perceived enlargement or contraction of space. These illusions include the Müller Lyer (Figure 1-1A), which was successfully modelled in our first two studies. Others illusions of distortion include the Ponzo illusion (Figure 5-5A), discussed in Section 5.4.2.2. This class of illusions may be the simplest to model within HMAX, since the classifier can be trained to distinguish between categories. HMAX may in fact provide an integrated way to model some of these illusions, given that filtering-only explanations, or accounts based on natural scene statistics, may be sufficient. HMAX is predominantly useful in this scenario for exploring the interaction between these two explanations. Referring to other feedforward models, the SpikeNET architecture (VanRullen *et al.*, 1998; Thorpe *et al.*, 2001; VanRullen and Thorpe, 2001a, 2001b, 2002; VanRullen *et al.*, 2005) would also be capable of modelling this type of illusion. However, as for HMAX,

emulation of the temporal dynamics of viewing the Müller-Lyer (Coren & Porac, 1984; Predebon et al. 1993; Predebon, 1998, 2006) would not be possible in such a feed-forward architecture. A generative model would be needed to capture the rates of decrement over time for both the arrowhead and arrow-tail forms of the illusion, as demonstrated in Predebon (2006). It would only be possible to model top-down effects, such as those of selective attention, using feedback models. Goryo *et al.* (1984) measured the effect of selective attention on Müller-Lyer figures, finding that attentional mechanisms increase bias to a greater extent for distortions of contraction versus expansion. A generative model would be needed to simulate the effect of top-down activation signals on the magnitude of illusory bias to compare with human results.

Illusions of paradox are those that present an overall percept that is incoherent or impossible despite coherency being demonstrated at a local level. Paradoxes include the Penrose triangle (Chapter 1, Figure 1-1B) which illustrates coherent individual edges and corners of a 3D triangle that join to form an impossible 3D object (Penrose & Penrose, 1958). In order to model paradoxes, a network would need to read in an image, activate nodes to generate a high-level representation of this input (e.g. a solid, 3D triangle) and then reactivate lower-level nodes using feedback. A mismatch between nodes that are activated directly by the stimulus and nodes that are activated from higher levels in the network (also originating from that same stimulus) could indicate a paradox. Again, probabilistic generative models appear to be the most suitable solution for emulating human perception of this class of illusion. Using this type of model, lower-level activations would be generated based on higher-level representations in the network. The top-down activations or “fantasy” would then be used to assess the extent to which it matches up with the initial stimulus. The discrepancy between the fantasy and the stimulus would be quantified as a mismatch error. With coherent objects presented as input, there would be a small error term. For a paradoxical illusion, there will

always be a mismatch between input and fantasy, producing a recurring error given the same stimulus. For paradoxical objects, the error term would remain high. Feedback models would allow for the error term to be measured over time, propagating back and forth through the network and showing no resolution between top-down representations activated at higher-levels and the bottom-up, stimulus-driven representations. Feed-forward models such as HMAX would not be capable of activating representations from higher to lower levels of the network and would therefore be unsuitable for demonstrating paradoxical illusions.

Illusions of fiction are those that generate imaginary contours and surfaces that are not present in the stimulus. The Kanisza triangle is one example of an illusion of fiction, where three circles, each having a triangular cutout, are strategically placed so that the apices of their cutouts align like three corners of a triangle (Chapter 1, Figure 1-1D). Fictional illusions would once more be most suitably modelled in a probabilistic generative network, given that they would be able to generate phantom percepts through feedback mechanisms. Taking the Kanisza triangle as our example, observers of the illusion perceive a large triangle that is the same colour as the background occluding the three circles. A model able to reproduce this percept would need to be capable of storing multiple object representations at once, would need to recognise that perceiving fully closed circles is more likely than perceiving circles with missing wedges, and then fantasise that the most likely explanation for viewing these objects is for a triangle to be present that occludes the other three objects. Feed-forward networks such as HMAX would be unable to generate fictional representations. However, it may be possible to train HMAX into recognising triangles, circles and other shapes and then test its ability to classify an object that is fictional.

In summary, we have analysed example illusions of fiction, paradox, distortion and ambiguity that constitute Gregory's schema. As made evident in Chapters 2 & 3 (Zeman *et*

al., 2013, 2014), HMAX is able to model an illusion of distortion, namely the Müller-Lyer, and could be adapted to account for other variations of the Müller-Lyer (Section 5.4.2) as well as the Ponzo illusion (Section 5.4.2.2). As pointed out in the preceding paragraphs, HMAX however, would not be suitable for modelling other illusion classes. Instead, probabilistic generative networks or other feedback models would be required to model illusions of fiction, paradox and ambiguity.

5.4.1. Extensions and limitations of the exponential filter model

Our exponential filter model was able to demonstrate a range of lightness illusions (Chapter 4, Zeman *et al.*, *in submission*), raising some follow-up questions: What other illusions could be predicted by this exponential filter model? Are there certain limitations within the model that would prevent the simulation of particular illusions? To address these questions, we first address some of the constraints of pre-cortical models before looking at certain classes and examples of illusions that could be reproduced within these networks.

An obvious limitation of low-level visual models is in their capability to model complex features, for example, oriented edges and shapes. Interestingly, many of the illusions successfully emulated in the exponential filter model contained edges or enclosed shapes, such as the simultaneous contrast effect (Blakeslee & McCourt, 1999), the checkerboard illusion DeValois and DeValois (1988), radial White's illusion (Anstis, 2003) and the Todorovic (1997) illusion. This highlights the independence between assessing lightness of a stimulus patch and more complex properties of the stimulus, such as the presence of edges or gratings. Therefore, illusions that contain edges or shapes can still be simulated using

exponential filters. We turn again to Gregory's (1997) categories of illusions to explore whether certain classes of illusion are suited to this particular type of model.

Gregory's (1997) classification schema defines illusions not just by their appearance, but also by their proposed aetiology, separating physical explanations from cognitive ones. As mentioned previously in Chapter 1, Gregory defines four causes of illusions, divided into explanations that are based on optics, signal processing, the recruitment of domain-specific rules or object knowledge. Rule and knowledge-based explanations would involve cortical mechanisms and so would not be applicable here. We also rule out explanations that are based on optics (such as those that rely on the physics of light waves), since they are created outside of the system we are trying to model. Hence, illusions driven by signal processing are most suitable for modelling by low-level models such as our exponential filter model. We now look at specific examples of these illusion cases. Many of these examples are ruled out in the models that we recruit, since the scope of these networks does not involve motion or colour and would not receive multiple input images. However, some illusions that Gregory lists can be simulated within the exponential filter model, and we describe these below.

We now turn to Gregory's examples of signal processing illusions, which involve low-level mechanisms. These include after images (where an image continues to appear in one's vision after exposure to the image has ceased) and the Café wall (where alternating black and white tiles separated by grey mortar induce tilt, as illustrated in Figure 5-10). Using a pre-cortical network, it may be possible to model both after images and the Café wall illusion. In order to simulate after images, it would be necessary to include a temporal component in a model in order to account for adaptation with the overstimulation of neurons. Therefore, the exponential filter model would not be suitable for modelling after images. The Café wall illusion (Gregory & Heard, 1979) provides an interesting example of induced perceived tilt

that is dependent upon the luminance of the mortar lines (see Figure 5-10). A closely related effect is the illusion of striped cords (Kitaoka, 1998), Figure 5-11, with additional variations of this ensemble of illusions being presented in Kitaoka et al. (2004). All of these stimuli could be presented to the model and filter values observed along tile intersections. Having an orientation read-out in a model would be necessary for simulating these tilted illusions. Taking a network such as HMAX, an image would be presented to it and the maximally activated neurons would indicate the most likely orientations within the image.

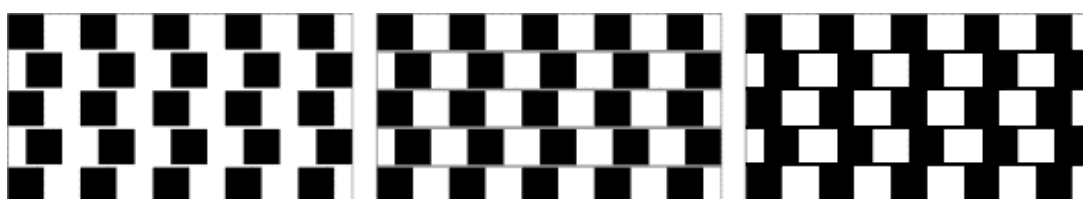


Figure 5-10: The café wall illusion, reproduced from Gregory & Heard (1979), demonstrating how changes in luminance of mortar between the black and white tiles affects perceived tilt.

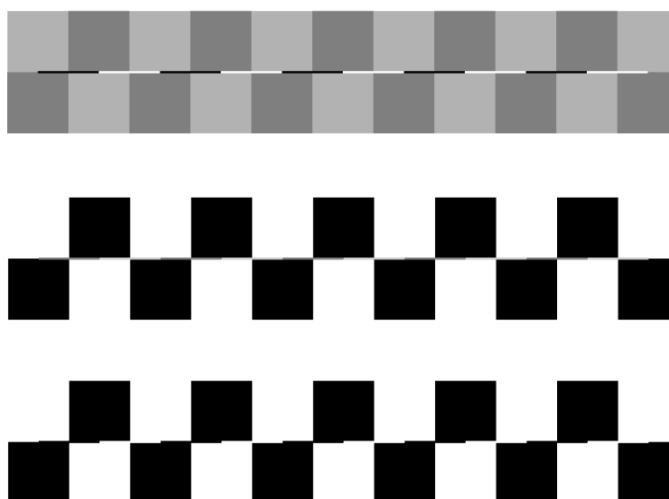


Figure 5-11: Striped chord illusion, reproduced from Kitaoka (1998), demonstrating changes in luminance of the tiles or of the chord affects the magnitude of perceived tilt.

Moving away from Gregory's schema and turning to other examples of illusions that could be modelled in the exponential filter model, we turn our attention to what differentiates our model from other existing pre-cortical models. This allows us to distinguish which particular illusion predictions would differ in our model compared to others. The filters used in our model (Chapter 4, Zeman *et al.*, *in submission*) are differentiated from DOG filters in one key way – the exponential filters that we recruit are anisotropic (Basu and Su, 2001). This may have an impact on oriented forms of lightness illusions, some of which have been shown to be influenced by the angle at which they are presented. Examples of tilted lightness illusions were mentioned earlier in Section 5.3.3, including the tilted Hermann Grid (de Lafuente and Ruiz, 2004), the scintillating grid illusion (Qian et al., 2009) and the jaggy diamonds illusion (Kawabe et al., 2010). Let's take as an example the Hermann Grid illusion (Figure 5-2), where a set of vertically and horizontally aligned black squares on a white background can induce phantom grey areas at intersections (Hermann, 1870). The Hermann Grid illusion has been shown to be orientation dependent, with the effect at oblique angles being roughly a third of the magnitude compared to presentations with horizontal and vertical configurations (de Lafuente and Ruiz, 2004). A model with isotropic filters would fail to predict differences in magnitude for tilted configurations. Blakeslee and McCourt (1997) demonstrate predictions for the standard form of the Hermann Grid in their original DOG model, but they did not test the Hermann Grid at oblique orientations, nor did they assess any form of the illusion in subsequent ODOG models. It would be interesting to see whether the ODOG model with oriented filters would be capable of predicting both the standard and tilted forms of the Hermann Grid.

5.5 Applications for computer modelling of illusions

5.5.1. Autonomous systems and engineering applications

The possibility for computer vision systems to misjudge the size or length of objects has implications for distance judgments in computerised systems that depend upon camera information. Automated navigation and landing systems rely on sensory information to calculate distances to known objects and orientate themselves within the environment. In recent years, driverless cars, such as the Google car, have been heavily researched and tested, motivated by providing one of the safest transport options, removing dangers for motorists and pedestrians on top of reducing traffic congestion and fuel intake (Guizzo, 2011). Driverless vehicles use mounted camera images alongside radars, GPS and wheel sensors to calculate the position and orientation of the vehicle (Guizzo, 2011; Whitwam, 2014). The heights of objects that are aligned with straight-edged features, such as roads, trees or buildings, could be calculated incorrectly by an artificial visual system and therefore provide erroneous information to the navigation systems.

In our second study (Chapter 3, Zeman *et al.*, 2014), we demonstrated a reduction in errors (measured as bias and uncertainty) through modelling the MLI in HMAX. We found that operations performed by complex cells reduced bias and uncertainty errors in the majority of cases. We hypothesised that by capitalising on the properties of complex cells, that is, increasing the level of positional variance in input images, we would be able to produce a similar pattern of error reduction. This hypothesis was confirmed, as both bias and uncertainty decreased. Other artificial systems that require accurate length estimations may also benefit from increasing variance in their training images and recruiting complex cell operations within their network. By taking into account potential misjudgements related to illusions of line length, these systems could learn to compensate for such errors and provide systems that are more accurate and more reliable, and therefore safer.

5.5.2. Links with psychological disorders

Visual illusions are a useful tool for assessing the integrity of basic sensory processing mechanisms for people with psychological disorders (Happé, 1996; Ropar and Mitchell, 1999; Silverstein & Keane, 2011; Notredame *et al.*, 2014). Furthermore, they also present a way to quantify the effectiveness of ongoing treatment. By simulating illusions using computer models of the visual system, it may be possible to manipulate parameters within an artificial network to reflect some of the compromised visual processing mechanisms associated with particular cognitive disorders. By matching models to the visual experiences of clinical populations, we can gain a deeper understanding of these disorders, make predictions for other visual experiences of clinical populations and ultimately provide support for and against theories of these disorders.

In this section, we look at schizophrenia as our main example, since both the Müller-Lyer illusion and lightness illusions have been extensively studied in these populations. People with schizophrenia have been shown to have greater susceptibility to the Müller-Lyer illusion compared to non-schizophrenic mental patients and normal populations (Weckowicz and Witney, 1960; Letourneau (1974), Capozzoli & Marsh, 1994; Pessoa *et al.*, 2008; Kantrowitz *et al.*, 2009). By greater susceptibility, we mean that schizophrenics demonstrate larger biases (measured as PSEs) compared to non-schizophrenics. As a result of these early studies, some visual illusions, including the Müller-Lyer, were considered as effective diagnostic tools of schizophrenia, over more subjective, verbal assessments (Cromwell, 1975; Cromwell & Pithers, 1981; Pessoa *et al.*, 2008).

In the last decade, however, most of the findings that schizophrenics have increased levels of bias for visual illusions have been overturned. Researchers have shown that schizophrenics can also demonstrate reduced susceptibility, or even a complete eradication of bias, for other

illusions than those mentioned previously (Dakin *et al.*, 2005; Dima *et al.*, 2009; Kantrowitz *et al.*, 2009; Barch *et al.*, 2012; Notredame *et al.*, 2014). Dakin *et al.* (2005) was one of the first studies to demonstrate reduced bias in a contrast-contrast illusion, where a textured annulus perceptually appears to have a reduced contrast level when surrounded by an outer textured annulus, compared with no surround. Follow-up neurological studies demonstrated that reduced bias for the contrast-contrast illusion was predominantly a result of reduced contrast surround suppression for schizophrenics (Fogelson *et al.*, 2011, Seymour *et al.*, 2013). Dima *et al.* (2009) found that schizophrenics do not experience the hollow-face illusion. Later follow-up studies explored the neurological components of this finding, that it was due to impaired top-down processing (Dima *et al.*, 2010, 2011). These recent studies investigating illusions in schizophrenics demonstrate two key points. Firstly, that illusions provide a useful diagnostic tool for schizophrenia, taking into account increased or reduced bias levels that is dependant upon the illusion presented. Secondly, that illusions can provide insight into the neural processing of clinical populations.

We now consider the models and illusions that have been studied in this thesis in light of schizophrenia. Reflecting on our work in modelling the Müller-Lyer illusion in HMAX, it is possible to manipulate levels of lateral inhibition and decrease these to reflect the reduced levels that have been proposed to be the main cause by Must *et al.* (2004), Dakin *et al.* (2005) and Robol *et al.* (2013). Robol *et al.* (2013) suggest two main forms of inhibition to be considered for visual modelling of schizophrenics: local, tuned suppression and long-range intrinsic inhibition. In some preliminary experiments that were run to determine the optimal levels of long-range inhibition between neurons in HMAX, we found that the default setting provided the highest level of performance for the control task. When levels of long-range inhibition were lower or higher than the default setting in HMAX, we found that the percentage of correct classifications fell for the control task in the model. Interestingly,

reduced levels of long-range inhibition may also reflect the general poor performance of schizophrenics in performing low-level visual tasks (Robol *et al.*, 2013). After reducing lateral inhibition levels in the model, we could observe the levels of bias and uncertainty for the Müller-Lyer and determine whether these increase, in line with the work of Weckowicz and Witney (1960), Capozzoli & Marsh (1994), Pessoa *et al.* (2008) and Kantrowitz *et al.* (2009). Kantrowitz *et al.* (2009) present one of the most recent studies on the MLI in schizophrenics, demonstrating the effect of contrast on a series of line length illusions, including the MLI and Ponzo illusions for schizophrenics versus controls. Kantrowitz *et al.* (2009) reported that increasing contrast levels for the MLI decreased susceptibility to the illusion and that increasing contrast levels for the Ponzo increased susceptibility for healthy populations. Comparing this result to schizophrenics, increasing the contrast levels of these illusions showed an increased bias for the Müller-Lyer and a decreased bias for Ponzo, when compared with the control group. It is possible to manipulate contrast levels for the training and test stimuli in HMAX, allowing us to see if the effect of increasing contrast shows a similar trend in the machine learner as it does in humans. This would allow us to compare model results for both healthy and clinical populations. In HMAX, lateral inhibition levels can also be altered, opening up further experiments to test the relationship between illusion susceptibility and modified neural mechanisms using a computational model. This would allow us to investigate not just the effect of compromised forms of neural inhibition on illusory bias, but it would also allow us to observe the interaction between statistical properties inherent in images and the operations on these. This may provide insight into the visual functioning of schizophrenics, in determining whether image statistics, neural operations, or a combination of these two factors lead to alterations in illusion susceptibility.

Regarding the exponential filter model, one feature that we can manipulate is the gain control that forms part of the normalisation step (Chapter 4, Zeman *et al.*, *in submission*). We now

consider lightness illusions that are thought to be dependant upon gain control mechanisms to see whether modelling these can provide insight into the visual mechanisms of schizophrenics. Schizophrenics have been shown to be less vulnerable to one such phenomenon known as the contrast-contrast effect or Chubb illusion (Dakin *et al.*, 2005 and Barch *et al.*, 2012). The Chubb illusion (Chubb et al, 1989) involves a textured patch with a uniform grey surround (Figure 5-12A), which appears to have a higher contrast than an identical patch with a high contrast surround (Figure 5-12B). The spatial frequency of the surround affects the lightness perception of the target patch, in that changing the surround from high (Figure 5-12B) to low spatial frequency (Figure 5-12C) eliminates the effect. Olzak and Laurinen (1999) demonstrate that multiple gain control processes are present in the Chubb illusion and Zenger-Landolt & Heeger (2003) link the Chubb illusion to levels of gain control in V1 using fMRI. Dakin *et al.* (2005), Barch *et al.* (2012) and Robol *et al.* (2013) have studied the contrast-contrast illusion in schizophrenics and have put forward weak gain control mechanisms as an explanation for why they are less vulnerable to the illusion.

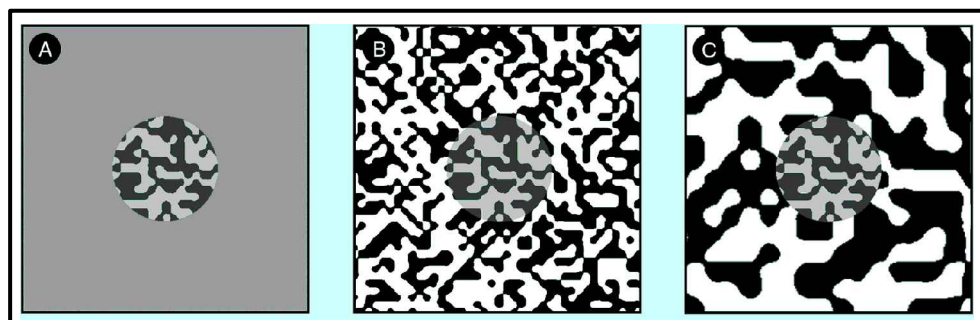


Figure 5-12: The Chubb effect, reproduced from Figure 1, Lotto and Purves (2001). The target patch in A is the same contrast as the centre patch in B, although the target patch in B appears to have a lower contrast. Manipulating the spatial frequency of the surround can eliminate the effect (C).

Within our model, it is possible to manipulate the levels of gain control and observe differences in luminance values of the centre target patch. Looking at examples of the Chubb illusion in Figure 5-12 (reproduced from Lotto and Purves, 2001), it would be possible to create stimuli using combinations of A, B and C to compare the effects of surround contrast and spatial frequency on the target patch. To assess the model's ability to account for the Chubb effect, we would measure the output lightness values for the two different luminances in the target patches, and calculate a single perceived contrast measure using these values. If the exponential filter model is successful in simulating the contrast-contrast effect, this would imply the possible involvement of pre-cortical areas. While levels of gain control for the Chubb illusion have been linked to V1 (Zenger-Landolt & Heeger, 2003), this does not eradicate the involvement of earlier areas such as LGN or the retina. To demonstrate the effect in a V1-like model, it would be more appropriate to use Gabor filters with surround suppression.

5.6 Closing Remarks

The interdisciplinary nature of this research lends itself to improving machine learning techniques as well as gaining a better understanding of psychological processes. While it should be acknowledged that demonstrating an effect in one model does not guarantee that the same causes are responsible for the same effect in another system, by simulating visual illusions in computational models we have eliminated the necessity of some causes. For example, depth information is not necessary to bring about the MLI, because we can simulate the illusion in a model that does not account for depth (Chapter 2, Zeman *et al.*, 2013).

In addition to eliminating some of the necessary causes behind an illusion (Chapter 2, Zeman *et al.*, 2013), we have identified potential sources of bias that may help to improve automated computer vision systems (Chapter 3, Zeman *et al.*, 2014). We hypothesise how potential biases can be overcome, which we then demonstrate in practice (by increasing positional variance in the stimuli to reduce illusory errors of bias and uncertainty). The ability of this technique to reveal potential flaws in, for example, automated navigation systems, and to suggest ways to reduce and even eliminate these biases is just one future application of the work presented in this thesis.

In our final study, we demonstrate how low-level filtering techniques, inspired by the contrast statistics of natural images, can account for a large repertoire of lightness illusions (Chapter 4, Zeman *et al.*, *in submission*). By applying filters of different shapes as well as different sizes, we expand on the current literature that already demonstrates the large influence of low-level processing on lightness illusions. The exponential filters we employ are not orientation selective, demonstrating that V1-style operations are not required to account for many lightness illusions.

As a whole, this dissertation highlights many advantages in applying computational modelling to the study of visual illusions. Together, the studies within this thesis demonstrate that the proposed causes of visual illusions can be separately tested and assessed within a model for their impact on illusory bias and uncertainty. Importantly, through the use of computational models, we are able to identify some of the factors that are not necessary for bringing about certain visual illusions.

5.7 References

- Ahn, L. von, Blum, M., Hopper, N. J., and Langford, J. (2003) CAPTCHA: Telling humans and computers apart. In *Advances in Cryptology, Eurocrypt '03, Lecture Notes in Computer Science, 2656*, 294–311.
- Ahn, L. von, Blum M. and Langford, J. (2004) Telling Humans and Computers Apart Automatically. In *Communications of the ACM, 47(2)*, 56-60.
- Ahn, L. von., Maurer, B., McMillen, C., Abraham, D. and Blum, M. (2008). reCAPTCHA: Human-Based Character Recognition via Web Security Measures. *Science, 321(5895)*, 1465-1468.
- Anderson, B. L. (1997). A theory of illusory lightness and transparency in monocular and binocular images: The role of contour junctions, *Perception, 26(4)*, 419–454.
- Anstis, S. (2003). White's effect radial. Online demonstration retrieved from:
<http://www.cogsci.ucsd.edu/stanonik/illusions/wer0.html>
- Barch, D. M., Carter, C. S., Dakin, S. C., Gold, J., Luck, S. J., MacDonald III, A., Ragland, J. D., Silverstein, S., and Strauss, M. E. (2012). The Clinical Translation of a Measure of Gain Control: The Contrast-Contrast Effect Task. *Schizophrenia Bulletin, 38(1)*, 135–143.
- Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P., & Friston, K. J. (2012). Canonical microcircuits for predictive coding. *Neuron, 76(4)*, 695-711.
- Basu, M. and Su, M. (2001). Image smoothing with exponential functions, *International Journal of Pattern Recognition and Artificial Intelligence, 14(4)*, 735–752.
- Benary, W. (1924). Beobachtungen zu einem Experiment über Helligkeitskontrast. *Psychologische Forschung, 5(1)*, 131–142.
- Bertulis, A., & Bulatov, A. (2001). Distortions of length perception in human vision. *Biomedicine (Lithuania), 1(1)*, 3 - 25.
- Bertulis, A. & Bulatov, A. (2005). Distortions in length perception: Visual field anisotropy and geometrical illusions. *Neuroscience and Behavioral Physiology, 35(4)*, 423-434.
- Blakeslee, B. and McCourt, M. E. (1997). Similar mechanisms underlie simultaneous brightness contrast and grating induction. *Vision Research, 37(20)*, 2849–2869.
- Blakeslee, B. and McCourt, M. E. (1999). A multiscale spatial filtering account of the White effect, simultaneous

brightness contrast and grating induction. *Vision Research*, 39, 4361–4377.

Blakeslee, B. and McCourt, M. E. (2001). A multiscale spatial filtering account of the Wertheimer-Benary effect and the corrugated Mondrian. *Vision Research*, 41(19), 2487–2502.

Blakeslee, B. and McCourt, M. E. (2004). A unified theory of brightness contrast and assimilation incorporating oriented multi-scale spatial filtering and contrast normalization, *Vision Research*, 44(21), 2483–2503.

Blakeslee, B. and McCourt, M. E. (2008), Nearly instantaneous brightness induction, *Journal of Vision*, 8(2), 1–8

Blakeslee, B., Pasioka, W., and McCourt, M. E. (2005), Oriented multiscale spatial filtering and contrast normalization: a parsimonious model of brightness induction in a continuum of stimuli including White, Howe and simultaneous brightness contrast. *Vision Research*, 45(5), 607–615.

Blakeslee, B., Reetz, D., and McCourt, M. E. (2008), Coming to terms with lightness and brightness: Effects of stimulus configuration and instructions on brightness and lightness judgments. *Journal of Vision*, 8(11), 1–14.

Bonin, V., Mante, V., and Carandini, M. (2005). The suppressive field of neurons in lateral geniculate nucleus, *The Journal of Neuroscience*, 25(47), 10844–10856.

Brentano, F. (1892). Über ein optisches Paradoxon. *Zeitschrift für Psychologie*, 3, 349–358.

Brown, H., & Friston, K. (2012). Free energy and illusions: the Cornsweet Effect. *Frontiers in Psychology*, 3(43).

Brown, H., Adams, R. A., Parees, I., Edwards, M., & Friston, K.. (2013). Active inference, sensory attenuation and illusions. *Cognitive Processing*, 14(4), 411–427.

Capozzoli, N. J., & Marsh, D. (1994). Schizophrenia and geometric illusions. Report of perceptual distortion. *Schizophrenia Research*, 13(1), 87–89.

Carandini, M. and Heeger, D. J. (2012). Normalization as a canonical neural computation. *Nature Reviews Neuroscience*, 13, 51 – 62.

Carrasco, M., Figueroa, J. G., Willen, J. D. (1986). A test of the spatial-frequency explanation of the Müller-Lyer illusion. *Perception*, 15, 553–562.

Chubb, C., Sperling, G., & Solomon, J. A. (1989). Texture interactions determine perceived contrast. *Proc Natl Acad Sci U. S. A.*, 86(23), 9631–9635.

- Cooper, M. R. & Runyon, R. P. (1970). Error increase and decrease in minimal form of Mueller-Lyer illusion. *Perceptual and Motor Skills*, 31, 535-538.
- Cope, D., Blakeslee, B. & McCourt, M. E. (2013). Analysis of multidimensional difference-of-Gaussians filters in terms of directly observable parameters. *Journal of the Optical Society of America, A*, 30(5), 1002-1012.
- Cope, D., Blakeslee, B., and McCourt, M. E. (2014a). Modeling lateral geniculate nucleus response with contrast gain control. Part 1: formulation. *Journal of the Optical Society of America A*, 30(11), 2401-2408.
- Cope, D., Blakeslee, B., and McCourt, M. E. (2014b). Modeling lateral geniculate nucleus response with contrast gain control. Part 2: analysis. *Journal of the Optical Society of America A*, 31(2), 348-362.
- Coren S., Girgus J.S., & Day, R. H. (1973). Visual Spatial Illusions: Many Explanations. *Science*, 179:4072, 503-504.
- Coren, S. & Porac, C. (1984) Structural and cognitive components in the Müller- Lyer illusion assessed via cyclopean presentation. *Perception and Psychophysics*, 35(4), 313–318.
- Corney, D. & Lotto, R. B. (2007). What Are Lightness Illusions and Why Do We See Them? *PLoS Computational Biology*, 3(9): e180. doi:10.1371/journal.pcbi.0030180.
- Cromwell, R. L. (1975). Assessment of schizophrenia. In: *Annual Review of Psychology*. Palo Alto, CA: Annual Reviews, Inc., 593-619.
- Cromwell, R. L., & Pithers, W. D. (1981). Schizophrenic/paranoid psychoses: determining diagnostic divisions. *Schizophrenia Bulletin*, 7(4), 674-688.
- Dakin, S. C. & Bex, P. J. (2003) Natural Image Statistics Explain Brightness "Filling-In". *Proceedings of the Royal Society of London, Biological Sciences*, 270(1531), 2341-2348.
- Dakin, S., Carlin, P. & Hemsley, D. (2005). Weak suppression of visual context in chronic schizophrenia. *Current Biology*, 15(20), R822–R824.
- DeValois, R. L., & DeValois, K. K. (1988). Spatial vision. New York: Oxford University Press.
- Dima, D., Roiser, J. P., Dietrich, D. E., Bonnemann, C., Lanfermann, H., Emrich, H. M., & Dillo, W. (2009). Understanding why patients with schizophrenia do not perceive the hollow-mask illusion using dynamic causal modelling. *NeuroImage*, 46(4), 1180-1186.
- Dima, D., Dietrich D. E., Dillo W., & Emrich H. M. (2010). Impaired top-down processes in schizophrenia: A DCM study of ERPs. *NeuroImage* 52(3), 824–832.

- Dima, D., Dillo, W., Bonnemann, C., Emrich, H.M., & Dietrich D.E. (2011). Reduced P300 and P600 amplitude in the hollow-mask illusion in patients with schizophrenia. *Psychiatry Research: Neuroimaging*, 191(2), 145–151.
- Engber, D. (2014, January 19). Who made that Captcha? *The New York Times*. Retrieved from: http://www.nytimes.com/2014/01/19/magazine/who-made-that-captcha.html?_r=0
- Fisher, G. H. (1968). Gradients of distortion seen in the context of the Ponzo illusion and other contours. *Quarterly Journal of Experimental Psychology*, 20(2), 212–217, doi: 10.1080/14640746808400153.
- Fogelson, N., Ribolsi, M., Fernandez-Del-Olmo, M., Rubino, I. A., Romeo, D., Koch, G and Peled, A. (2011). Neural correlates of local contextual processing deficits in schizophrenic patients. *Psychophysiology*, 48(9): 1217–1226.
- Friston, K. (2005a) A theory of cortical responses. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 360(1456):815–36.
- Friston, K. J. (2005b) Hallucinations and perceptual inference. *Behavioral and Brain Sciences*, 28(6), 764–766, doi: 10.1017/S0140525X05290131.
- Friston, K. (2008). Hierarchical models in the brain. *PLoS Computational Biology*, 4(11): e1000211. doi:10.1371/journal.pcbi.1000211.
- Friston K. (2010). The free-energy principle: a unified brain theory? *Nat Rev Neuroscience*, 11(2), 127–38.
- Friston K. (2012). A Free Energy Principle for Biological Systems. *Entropy*, 14, 2100–2121. doi:10.3390/e14112100.
- Ghebreab, S., Smeulders, A. W. M., Scholte, H. S., and Lamme, V. A. F. (2009). A Biologically Plausible Model for Rapid Natural Image Identification. In Y. Bengio, D. Schuurmans, J. Lafferty, C. Williams & A. Culotta (Eds.), *Advances in Neural Information Processing Systems 22: Proceedings of the 2009 conference*, 629–637. NIPS Foundation.
- Gilchrist, A. L. (1977). Perceived lightness depends on perceived spatial arrangement, *Science*, 195(4274), 185–187.
- Gilchrist, A., Kossyfidis, C., Bonato, F., Agostini, T., Cataliotti, J., Li, X., Spehar, B., Annan, V. & Economou, E. (1999). An anchoring theory of lightness perception, *Psychological Review*, 106(4), 795–834.
- Ginsburg, A. (1978). *Visual Information Processing Based on Spatial Filters Constrained by Biological Data*.

Ph.D. thesis, Aerospace Medical Research Laboratory, Aerospace Medical Division, Air Force Systems Command.

Ginsburg, A. P. (1984). Visual form perception based on biological filtering. In *Sensory Experience, Adaptation and Perception*, Eds. Spillman, L. and Wooton, B. R., 53-72. New Jersey: Erlbaum.

Goryo, K., Robinson, J. O., & Wilson, J. A. (1984). Selective looking and the Müller-Lyer illusion: the effect of changes in the focus of attention on the Müller-Lyer illusion. *Perception*, 13(6), 647 – 654.

Golz, J. and MacLeod, D. I.A. (2002). Influence of scene statistics on colour constancy. *Nature*, 415, 637-640.

Greist-Bousquet, S. & Schiffman, H. R. (1981). The many illusions of the Müller-Lyer: comparisons of the wings-in and wings-out illusions and manipulations of standard and dot forms. *Perception*, 10(2), 147-154.

Gregory, R. L. (1963). Distortion of visual space as inappropriate constancy scaling. *Nature*, 199, 678–680.

Gregory R. L. (1966). *Eye and Brain: the psychology of seeing*. London: Weidenfeld & Nicolson; 5th edition 1997, Oxford University Press/Princeton University Press.

Gregory, R. L. (1968). Perceptual illusions and brain models. *Proceedings of the Royal Society of London B Biological Sciences*, 171(1024), 279-96.

Gregory, R. L & Heard, P. (1979). Border locking and the café wall illusion. *Perception*, 8(4), 365-380.

Gregory, R. L. (1997). Knowledge in perception and illusion, *Phil. Trans. R. Soc. Lond. B*, 352, 1121–1128

Gregory, R. L. (2005). The Medawar Lecture 2001: Knowledge for Vision: Vision for Knowledge. *Philosophical Transactions: Biological Sciences*, 360(1458), 1231-1251.

Groen, I. I. A., Ghebreab, S., Lamme, V. A. F. & Scholte, H. S. (2012a). Low-level contrast statistics are diagnostic of invariance of natural textures. *Frontiers in Computational Neuroscience*, 6(34). doi: 10.3389/fncom.2012.00034.

Groen, I. I. A., Ghebreab, S., Lamme, V. A. F., & Scholte, H. S. (2012b). Spatially Pooled Contrast Responses Predict Neural and Perceptual Similarity of Naturalistic Image Categories. *PLoS Computational Biology*, 8(10): e1002726. doi:10.1371/journal.pcbi.1002726

Guizzo, E. (2011, October 18). How Google's Self-Driving Car Works. *IEEE Spectrum*. Retrieved from: <http://spectrum.ieee.org/automaton/robotics/artificial-intelligence/how-google-self-driving-car-works>

Happé, F. G. E. (1996). Studying Weak Central Coherence at Low Levels: Children with Autism do not Succumb to Visual Illusions. A Research Note. *Journal of Child Psychology and Psychiatry*, 37(7), 873-877.

- Hamburger, K., & Shapiro, A. G. (2009). Spillmann's weaves are more resilient than Hermann's grid. *Vision Research*, 49, 2121-2130.
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010) The weirdest people in the world? *Behavioral and Brain Sciences*, 33(2-3), 61-83; discussion 83-135. doi: 10.1017/S0140525X0999152X.
- Hermann, L. (1870). Eine Erscheinung simultanen Contrastes. *Pflügers Archiv für die gesamte Physiologie*, 3, 13–15. doi:10.1007/BF01855743.
- Hinton, G. E. & Zemel, R. S. (1994) Autoencoders, minimum description length and Helmholtz free energy. In: *Advances in neural information processing systems* 6, Eds. J. Cowan, G. Tesauro & J. Alspector. Morgan Kaufmann.
- Hodgkin, A. L. & Huxley, A. F. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *The Journal of Physiology*, 117(4), 500–544.
- Hubel, D. H. (1959). Single unit activity in striate cortex of unrestrained cats. *The Journal of Physiology*, 147, 226-238.
- Hubel, D. H., & Wiesel, T. N. (1959). Receptive fields of single neurones in the cat's striate cortex. *Journal of Physiology*, 148, 574– 591.
- Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology*, 160, 106-154.
- Hubel, D. H., & Wiesel, T. N. (1965). Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat. *Journal of Neurophysiology*, 28, 229-289.
- Hohwy, J. (2013a). Delusions, illusions and inference under uncertainty, *Mind & Language*, 28(1), 57-71.
- Hohwy, J. (2013b) The predictive mind. Oxford University Press.
- Howe, C. Q. and Purves, D. (2002). Range image statistics can explain the anomalous perception of length. *Proceedings of the National Academy of Sciences*, 99(20), 13184-13188.
- Howe, C. Q. and Purves, D. (2004). Size Contrast and Assimilation Explained by the Statistics of Natural Scene Geometry. *Journal of Cognitive Neuroscience*, 16(1), 90-102.
- Howe, C.Q. and Purves, D. (2005a). Perceiving Geometry: Geometrical Illusions Explained by Natural Scene Statistics. New York: Springer.

Howe C. Q., & Purves D. (2005b). The Müller-Lyer illusion explained by the statistics of image–source relationships. *Proceedings of the National Academy of Sciences*, 102(4), 1234–1239.

Jordan, K. and Randall J. (1987). The effects of framing ratio and oblique length on Ponzo illusion magnitude. *Perceptual Psychophysics*, 41(5), 435-9.

Kantrowitz, J. T., Butler, P. D., Schecter, I., Silipo, G., and Daniel C. Javitt, D. C. (2009). Seeing the World Dimly: The Impact of Early Visual Deficits on Visual Experience in Schizophrenia. *Schizophrenia Bulletin*, 35(6), 1085–1094, doi:10.1093/schbul/sbp100

Kar, B. C. (1967). *Muller-Lyer illusion in schizophrenics as a function of field distraction and exposure time*. M.A. thesis, George Peabody College for Teachers, Nashville, TN.

Kawabe, T., Qian, K., Yamada, Y., & Miura, K. (2010). The jaggy diamonds illusion. *Perception*, 39, 573-576.

Kersten, D., Mamassian, P., & Yuille, A. (2004). Object perception as Bayesian inference. *Annual Review of Psychology*, 55, 271-304.

Kitaoka, A. (1998). Apparent contraction of edge angles. *Perception*, 27(10), 1209-1219.

Kitaoka, A., Pinna, B., & Brelstaff, G. (2004). Contrast polarities determine the direction of café wall tilts. *Perception*, 33(1), 11-20.

Klette, R. 2014. *Concise computer vision, an introduction into theory and algorithms*. Springer-Verlag London 2014.

Knill, D. C. and Kersten, D. (1991). Apparent surface curvature affects lightness perception. *Nature*, 351(6323), 228-30.

Kuffler, S. W. (1953). Discharge patterns and functional organization of mammalian retina. *J. Neurophysiology*, 16, 37-68.

Kuffler, S.W. (1973). The single-cell approach in the visual system and the study of receptive fields. *Invest Ophthalmol*, 12(11), 794-813.

de Lafuente, V. and Ruiz, O. (2004). The orientation dependence of the Hermann Grid illusion. *Exp Brain Res*, 154, 255–260.

Lee, A. B., Mumford, D., and Huang, J. (2001). Occlusion models for natural images: A statistical study of a scale-invariant dead leaves model. *International Journal of Computer Vision*, 41(1/2), 35–59.

- Lee, T. S. & Mumford, D. (2003) Hierarchical Bayesian inference in the visual cortex. *Journal of Optical Society of America, A*, 20(7), 1434 – 48.
- Letourneau, J. E. (1974). The Oppel-Kundt and the Müller-Lyer illusions among schizophrenics. *Perceptual and motor skills*, 39, 775-778.
- Lotto, R. B., & Purves, D. (2001). An empirical explanation of the Chubb illusion. *Journal of Cognitive Neuroscience*, 13(5), 547-555.
- Masquelier T., & Thorpe, S. J. (2007). Unsupervised learning of visual features through spike timing dependent plasticity. *PLoS Computational Biology*, 3(2): e31. doi: 10.1371/journal.pcbi.0030031
- Müller-Lyer, F. C. (1889). Optische Urteilstäuschungen. *Archiv für Anatomie und Physiologie*, 2, 263–270.
- Müller-Lyer, F. C. (1896). Über Kontrast und Konfluxion. (Zweiter Artikel). *Zeitschrift für Psychologie und Physiologie der Sinnesorgane*, 10, 421–431.
- Mundy, M. E. (2014). Testing day: The effects of processing bias induced by Navon stimuli on the strength of the Müller-Lyer illusion. *Advances in Cognitive Psychology*, 10(1), 9-14. doi: 10.2478/v10053-008-0151-8.
- Must, A., Janka, Z., Benedek, G., and Kéri, S. (2004). Reduced facilitation effect of collinear flankers on contrast detection reveals impaired lateral connectivity in the visual cortex of schizophrenia patients. *Neuroscience Letters*, 357(2), 131–134.
- Mutch, J., & Lowe, D. G. (2008) Object class recognition and localization using sparse features with limited receptive fields. *International Journal of Computer Vision* 80: 45–57.
- Necker, L. A. (1832). Observations on some remarkable optical phaenomena seen in Switzerland; and on an optical phaenomenon which occurs on viewing a figure of a crystal or geometrical solid. *London and Edinburgh Philosophical Magazine and Journal of Science*, 1(5), 329–337.
- Notredame, C-E., Pins, D., Deneve, S., & Jardri, R. (2014). What visual illusions teach us about schizophrenia. *Frontiers in Integrative Neuroscience*, 8(63). doi: 10.3389/fnint.2014.00063
- Olzak, L. A., & Laurinen, P. I. (1999). Multiple gain control processes in contrast - contrast phenomena. *Vision Research*, 39(24), 3983–3987.
- Penrose, L. S., & Penrose, R. (1958). Impossible objects: A special type of visual illusion. *British Journal of Psychology*, 49(1), 31–33.
- Pessoa, V. F., Monge-Fuentes, V., Simon, C. Y., Suganuma, E., Tavares, M. C. (2008). The Müller-Lyer illusion

as a tool for schizophrenia screening. *Reviews in the Neurosciences*, 19(2-3), 91–100.

Ponzo, M. (1911). Intorno ad alcune illusioni nel campo delle sensazioni tattili sull'illusione di Aristotele e fenomeni analoghi. *Arch. Gesamte Psychol*, 16, 307–345.

Predebon, J. (1992). Framing effects and the reversed Müller-Lyer illusion. *Perception & Psychophysics*, 52(3), 307-314.

Predebon, J., Stevens, K., & Petocz, A. (1993). Illusion decrement and transfer of illusion decrement in Müller-Lyer figures. *Perception*, 22(4), 391-401.

Predebon, J. (1994). The reversed Müller-Lyer illusion in conventional and in wing-amputated Müller-Lyer figures. *Psychological Research*, 56(4), 217-223.

Predebon, J. (1998). Decrement of the Brentano Müller-Lyer illusion as a function of inspection time. *Perception*, 27(2), 183 – 192.

Predebon, J. (2004). Selective attention and asymmetry in the Müller-Lyer illusion. *Psychonomic Bulletin and Review*, 11(5), 916-920.

Predebon, J. (2005). A comparison of length-matching and length-fractionation measures of Müller-Lyer distortions. *Perception & Psychophysics*, 67(2), 264-273.

Predebon, J. (2006). Decrement of the Müller-Lyer and Poggendorff illusions: the effects of inspection and practice. *Psychological Research*, 70(5), 384-94.

Prinzmetal, W., and Beck, D. M. (2001). The tilt-constancy theory of illusions. *Journal of Experimental Psychology: Human Perception and Performance*, 27(1), 206-217.

Prinzmetal, W., Shimamura, A. P., & Mikolinski, M. (2001). The Ponzo illusion and the perception of orientation. *Perceptual Psychophysics*, 63(1), 99-114.

Qian, K., Yamada, Y., Kawabe, T., & Miura, K. (2009). The scintillating grid illusion: influence of size, shape, and orientation of the luminance patches. *Perception*, 38(8), 1172-82.

Restle, F., & Decker, J. (1977) Size of the Mueller-Lyer illusion as a function of its dimensions: Theory and data. *Perception and Psychophysics*, 21, 489–503.

Riesenhuber, M., and Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience*, 2, 1019–1025.

- Robinson, A. E., Hammon, P. S., and de Sa, V. R. (2007). Explaining brightness illusions using spatial filtering and local response normalization. *Vision Research*, 47(12), 1631–1644.
- Robol, V., Tibber, M. S., Anderson, E. J., Bobin T., Carlin, P., Shergill, S. S. and Dakin, S. C. (2013) Reduced Crowding and Poor Contour Detection in Schizophrenia Are Consistent with Weak Surround Inhibition. *PLoS ONE* 8(4): e60951. doi:10.1371/journal.pone.0060951
- Ropar, D. & Mitchell, P. (1999). Are individuals with autism and Asperger's syndrome susceptible to visual illusions? *J Child Psychol Psychiatry*, 40(8), 1283-1293.
- Rubin, E. (1915). *Synsoplevede Figurer*. Doctoral thesis, Copenhagen: Gyldendalske.
- Salmela, V. R & Laurinen, P. I. (2009). Low-level features determine brightness in White's and Benary's illusions. *Vision Research*, 49(7), 682-690. doi: 10.1016/j.visres.2009.01.006.
- Schmack, K., Castro, A. G.-C., de, Rothkirch, M., Sekutowicz, M., Rössler, H., Haynes, J.-D., Heinz, A., Petrovic, P., & Sterzer, P. (2013). Delusions and the role of beliefs in perceptual inference. *Journal of Neuroscience*, 33(34), 13701–13712.
- Schmid, A. C. & Anderson, B. L. (2014). Do surface reflectance properties and 3-D mesostructure influence the perception of lightness? *Journal of Vision*, 14(8):24, 1–24.
- Scholte, H. S., Ghebreab, S., Waldorp, L., Smeulders, A. W. M. & Lamme, V. A. F. (2009). Brain responses strongly correlate with Weibull image statistics when processing natural images. *Journal of Vision*, 9(4):29, 1-15, <http://journalofvision.org/9/4/29/>, doi:10.1167/9.4.29.
- Segall, M. H., Campbell, D. T. and Herskovits, M. J. (1966). *The Influence of Culture on Visual Perception*. The Bobbs-Merrill Company, Inc.
- Serre, T., Wolf, L., and Poggio T. (2005a). Object recognition with features inspired by visual cortex. In: *Proceedings of 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005)*. San Diego: IEEE Computer Society Press, 886–893.
- Serre, T., Kouh, M., Cadieu, C., Knoblich, U., Kreiman, G., and Tomaso Poggio. (2005b). A theory of object recognition: computations and circuits in the feedforward path of the ventral stream in primate visual cortex. Technical report, Massachusetts Institute of Technology, Cambridge, MA, December 2005.
- Serre, T., Oliva, A., & Poggio, T. (2007) A feedforward architecture accounts for rapid categorization. *Proceedings of the National Academy of Science*, 104(15), 6424–6429.
- Serre, T., & Poggio, T. (2010). A neuromorphic approach to computer vision. *Communications of the ACM*

(online), 53(10), October 2010.

Serre, T. (2014). Hierarchical Models of the Visual System. In *Encyclopedia of Computational Neuroscience* (pp. 1-12). Springer New York.

Seymour, K., Stein, T., Sanders, L. L. O., Guggenmos, M., Theophil, I., & Sterzer, P. (2013). Altered contextual modulation of primary visual cortex responses in schizophrenia. *Neuropsychopharmacology*, 38(13), 2607-2612.

Silverstein, S. M. & Keane, B. P. (2011). Perceptual Organization Impairment in Schizophrenia and Associated Brain Mechanisms: Review of Research from 2005 to 2010. *Schizophrenia Bulletin*, 37(4), 690-699

Simoncelli, E. P. and Olshausen, B. (2001). Natural image statistics and neural representation. *Annual Review of Neuroscience*, 24(1), 1193-1216.

Srinivasan, Laughlin and Dubs. 1982. Predictive coding: a fresh view of inhibition in the retina.

Thiéry, A. (1896), Über geometrisch-optische Täuschungen, *Philosophische Studien*, 12, 67-126.

Thorpe, S., Delorme, A., & VanRullen, R. (2001). Spike-based strategies for rapid processing. *Neural Networks*, 14(6-7), 715-725.

Todorovic, D. (1997). Lightness and junctions. *Perception*, 26(4), 379-395.

VanRullen, R., Gautrais, J., Delorme, A., & Thorpe, S. (1998). Face processing using one spike per neurone. *BioSystems*, 48(1-3), 229-239.

VanRullen, R. & Thorpe, S. J. (2001a). Rate coding versus temporal order coding: What the retinal ganglion cells tell the visual cortex. *Neural Computation*, 13(6), 1255-1283.

VanRullen, R. & Thorpe, S. J. (2001b). Is it a bird? Is it a plane? Ultra-rapid visual categorisation of natural and artificial objects. *Perception*, 30(6), 655-68.

VanRullen, R. & Thorpe, S. J. (2002). Surfing a spike wave down the ventral stream. *Vision Research*, 42(23), 2593-2615.

VanRullen, R., Guyonneau, R., and Thorpe, S. J. (2005). Spike times make sense. *Trends in Neuroscience*, 28(1).

Weckowicz, T. E., and Witney, G. (1960). The Muller-Lyer illusion in schizophrenic patients. *The British Journal of Psychiatry*, 106, 1002-1007.

White, M. (1979). A new effect of pattern on perceived lightness, *Perception*, 8(4), 413-416.

Whitwam, R. (2014, September 8). How Google's self-driving cars detect and avoid obstacles. *Extreme Tech*. Retrieved from:

<http://www.extremetech.com/extreme/189486-how-googles-self-driving-cars-detect-and-avoid-obstacles>

Woloszyn, M. R. (2010). Contrasting three popular explanations for the Müller-Lyer illusion. *Current Research in Psychology, 1*, 102-107. doi:10.3844/crpsp.2010.102.107

Woodworth, R. S. (1938). *Horizontal-vertical illusion due to actual misperception of depth*. Experimental Psychology. New York: Holt.

Xu, X., Bonds, A. B., and Casagrande, V. A. (2002). Modeling receptive-field structure of koniocellular, magnocellular, and parvocellular LGN cells in the owl monkey (*Aotus trivigatus*). *Visual Neuroscience, 19*(6), 703-711.

Zeman, A., Obst, O., Brooks, K. R., & Rich, A. N. (2013). The Müller-Lyer Illusion in a computational model of biological object recognition. *PLoS ONE, 8*(2), e56126. doi:10.1371/journal.pone.0056126

Zeman, A., Obst, O., & Brooks, K. R. (2014). Complex cells decrease errors for the Müller-Lyer Illusion in a computational model of the visual ventral stream. *Frontiers in Computational Neuroscience, 8*(112). doi:10.3389/fncom.2014.00112

Zeman, A., Brooks, K. R & Ghebreab, S. (in submission). An exponential filter model predicts lightness illusions. *In submission*.

Zenger-Landolt, B., & Heeger, D. J. (2003). Response suppression in V1 agrees with psychophysics of surround masking. *Journal of Neuroscience, 23*(17), 6884-6893.

Zhou, H., Friedman, H. S., & von der Heydt, R. (2000). Coding of border ownership in monkey visual cortex. *Journal of Neuroscience, 20*(17), 6594-611.

Zhu, S. C., and Mumford, D. B. (1997). Learning generic prior models for visual computation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition: June 17 - 19, San Juan, Puerto Rico*, ed. IEEE Computer Society, 463-469. Los Alamitos, CA : IEEE Computer Society.