

Scale-Space and the Implicit Coding of Luminance in V1

Alex Ioannides

University College London

Thesis submitted to the University of London for the degree of Doctor of Philosophy.



UMI Number: U592057

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



UMI U592057

Published by ProQuest LLC 2013. Copyright in the Dissertation held by the Author.
Microform Edition © ProQuest LLC.

All rights reserved. This work is protected against
unauthorized copying under Title 17, United States Code.



ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

I hereby declare that the work presented in this thesis is entirely my own.

Alex Ioannides

16 / 07 / 07

Abstract

This thesis pursues a single line of enquiry: lightness, brightness, and visual illusions. In particular, it focuses on White's effect, simultaneous brightness contrast, and low-level theories that can account for both phenomenon.

In the first part (Chapters 1-2), the problem-space is defined before a review of lightness and brightness theories from both low- and high-level vision.

In the second part (Chapter 3), the only two low-level V1 models of brightness, capable of accounting for both White's effect and simultaneous brightness contrast, are shown to be reliant on the amplification of low spatial frequency information derived for large-scale RFs, to accurately reconstruct images and account for the illusory brightness apparent in both effects. It is argued that these large-scale RFs do not exist in V1, and that the global re-weighting and re-normalisation schemes employed by these models are not constrained by the known local nature of intra-cortical connections. Hence, it was concluded that these models are not biologically plausible.

In the third part (Chapter 4), the issue of recovering low spatial frequency and local mean luminance information without explicitly sampling it, is considered. The problem is formally defined in the Scale-Space framework and solved analytically. That is, an algorithm for recovering local mean-luminance (and low spatial frequencies), from the information implicit in contrast coding cells typically found in V1, is constructed, and is referred to as the Implicit Luminance Coding (ILC) model.

It is argued that the ILC model is not biologically-plausible, by virtue of its global optimisation framework being unconstrained by the known local nature of intra-cortical connections. Subsequently, a new algorithm is proposed, based on a numerical approximation to the analytical solution. The biologically-plausible ILC algorithm is developed into a complete low-level model of brightness, which makes use of the information present in multiple scale channels. The model is shown to be capable of accounting for both White's effect and simultaneous brightness contrast, by means of an interplay between two independent assimilation and contrast mechanisms.

The final part (Chapter 5), is concerned with the application of the model to visual phenomenon synonymous with lightness and brightness, including all known variants of White's effect and simultaneous brightness contrast, and some effects that are traditionally accounted for by appealing to mechanisms from high-level vision, thus facilitating the delineation of low-level from higher-level phenomena. The biologically-plausible ILC model is shown to be in good accordance with this experimental data. Furthermore, qualitative accounts for the temporal evolution of the filling-in process were provided and shown to be in agreement with experiment, and novel predictions as to the temporal evolution of White's effect relative to simultaneous brightness contrast are described.

Acknowledgements

Firstly, I would like to thank my supervisor, Alan Johnston, for the inspiring discussions, invaluable guidance, and for believing in me when others didn't. I also thank Lewis Griffin for the many insights into my work.

Special thanks go to Ali Jeewajee, for being there to share the 'academic experience' for the past five years, and to Alex Lewis for the assistance and motivation in the first year.

Furthermore, I would like to thank my life-partner Bianca Levy, who has put-up with my 'ups and downs' and helped me to look at things positively. It remains to thank my parents who, as always, have been a constant source of encouragement and support, and who engendered my academic curiosity from a young age.

Contents

Abstract	2
Acknowledgements	3
Chapter 1: Lightness and Brightness	7
Luminance, Illuminance, Surface Reflectance, and Intrinsic Images	7
Lightness and Brightness in Complex Scenes	8
Lightness Constancy	10
From Lightness Constancy to Colour Constancy	11
Lightness and Brightness Illusions	12
Chapter 2: Models of Lightness and Brightness	16
Models of Low-Level Vision	19
Retinal Adaptation and the Generic Channels Model	19
Low-Dimensional Models and Colour Constancy	20
Von Kries Algorithms and the Retinex Theory of Colour Vision	23
Receptive Fields and Filter-Bank Models	27
Multi-Scale Filter Banks	30
Local Energy Models	36
Lightness and Neural Networks	38
Models of Higher-Level Vision	41
The Bayesian Approach: Inferring the Most Likely Origin of the Luminance	41
Local Junction Analysis	43
The Anchoring Theory of Lightness	47
Summary	48
Chapter 3: An Analysis of Current Neural Models	52
White's Effect: Test Stimuli and Psychophysical Data	53
The Natural Scene Statistics Model	56
The Log-Gabor Code	58

Results	61
The Oriented Difference of Gaussians Model	66
The DoG Code	66
Results	69
Discussion	73
The Spatial Extent of Information Integration in V1	73
The Coding of Luminance in V1	77
Low Spatial Frequencies and the Fourier Representation of Images in V1	79
Chapter 4: The Implicit Coding of Luminance in V1	82
Scale-Space, Local Jets, and the Known Functionality of V1	82
The Representation of Orientation Information	84
The Local-Jet: From Filter Output to Image	85
Local Jets and the Role of ‘Blob’ Cells in V1	86
Recovering Global Luminance Structure at a Single Scale: An Analytical Solution	88
A Biologically Plausible Implementation of the ILC Model	95
Towards a Complete Model of Luminance Coding in V1	99
Multiple Scale Channels	99
Robustness to Image Noise	102
Determining the Number of Iterations	103
The Spatial Frequency Content of the Biological ILC Model’s Output	104
Modelling Lightness and Brightness Illusions	107
From Model Output to Predicted Brightness	107
The Minimum Discriminable Response	108
SBC and White’s Effect: A Tale of Contrast and Assimilation	109
White’s Effect and the Choice of Spatial Scales	112
Comparing the Biological ILC Model to the ODOG and NSS Models	113
Chapter 5: Results & Discussion	116
Static Lightness and Brightness Illusions	118
Simultaneous Brightness Contrast Effects	118

Lightness Illusions	129
Assimilation Effects	135
Perceptual Filling-In and Edge-Driven Surface Perception	146
The Craik-Cornsweet-O'Brien Effect	147
Grating Induction	150
The Temporal Dynamics of Filling-In and Brightness	152
The Temporal Evolution of SBC and White's Effect	155
A Comparison with Other Models	156
Summary	157
Conclusions	159
References	165

Chapter 1: Lightness and Brightness

Before anything pertaining to lightness and brightness can be discussed it is necessary to define these terms and establish their connection to the physical world.

Luminance, Illuminance, Surface Reflectance, and Intrinsic Images

The 2-D distribution of light-intensity, incident on the retina from a scene, is referred to as the luminance (or retinal image). The physics which governs the nature of luminance is dependent on two factors: the source of light in the scene, and how the surfaces of objects within the scene reflect light towards the observer.

A more rigorous description of the relationship between luminance and environment requires a formal definition of the latter. Consider a simplified model of the environment consisting of a planar world embedded in a 3-D space, together with an observer and a light source. The light source is assumed to be a point at infinity so that illumination incident on the flat world is uniform. Furthermore, all surfaces in this world are assumed to be Lambertian, that is, they scatter light equally in all directions such that the luminance is independent of the observers position. The luminance at point (x, y) in the retinal image, $L(x, y)$, arising from this scene, analogous to that when viewing a painting in an art gallery, is given by,

$$L(x, y) = R(x, y)I(x, y), \quad (1.1)$$

where $R(x, y)$ is the surface reflectance (or reflectance) and is defined as the fraction of the incident illumination, $I(x, y)$, that is reflected towards the observer. It should be noted that for uniform illumination, $I(x, y)$ is a constant for all x and y , but it is instructive to cast both the reflectance and illumination in this way as it motivates the notion that luminance may be regarded as the product of two distinct reflectance and illuminance ‘intrinsic images’ (Barrow & Tenenbaum, 1978).

Although idealised environments such as the planar world represent only a trivial subset of all possible scenes, by abstracting away the 3-D layout of surfaces it is possible to negate the effects of shading and study lightness and brightness in isolation from any confounding factors. It is for this reason that so much attention has been focused on this paradigm, and in particular to what are referred to as ‘Mondrian’ scenes; planar worlds comprising of an array of ‘overlapping’ rectangular surfaces of varying reflectance, so called because of their apparent resemblance to the paintings of Piet Mondrian.

Within such a context, lightness is defined as the perceived reflectance of a surface. It can be considered as representing the visual system’s attempt at extracting the physical reflectance based on the luminance from a scene. Brightness, however, is defined purely as the perceived luminance of a surface or region, irrespective of the underlying physical properties that generated it. Unlike luminance, illuminance, and reflectance, which are all physically measurable quantities, lightness and brightness are intrinsically subjective variables (Adelson, 2000). The two are intimately linked, and the difference between lightness and brightness judgements is subtle. For example, in Mondrian scenes under isotropic illumination the two perceptual dimensions are essentially equivalent.

A greater importance is ascribed to lightness over brightness, and often the latter is seen as preceding the former. This stance is motivated by the wealth of information about the environment which is implicit in the reflectance of a surface, and hence in the perception of this reflectance. Reflectance is physically determined by the amount of electromagnetic energy absorbed - as opposed to reflected - by the electronic structure in the surface layers of a material. Hence, a change in reflectance can be attributed to a change in material - information necessary for object recognition.

Lightness and Brightness in Complex Scenes

Despite the additional complexity, 3-D worlds can also be idealised to facilitate the study of lightness and brightness in ‘real world’ settings. The simplest physical construct considers

objects as ensembles of Lambertian surfaces in the same 3-D space as the observer and a source of uniform illumination. The luminance in this world is then given as a modification to (1.1),

$$L(x, y) = R(x, y)I(x, y)\mathbf{n}(x, y) \cdot \mathbf{s}, \quad (1.2)$$

where $\mathbf{n}(x, y)$ is the unit surface-normal (in the 3-D space) of the surface, and \mathbf{s} is the unit direction of the illumination (Adelson & Pentland, 1996). The luminance from a surface in this 3-D space is not only dependent on reflectance, but also on its orientation relative to the direction of illumination. That is, the luminance may now be decomposed into three intrinsic images: reflectance, illuminance, and shading.

Shading is of no consequence in the flat world where all surfaces are constrained to lie on the same plane, and where a spatial variation in luminance can only be attributed to a change in reflectance. Here, there is a one-to-one relationship between scene configuration and luminance. However, in the real world paradigm there is a many-to-one relationship between possible scene configurations and a single luminance image. A change in luminance at any point can now be attributed to either a change in reflectance or a change in surface orientation.

Accounting for the biological substrates of lightness now becomes intrinsically more involved than any discussion of a simple neural representation of the retinal image (although this may still apply to brightness); additional information about the configuration of the scene has to be inferred or extracted from the luminance. Regardless of this increase in complexity, our brains still perform the task near-perfectly, especially remarkable given that scenes in the natural world are not entirely composed of opaque Lambertian surfaces, but include surfaces with specular components, translucent, and transparent surfaces; for which illumination is not uniform and does not pass unadulterated through the atmosphere, and surfaces are illuminated by the light reflected from other surfaces (ambient illumination),

together with that from well-defined sources.

In this setting, the symmetry between lightness and brightness judgements is broken. Lightness is involved with determining the underlying reflectance of a surface, and hence aims to remain invariant under variations in illuminance and shading, whereas brightness will be an explicit function of these parameters.

Lightness Constancy

Given the fundamental role of lightness information in visual perception, it is crucial that the lightness associated with a surface, is *always* associated with that surface, regardless of any change in the environment; it must be invariant to produce the stable visual world we are accustomed to. For example, a change in illuminant resulting in an altered luminance should not affect the lightness attributed to any surface, despite the difference in information provided to the visual system. Such invariance in lightness is referred to as ‘lightness constancy’.

Although lightness constancy may seem an obvious and simple desire, the means by which the visual system achieves it is not fully understood for even the simplest environments. For example, consider the flat world environment, but without uniform illumination, such that $I(x, y)$ is no longer a constant for all x and y . Reflectance is given from (1.1),

$$R(x, y) = \frac{L(x, y)}{I(x, y)}. \quad (1.3)$$

It is clear that from the luminance alone it is not possible to determine the reflectance. Put another way, given the reflectance and illuminant intrinsic images it is possible to calculate the luminance, but it is impossible to calculate (or ‘unmultiply’) the precise reflectance and illuminance images that gave rise to the luminance; once again there is a many-to-one relationship between scene configuration and luminance as soon as any realism is intro-

duced into our models of the environment. By inspection of (1.3) it can be seen that this problem may be tackled by either determining or inferring the illuminance image from the information implicit in the luminance, or by discounting its effects altogether; both kinds of approach will be discussed later.

One important issue which needs to be highlighted, is whether exact reflectance values need to be recovered, or if an alternative representation that maintains all proportionalities between the reflectances of different surfaces, is adequate for assigning lightness. As an example, consider the original flat world environment with uniform illumination, $I(x, y) = C$, where C is an arbitrary constant. Reflectance is now given by,

$$R(x, y) = \frac{L(x, y)}{C}. \quad (1.4)$$

Although reflectance can be analytically recovered with knowledge of the illuminant, it is now possible to assign lightness to a surface through its scaling of $1/C$ relative to other surfaces. This example illustrates what may be referred to as a ‘representation’ of the visual world. In this context the visual system is not reverse-engineering the physics of a scene, but instead attributes lightness by defining its own reflectance metric, which is also invariant under transformation of scene configuration.

From Lightness Constancy to Colour Constancy

In the formalism developed above, reflectance (and hence lightness) is not dependent on the wavelength of the incoming light. Reflectance has corresponded to the fraction of illumination in the entire visible region of the electromagnetic spectrum, that is reflected by a surface. It is possible to extend this formalism to deal with the perception of colour and colour constancy, by incorporating a dependence on wavelength, λ , into the formalism described above, such that (1.1) becomes,

$$L(x, y, \lambda) = R(x, y, \lambda)I(x, y, \lambda). \quad (1.5)$$

There now exists a luminance, reflectance, and illuminance image for every wavelength. In this context, colour perception may be viewed as a generalisation of lightness perception; hence, any successful theory of colour constancy is a successful theory of lightness constancy, by default. This can be made explicit by showing that (1.1) and (1.3) can be recovered via (1.5) through integration,

$$\begin{aligned} L(x, y) &= \int d\lambda L(x, y, \lambda) \\ &= \int d\lambda R(x, y, \lambda)I(x, y, \lambda), \end{aligned} \quad (1.6)$$

$$R(x, y) = \int d\lambda \frac{L(x, y, \lambda)}{I(x, y, \lambda)}, \quad (1.7)$$

where limits are determined by the range of the electromagnetic spectrum that pigment in human photo-receptors cells can respond to. It should be noted, that the inverse operation is not possible; the relationship between lightness and colour is not symmetric. However, concepts from the relatively simpler study of lightness may nevertheless facilitate a greater understanding of colour, and this alone is motivation enough to justify its further investigation.

Lightness and Brightness Illusions

However perfect we assume our ability to remain lightness constant to be, there is an upper limit to performance. In particular, failures in lightness constancy could be manifest in certain visual illusions. The nature of these inadequacies in visual perception betray the underlying mechanisms through which lightness and brightness are realised by the brain, and thus warrant study.

Particularly important illusions in the study of lightness and brightness are simultaneous brightness contrast, Whites effect, and Adelson's 'wall of blocks' demonstration. In simultaneous brightness contrast (Figure 1.1, left), the grey patch on the right appears darker than the one on the left, although in reality both patches are equi-luminant and only differ in the luminance of their surrounds (Helmholtz, 1887; Hering, 1964; Adelson, 2000). In White's effect (Figure 1.1, right), the nature of the simultaneous brightness contrast effect is reversed through the introduction of a grating pattern in the stimulus. The left patch is perceived as being darker than the right patch, despite the local luminance being darker and not lighter as in simultaneous brightness contrast (White, 1979; Blakeslee & McCourt, 1999). In Adelson's 'wall of blocks' (Figure 1.2) the same simultaneous brightness contrast effect is observed between the grey diamonds in both variants. However, the strength of the effect for the stimulus on the left is greatly reduced when the configuration of some of the edges are altered with respect to the stimulus on the right, despite leaving the local luminance relations of each grey diamond unchanged (Adelson, 1993).

These effects are thought by some to result from the brain attributing different interpretations as to the origins of the luminance for the different surfaces under scrutiny, and hence the reflectance of these surfaces. That is, these effects are the direct result of the brain trying to remain lightness constant. Exactly how this may be facilitated remains unknown, and a contentious issue. Alternatively, there is a school of thought which suggests that these phenomena arise simply from the encoding of luminance information in the early (or low-level) visual areas of the brain, and make no reference to the involvement of any higher-level processing. In this context, the illusions are a direct consequence of the biological mechanisms which facilitate brightness judgements. This work will primarily focus on this latter philosophy.

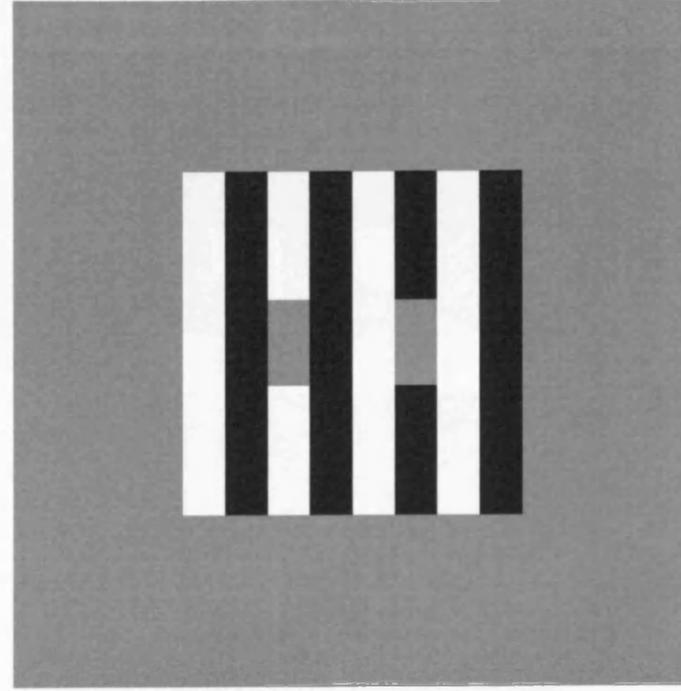
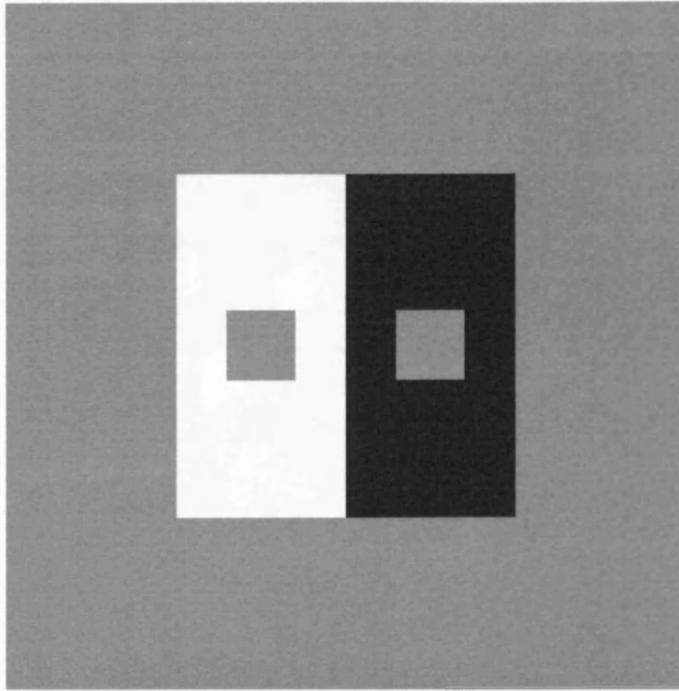


Figure 1.1: An example of the simultaneous brightness contrast effect (left), and White's effect (right). In simultaneous brightness contrast the grey square on the white background appears darker than it's counterpart on the black background. In White's effect, the grey patch on the left (surrounded by more black than white), also appears darker than it's counterpart on the right (surrounded by more white than black). For both of the illusions, the strength of the effects depend on spatial scale: White's effect should appear stronger as the grating increases in spatial frequency (as the reader moves the page away from them) (White, 1981); and simultaneous brightness contrast will appear stronger when the grey test areas decrease in size relative to their surrounds (Yund & Armington, 1975).

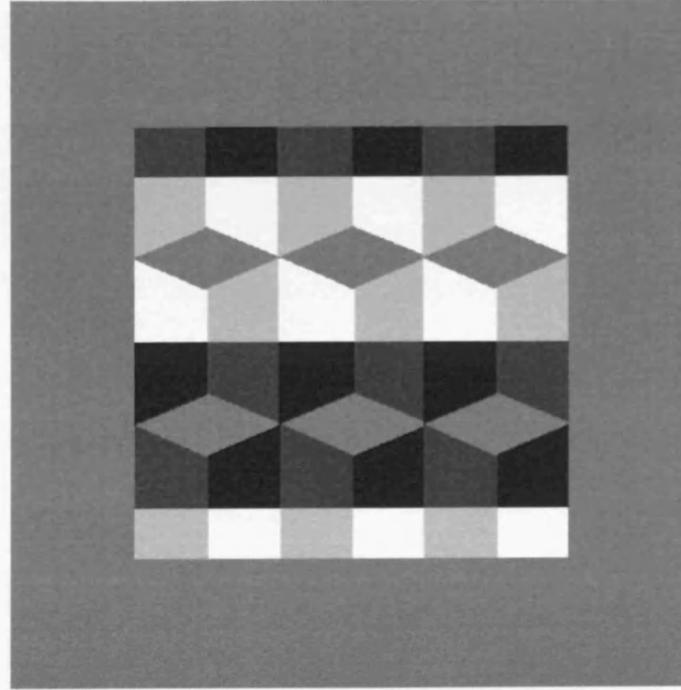
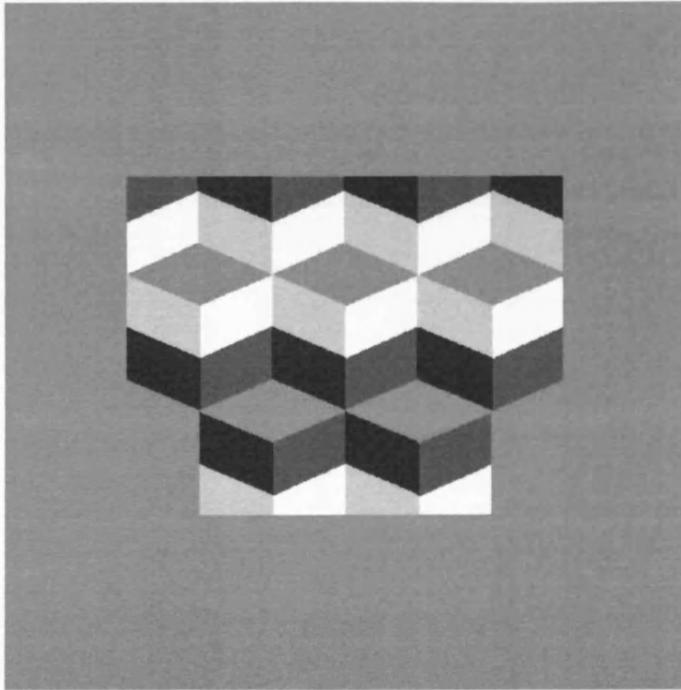


Figure 1.2: Adelson's 'Wall of Blocks' demonstration. The grey diamonds in both pictures are of the same luminance as the background and are framed by surrounds of either high or low luminance, the areas of which are equal between stimuli. Yet, the magnitude of the simultaneous brightness contrast effect on the right ought to appear far stronger than that on the left.

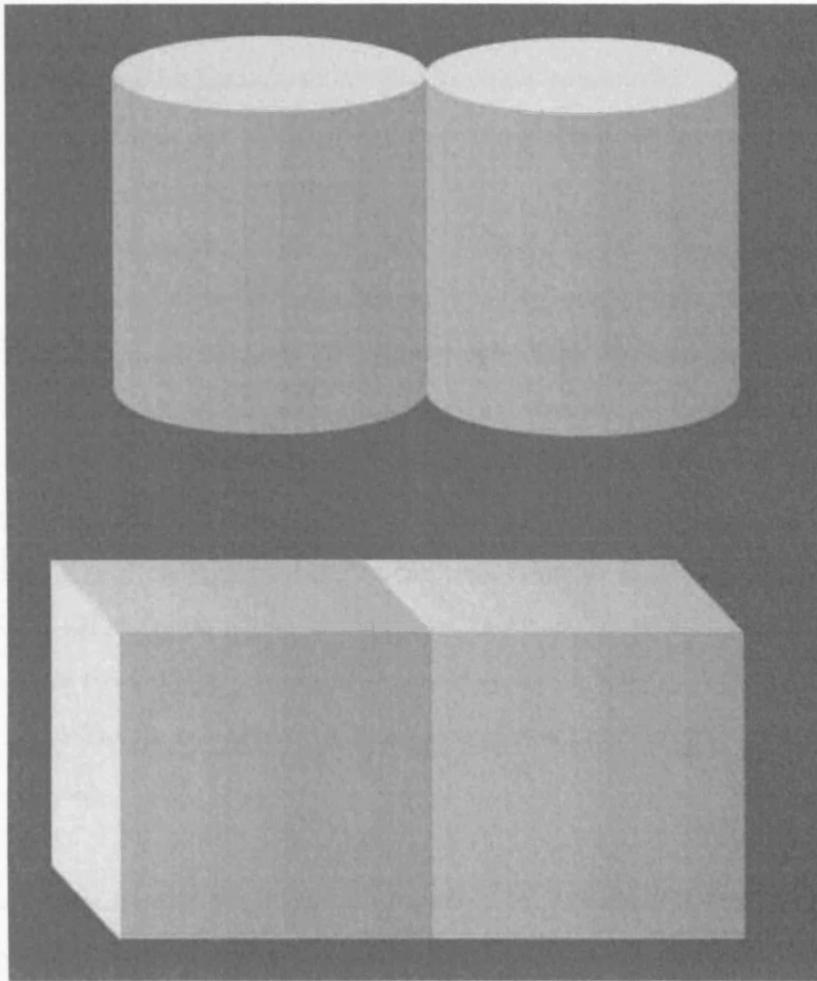


Figure 2.1: Knill and Kersten's variation on the Craik-Cornsweet-O'Brien effect.

Chapter 2: Models of Lightness and Brightness

There exists no explicit and formal definition of what constitutes low- and high-level vision. Low-level vision has come to be regarded as an umbrella term for the relatively simpler functions of the visual system, such as light adaptation and the identification of lines and edges, which lead to a direct neural representation of fundamental image structure. These mechanisms are widely thought to be confined to the retina, lateral geniculate nucleus, and primary visual cortex (V1), and are collectively referred to as the 'primary visual stream'. Such low-level processes do not involve cognition; the brain responds to low-level stimuli even when anaesthetised (for example, see Zeki (1983); Brewer, Press, Logothetis,

and Wandell (2002)). In contrast to this rather parsimonious definition of low-level vision, higher-level vision may be neatly summarised as ‘what comes next’. It is concerned with more complex tasks, such as the identification of objects and the interpretation of scenes, and as such is likely to involve cognition.

As alluded to at the end of the previous chapter, it is the belief of the author that a large set of lightness and brightness illusions are nothing more than artefacts derived from the encoding of luminance information in the primary visual stream, and the subsequent decoding (or image reconstruction). That is to say, it is believed that a lot of lightness and brightness illusions are derived from the type of low-level visual processes described above. However, it is *not* believed that *all* such phenomena can be attributed to the primary visual stream alone; it will be necessary for higher-level process, more concerned with inferring the origins of the luminance from the configuration of the scene as a whole, to enter into accounts for some illusions. What follows below is a discussion of when high-level processing is thought to become relevant.

A variation on the Craik-Cornsweet-O’Brien effect by Knill and Kersten (1991) gives an excellent demonstration of where the delineation between low- and high-level effects are hypothesised to occur, and is depicted in Figure 2.1 (for a more complete discussion of the effect refer to Chapter 6). The faces of the joined cubes at the bottom of the figure have both been generated with the same luminance ramp, increasing linearly in luminance from right to left. At their juncture there is a discontinuity in luminance as one ramp ends and the other begins. As a consequence of this border, the cube on the left appears darker than the one on the right, despite the fact that their luminance profiles are identical. It is widely held view that the illusion is manifest as a direct result of the edge between the luminance ramps, and thus belongs in the domain of low-level vision (Cornsweet, 1970). However, at the top of Figure 2.1 the effect is largely nulled after the cubes have been changed into cylinders through alteration of the lines at the top and bottom, regardless of the fact that the luminance profiles are identical to that of the cubes below. In the latter case, the luminance ramps are likely to be interpreted as the shading from the curved surfaces of the

cylinders, and as such the reflectance of the underlying surfaces, and hence the lightness, are assigned to be the same. In the case of the cubes (and any 1-D or 2-D version of this effect), there are no such cues in the scene and the effect remains.

This process is similar in a similar vein to how the latent simultaneous brightness contrast effect present in Adelson's 'Wall of Blocks' demonstration (Figure 1.2), is modified after a change in the configuration of the geometry in the scene, but not the underlying luminance values. For the stimulus on the right, the grey diamonds are interpreted as existing beneath light or dark transparency media. But, as they both have the 'same' luminance the diamonds under the darker transparency must have a higher surface reflectance, and hence appear brighter as the visual system attempts to remain lightness constant. On the other hand, for the stimulus on the left there are no such higher-level cues to alter the interpretation of the scene and subsequent lightness computation, and hence only the residual SBC effect is observed between the two sets of grey diamonds (Adelson, 1993).

Both of the above demonstrations represent examples of how higher-level scene interpretation, and inference of the origin of the luminance, can have an overriding affect on visual cognition. The latter part of this chapter is concerned with reviewing some of the more prominent higher-level theories of lightness, some of which take the rather extreme view that *all* such phenomenon are the direct result of the kind of higher-level processes described above (Helmholtz, 1887; Purves, Shimpf, & Beau-Lotto, 1999). The view held here, is that all phenomena which lack any cues that are relevant to inferring the 3-D configuration of the scene and the origin of the luminance, are all low-level effects, and that any modification of such effects (such as those discussed above) are the direct consequence of inference processing downstream, acting on the low-level brightness outputs from V1. The following part of this chapter, and indeed the underlying theme of this work, will be concerned with low-level theories and the extent to which they are capable of explaining lightness and brightness phenomena.

Models of Low-Level Vision

Low-level theories of lightness and brightness may be collected into two groups; those for which retinal mechanisms of light adaptation are the primary mechanism, and those which model local interactions (more specifically, the lateral inhibition between photoreceptor cells) in the retina through the receptive field properties of retinal ganglion cells, and their subsequent input to V1. What follows is a brief account of both approaches.

Retinal Adaptation and the Generic Channels Model

Adaptation in the retina refers to the modification of the response mechanism, such that the retina is always maximally sensitive to small variations in luminance relative to the mean level. This acts as an efficient method for extending the overall operating range of the visual system, without loss of precision (Goldstein, 2003). The physiology of the retina imposes a lower-bound on the time taken to make a transition between any two adaptational states, and this latency is apparent when one directly views a bright light and suddenly turns away to a darker region, resulting in a transient ‘after image’ of the light in the visual field, revealing the adaptational process at work. That adaptational effects in the retina may influence the perception of colour and lightness is an idea first proposed by von Kries (1902, 1905), which has since been developed into complete theories of colour and lightness perception.

More specifically, adaptation arises through a modification of the original photoreceptor cell (PRC) responses by subsequent layers of neural circuitry in the retina, the net result of which are the responses of retinal ganglion cells (RGCs) (Hecht, 1920; Shapley & Enroth-Cugell, 1984; Brown & Masland, 2001). At any one point on the retina in normal photopic viewing conditions, the mapping between the responses of the three classes of cone-PRC and the corresponding RGCs, is widely considered as being well-approximated by a series of successive linear operations: an initial multiplicative scaling, followed by an additive shift, a recombination into opponent colour signals, and an attenuation. Collectively this is referred to as the ‘generic channels’ framework (Maloney, 1999). This linear transformation from

PRC to RGC responses may be formally expressed as,

$$\mathbf{o}^{xy} = \mathbf{AHS}(\mathbf{t} + \mathbf{p}^{xy}), \quad (2.1)$$

where $\mathbf{p}^{xy} = [p_l^{xy}, p_m^{xy}, p_s^{xy}]$ are the responses from the L-, M-, and S-cones at point (x, y) on the retina, \mathbf{t} is the additive shift, $\mathbf{S} = \text{diag}[s_1, s_2, s_3]$ and $\mathbf{A} = \text{diag}[a_1, a_2, a_3]$ are the 3×3 diagonal matrices corresponding to the multiplicative scaling and attenuation, \mathbf{H} is the 3×3 opponent transformation matrix, and \mathbf{o}^{xy} are the RGC signals. Assuming that \mathbf{H} can be determined by arguments for the efficient coding of information in the optic nerve (Buchsbaum & Gottschalk, 1983; Zhaoping, 2002), then each adaptational state in the model is determined by nine parameters. The salient point here, is that the responses of RGCs are a linear function of the PRC responses, such that (2.1) can be generalised to,

$$\mathbf{o}^{xy} = f(\mathbf{p}^{xy}), \quad (2.2)$$

where f is a linear function. Hence, any abstract model of colour or lightness constancy that produces a representation of intrinsic colours or lightness, based on some linear transformation of the PRC responses, can be reconciled with the generic channels model, and understood through the architecture of the retina.

Low-Dimensional Models and Colour Constancy

Physics-based approaches to understanding colour constancy have focused on constructing low-dimensional linear models of illuminants and reflectances in the natural environment, and then estimating the parameters required to recover reflectance, given the information present in cone-PRC responses. What follows is a summary of the low-dimensional models framework described by Maloney (1999).

Within this framework, the physical response of a PRC of class k at a point (x, y) on the

retina, p_k^{xy} , is given by,

$$p_k^{xy} = \int d\lambda S_k(\lambda) R^{xy}(\lambda) I^{xy}(\lambda), \quad (2.3)$$

where $k = 1, 2, 3$, and $S_k(\lambda)$ defines the relative PRC response to an illumination as a function of wavelength λ , and are functions that have been empirically determined (Stockman & Sharpe, 2000). Mathematically, it is possible to regard the reflectance as a vector in an infinite-dimensional vector space (or function space), where the inner-product between any two such vectors, $\mathbf{a} = A(\lambda)$ and $\mathbf{b} = B(\lambda)$ is defined as,

$$\mathbf{a} \cdot \mathbf{b} = \int_{-\infty}^{\infty} d\lambda A(\lambda) B(\lambda).$$

Any vector in this space may be expressed in terms of a set of basis functions, $b_i(\lambda)$,

$$B(\lambda) = \sum_{i=1}^{\infty} \beta_i b_i(\lambda), \quad (2.4)$$

where the β_i are the projections of $B(\lambda)$ onto $b_i(\lambda)$, given by $\mathbf{B} \cdot \mathbf{b}_i$. Expressing reflectance in this form and substituting into (2.3) gives,

$$p_k^{xy} = \sum_{p=1}^{\infty} q_p^{xy} \int d\lambda S_k(\lambda) r_p^{xy}(\lambda) I^{xy}(\lambda). \quad (2.5)$$

where the q_p^{xy} are the projections of a surface's reflectance onto the respective basis functions $r^{xy}(\lambda)$, of that space, and $p = 1, 2, 3, \dots, N$. This may be rewritten as the matrix equation,

$$\mathbf{p}^{xy} = \mathbf{L} \mathbf{q}^{xy}, \quad (2.6)$$

where $L_{kp} = \int d\lambda S_k(\lambda)r_p^{xy}(\lambda)I^{xy}(\lambda)$. Thus, if \mathbf{L}^{-1} exists, then a surface's reflectance, characterised by \mathbf{q}^{xy} , may be recovered by,

$$\mathbf{q}^{xy} = \mathbf{L}^{-1}\mathbf{p}^{xy}. \quad (2.7)$$

The successful calculation of \mathbf{L}^{-1} from the available information in the luminance is the principle concern of this approach. However, before \mathbf{L} is inverted it has to be determined. Firstly, this requires the successful calculation of basis functions, which is possible by one of two methods: principle component analysis, or linear regression such as least-squares. The crucial point is that regardless of the method used, the basis functions will be dependent on the particular data set that the model is being fitted to, prompting the question of precisely how to define the 'natural environment' in the context of surface reflectances. Secondly, as there are only three classes of PRC, $\mathbf{p}^{xy} = [p_l^{xy}, p_m^{xy}, p_s^{xy}]$, L must be a 3×3 dimensional matrix for (2.7) to exist. Thus, the infinite series expansion of the reflectance has to be truncated to the three basis functions which capture the greatest variance in the data. Knowledge of these basis functions is assumed *a priori*. Finally, the illuminant has either to be determined or inferred from the luminance. Although many algorithms exist for determining the illuminant in any given scene, only the 'reference surface' and 'grey world' methods will be discussed here, and the reader is directed to Maloney (1999) for a comprehensive discussion.

The reference surface algorithm relies on *a priori* knowledge of the reflectance of a given (reference) surface in the scene, so that when combined with the information in PRC responses to this surface, the illuminant, and hence \mathbf{q}^{xy} , can be calculated via (2.7). Choice of reference surface is arbitrary, but may be pegged to the brightest in the scene for plausible biological implementation¹. The grey world algorithm builds on the reference surface algorithm by making the assumption that the reflectance of the reference surface is that of the mean reflectance in natural scenes, which is also assumed *a priori*. It was originally thought that this reference surface was grey (from which the name of the algorithm was

¹as this surface would be easy to determine, and it may be assumed that the brightest surface in the scene is 'white', and hence reflects all wavelengths equally.

derived), but this is not necessarily so. This approach has the advantage of not requiring a reference surface to be found in the scene, but will estimate the illuminant incorrectly if the reflectance composition in the scene is such that the mean reflectance is a significant distance away from that of the natural environment. Once again, this raises the question of how to define what is a natural environment.

It should be noted that (2.7), or any linear function of (2.7), is of equivalent form to (2.2), and hence it is possible to conceive that if the *a priori* knowledge assumed by the algorithms described above is encoded into the adaptation control mechanisms of the retina, then the generic channels model, or any subsection of, may represent a linear model for calculating reflectance.

Von Kries Algorithms and the Retinex Theory of Colour Vision

It was originally suggested by von Kries (1902, 1905), that adaptation acts through a scaling of the sensation caused by a light. Translated into the generic channels framework, this is equivalent to a scaling of the PRC responses,

$$\mathbf{o}^{xy} = \mathbf{S}\mathbf{p}^{xy}, \quad (2.8)$$

where $\mathbf{S} = \text{diag}[s_1, s_2, s_3]$, in alignment with the multiplicative scaling phase, and determination of the parameters, s_i , requires a theory for the adaptational control mechanism, such as the “Retinex theory” of Land and McCann (1971). Retinex is a theory of colour and lightness perception in the retina or cortex, originally developed to recover the intrinsic colour of surfaces in Mondrian scenes, independent of the illuminant. That is, it exhibits colour and lightness constancy.

The central postulate of Retinex is that each cone class corresponds to a separate ‘retinal system’, which independently calculates the lightness of surfaces based on the reflectance information in the part of the spectrum that each PRC is responsive to. Thus, Retinex

can be thought of as acting on three luminance images, each derived from a specific PRC class. The three lightness values from the three retinal systems then define a point in a 3-D response space, which defines the colour of a surface. This is an example of a representation of sensory information; reflectance is not calculated explicitly.

The theory works on the principle that between any two adjacent PRCs, the effect of any spatially dependent or constant illuminant may be factored out by taking the ratio of the luminance at those points. For example, consider the luminance between two adjacent points, $L_1 = r_1 I$ and $L_2 = r_2 I$, infinitesimally separated such that the effects of a spatially varying illuminant are negligible over this distance. The ratio of the luminances,

$$\frac{L_1}{L_2} = \frac{r_1}{r_2},$$

is by default independent of the illuminant. Similarly, the ratio of luminances between two distant points can be calculated by the successive multiplication of all ratios that exist between the first point and the last point. For example, consider the luminance between three adjacent points, $L_1 = r_1 I$, $L_2 = r_2 I$, and $L_3 = r_3 I$. The ratio between the first and the third is given by,

$$\frac{L_1}{L_3} = \frac{L_1}{L_2} \times \frac{L_2}{L_3} = \frac{r_1}{r_3}.$$

This may also be implemented by using the logarithm of the luminance ratio, such that,

$$\log \frac{L_1}{L_3} = (\log L_1 - \log L_2) + (\log L_2 - \log L_3), \quad (2.9)$$

which is easier to engineer with a biological network of cells (through addition/subtraction of responses undergoing saturation). Thus, if all ratios are calculated with respect to some reference surface (for example, the ‘white’ in any image), then only changes in reflectance,

or edges, will result in a change in ratios, and these values can be used to assign lightness to a surface. This demonstrates clearly demonstrates the importance of edges, a subject which will be discussed further at a later stage.

The array of PRCs in the retina do not sample the luminance continuously, so the PRCs may be thought of as producing pixel values in a luminance, or log-luminance image. In this context, Horn (1974) showed that the Retinex algorithm is treating the image as a 1-D structure, and that the log-luminance ratio between adjacent points is equivalent to the output of a 1-D derivative operator on the log-luminance image, at that point. Thus, the log-luminance ratios between any two non-adjacent points can be recovered by integrating the derivative of the log-luminance image between the two points. However, the assumption that the effect of a spatially dependent illuminant is negligible between two infinitesimally separated points has to be dropped as a consequence of discrete sampling. Between two neighbouring pixels the value of the illuminant will change, and with it the value of the corresponding ratio, even when no change in reflectance is present. This can be circumvented by applying a threshold operation (which is an inherently non-linear process) to each log-luminance ratio, setting all minor deviations from 0 back to 0, in order to discriminate between changes due to illumination and those due to changes in reflectance². However, the threshold has to be chosen carefully: too low, and the effects of the illuminant will not be cancelled; too high, and some low-contrast edges may be erroneously cancelled together with the illuminant.

The problem of choosing a reference surface was addressed by Land and McCann (1971), and biological networks for performing this function were proposed. However, their approach was later dropped for being computationally convoluted, and Land (1983) proposed a simpler version of Retinex that does not depend on locating a reference surface. Instead, all ratios are taken relative to the average reflectance in the scene. This is achieved by hypothesising that at some rate any given cell in the retina will emit a 'signal' in the direction of all adjacent cells, analogous to emission by a point source. The signal will comprise the

²or minor deviations from 1 back to 1, in the case of normal ratios.

logarithm of the output from the cell, and when it reaches the next cell, will have the output from that cell subtracted and thresholded, before that value, or signal, carries on through the rest of the cells in the retina. The result of each spontaneous cell emission, is that the log-ratio of every cell response relative to the emitting cell (referred to as the designator), is calculated at some stage. The final designator is the average designator value for many emissions. It is possible to consider this process occurring in three ‘layers’: one consisting of the photoreceptor cells which sample the retinal image and have a logarithmic output; one which stores the running differences between adjacent cells (the log-ratios), or the signal at each point; and one which calculates the average from many such emissions.

The version of Retinex was analysed by Brainard and Wandell (1986), who showed that if the number and duration of retinal signals are large enough, then the lightness values, l_k^{xy} , of the k -th retinal system, are given by,

$$l_k^{xy} = \log \frac{p_k^{xy}}{G_k}, \quad (2.10)$$

where G_k corresponds to the geometric mean of receptor responses for the k -th receptor class, and which is in accordance with (2.8). It was also shown that when this version of Retinex is applied to Mondrian scenes, colour constancy was achieved for a change in illumination (up to the same thresholds as human colour constancy), but failed under a change in environment. That is, when the colour of a few surfaces in the Mondrian were altered, so were the colour values attributed to the unaltered surfaces. This is not surprising, given that a change in constituent colours will cause a change in the value of G_k , and hence the lightness values attributed by (2.10). Hence, Retinex alone is not a sufficient theory of colour constancy. However, as a theory of lightness its capabilities need not be limited by this failure. Although lightness values may depend on the surfaces in the environment, so long as they are assigned correctly in relation to one another, then an accurate representation of lightness can still be achieved. That is, the *absolute* lightness value is not important, only the ‘distances’ to the lightness of other surfaces. However, it should be noted that

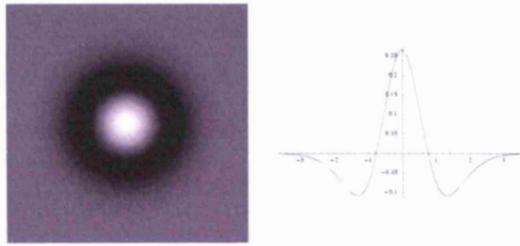


Figure 2.2: An example of a rotationally symmetric DoG RF (left) and its cross-section (right). The brighter regions respond maximally to higher luminance, while the darker regions have a suppressive response to higher luminance.

Retinex has not been shown to be capable of offering an explanation for lightness illusions such as simultaneous brightness contrast or White's effect.

In the mathematical analysis of Retinex as a theory for determining lightness (and not colour), Horn (1974) also extended the theory to work with 2-D images represented as 2-D images (unlike in Retinex). Rather than taking the first derivative as in the 1-D case, the Laplacian in 2-D was chosen as the simplest directionally invariant linear differential operator (unlike the directional derivative). In order to produce the differential image this approach convolves, or 'filters', the image with the Laplacian operator, and this naturally leads to the discussion of filter-based theories of lightness and brightness.

Receptive Fields and Filter-Bank Models

Modelling the local interactions of the retina through the receptive field (RF) properties of RGCs avoids requisite knowledge of the underlying neural circuits and their description in the generic channels framework. This facilitates an exploration of luminance information processing downstream from the retina through to V1. The RF structure of RGCs has been empirically mapped (for example, see Hubel and Wiesel (1977) or Hubel (1988)), and can be modelled, for example, by the rotationally symmetric Laplacian of Gaussian (LoG), or difference of Gaussian (DoG) functions (illustrated in Figure 2.2) (Rodieck, 1965; Marr, 1982).

The response of the RGC corresponding to point (x, y) on the retina is then given by the

weighted sum of all points on the retinal image, where the weighting is specified by the corresponding value of the RF centred on (x, y) . This may be formally expressed as the convolution of the RF with the retinal image,

$$L(x, y) \otimes RF(x, y) = \int \int d\xi d\tau \quad RF(\xi, \tau)L(x - \xi, y - \tau), \quad (2.11)$$

where integration is over the domain of the retinal image, and where the RF is considered a linear filter in this operation. It should be noted that the integrals will turn to sums as the retinal image is sampled discretely by the PRC mosaics. The form of a LoG or DoG RF, is such that it will give a positive weighting to a circular region about its centre and model lateral inhibitions through negatively weighting a surrounding annular region. This configuration is referred to as ‘centre-ON surround-OFF’, exists for both the achromatic and opponent colour channels, and is complemented by the existence of corresponding ‘centre-OFF surround-ON’ channels (Rodieck, 1998; Masland, 2001).

Within the achromatic channel alone, RGCs are found with RFs differing greatly in spatial extent, with mean RFs sizes increasing with distance away from the fovea (Bisti, Clement, Maffei, & Mecacci, 1977; Hubel, 1988). Each spatial configuration may be thought of as being optimally tuned to a specific spatial frequency (SF) in Fourier-space. It is often assumed that for each SF tuning there exists an ensemble of RFs, all tuned to the same SF, and arranged into a regular array (or ‘filter bank’), which covers the retinal image. If the retinal image is successively sampled in this manner, then each filter bank constitutes a distinct SF channel. Under this view, input from the retina to V1 (via the lateral geniculate nucleus³), and any subsequent processing, may be thought of in terms of these separate SF signals. Models based on filtering the retinal image in this way, although not explicit biological models of the primary visual stream, adequately represent the basic response properties of the underlying neurons and provide insight to psychophysical data. This alone is motivation to discuss them in the context of lightness and brightness.

³which shares the same RF properties (Hubel, 1988).

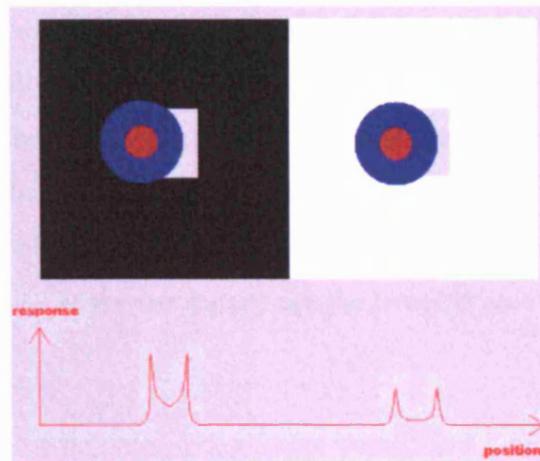


Figure 2.3: An example of how the RF properties of RGC and LGN cells can be used to assign the correct brightness to that seen in the simultaneous brightness contrast configuration. The centre-ON (red disc) surround-OFF (blue disc) RFs are depicted at the moment when the centre-ON units are entirely over the grey test patches. The plot below gives an indication of what the response from these cells would be like as they traversed the cross section of the stimulus.

Although many models based on filter banks have been constructed, even the simplest application of this paradigm has the power to explain lightness and brightness phenomena that many theories of light-adaptation cannot. Perhaps of greatest importance is the explanation that this approach can afford simultaneous brightness contrast (Figure 1.1, left). This effect can be understood by considering the response of a single bank of centre-ON surround-OFF RGCs, to a stimulus with a test patch just smaller than an individual RF, as depicted in Figure 2.3. The response of all cells with RFs straddling the border between background and test patch (with the centre-ON unit entirely over the test patch), will be greater for the darker background as it will elicit less inhibition than its brighter counterpart. Hence, if the responses of these cells are used to assign lightness and brightness then they correctly predict that the test patch on the darker background will appear brighter (Cornsweet, 1970; Goldstein, 2003).

An interesting point is raised when the test patch size relative to that of the RF is increased, to the extent that there will be RGCs responding solely to the grey test patches (and whose

response will be zero if modelled with LoG or DoG RFs with equally balanced excitatory and suppressive regions). Within the filter bank, only the cells at the edges of the test patch will exhibit the behaviour which accounts for simultaneous brightness contrast. Yet, the effect still persists. If the filter bank paradigm is to be successful, then this implies that the lightness or brightness attributed to the patch as a whole must have been determined by the responses at the edges of the test patch, and the bounded area ‘filled-in’ by some other mechanism.

It is important to note that modelling the response of the primary visual stream in this manner does not explicitly address the problem of recovering surface reflectance. The vast majority of filter-based approaches to low-level vision are simple models of the underlying physiology, whose output is more comparable to brightness rather than lightness judgements. This is not to say that such approaches are not capable of computing surface reflectance - reflectance calculations are performed on the output from neurons responsible for sampling the retinal image via their RF structure - just that they are rarely cast within this specific computational context. Regardless, one interesting product of this, is that it allows for a classification of visual phenomena based on the mechanism(s) responsible for their existence: those that are generated from the basic properties of neurons, and those that are a direct consequence of reflectance- and lightness constancy-based calculations. What follows is a brief discussion of various filter-based models and their ability to provide accounts of various lightness and brightness illusions.

Multi-Scale Filter Banks

The central theme for this class of model is the integration of SF information for use in the assignment of lightness or brightness. While a single filter bank is capable of processing a stimulus comprised of information on the same scale, real world scenes contain objects of various sizes, and hence information at all spatial scales, and any ‘realistic’ model of lightness or brightness perception must unify this information into a single interpretation.

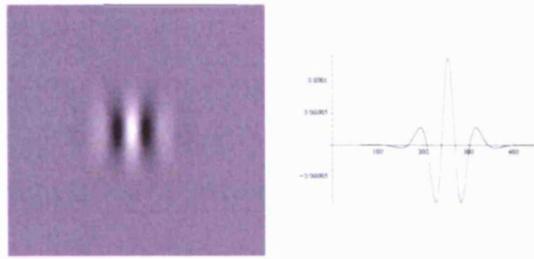


Figure 2.4: An example of an orientation-selective Log-Gabor RF (left) and its cross-sectional response profile (right).

Kingdom and Moulden (1992) implemented this approach in their Multiple Independent Descriptions Averaged Across Scale (MIDAAS) model of luminance coding. The basic idea underlying MIDAAS is that brightness assignments are made individually within each SF channel and then averaged to give the final brightness image. This is achieved by first incorporating the effects of light-adaptation through the application of an independent gain to each (LoG) filter. This is set to be inversely proportional to the average luminance in the region sampled by the filter⁴. Post-filtering, the SF channels (separated at octave intervals) are thresholded (introducing a non-linearity), so only the relatively strong signals expected from features at that scale are transmitted, and thus isolated, before a power transformation of the signal is applied. Zero-crossings in the resultant signals are then used to classify the response profiles of bars and edges, and the corresponding empirically determined brightness profile is then associated with each class of response. The brightness profiles derived from all channels are then recombined through a linear average to give the predicted luminance profile to the stimulus. The algorithm generates the correct psychophysically determined lightness assignments for various simple illusions, as well as assigning brightness correctly to features such as bars mounted on edges. However, the model is dependent on *a priori* knowledge of the brightness profiles associated with each configuration of zero-crossings. It is not clear how such information could be learnt. Furthermore, the model fails to assign brightness correctly in White's effect (Figure 1.1, right) (Kingdom & Moulden, 1992).

Although models based on the RFs of RGCs are capable of explaining some simple lightness

⁴divide-by-zero complications are bypassed through the introduction of a baseline gain

and brightness illusions, they fail on other more complex stimuli. In particular, White's effect easily foils the canonical filter bank approach used to account for simultaneous brightness contrast. In White's effect (Figure 1.1, right), the grey test patch on the left has twice as much border with black bars in the image, than the white bars, and vice-versa for the test patch on the right. Thus, one would expect the patch on the left to appear brighter than that on the right, based on the filter-derived account of simultaneous brightness contrast. However, the opposite effect is observed, as if the test patches have instead been assimilated with the mean luminance of their surrounds. Through the use of oriented filters (the non-rotationally symmetric RF structures usually associated with the simple cells of V1), it is possible to construct filter bank models to accommodate such anomalous illusions into a theory of the primary visual stream (Blakeslee & McCourt, 1999; Dakin & Bex, 2003).

The non-rotationally symmetric RFs of V1 may be modelled by a variety of functions such as the set of Gabor and Log-Gabor functions, the set of Gaussian derivatives, or oriented DoG filters. An example of such a filter is given in Figure 2.4. These RFs are characterised by an elongated centre-ON region with adjacent elongated centre-OFF lobes, or some repeated form of this structure. The orientation of the elongated regions determines the orientation of SFs that the filter will be most responsive to. This allows for the existence of separate filter banks at various orientations for every SF channel, allowing for an orientation-specific SF analysis of the retinal image.

RFs with these characteristics have been used by Kingdom, McCourt, and Blakeslee (1996) to provide a qualitative account of the grating induction illusion (see Figure 2.5), which is more effective than that afforded by rotationally symmetric filters. Consider a bank of filters oriented in the direction of the test stripe, with centre-ON regions set to the same width, such that the inhibiting lobes rest solely on the underlying grating. When the luminance of the grating is low, inhibition is least and the response of the cell to that point of the test stripe is high in comparison with that of a RF overlapping a region where the grating has a high luminance. Hence, the lightness assigned to the test stripe is also oscillatory, but out of phase with the inducing grating by factor of π .

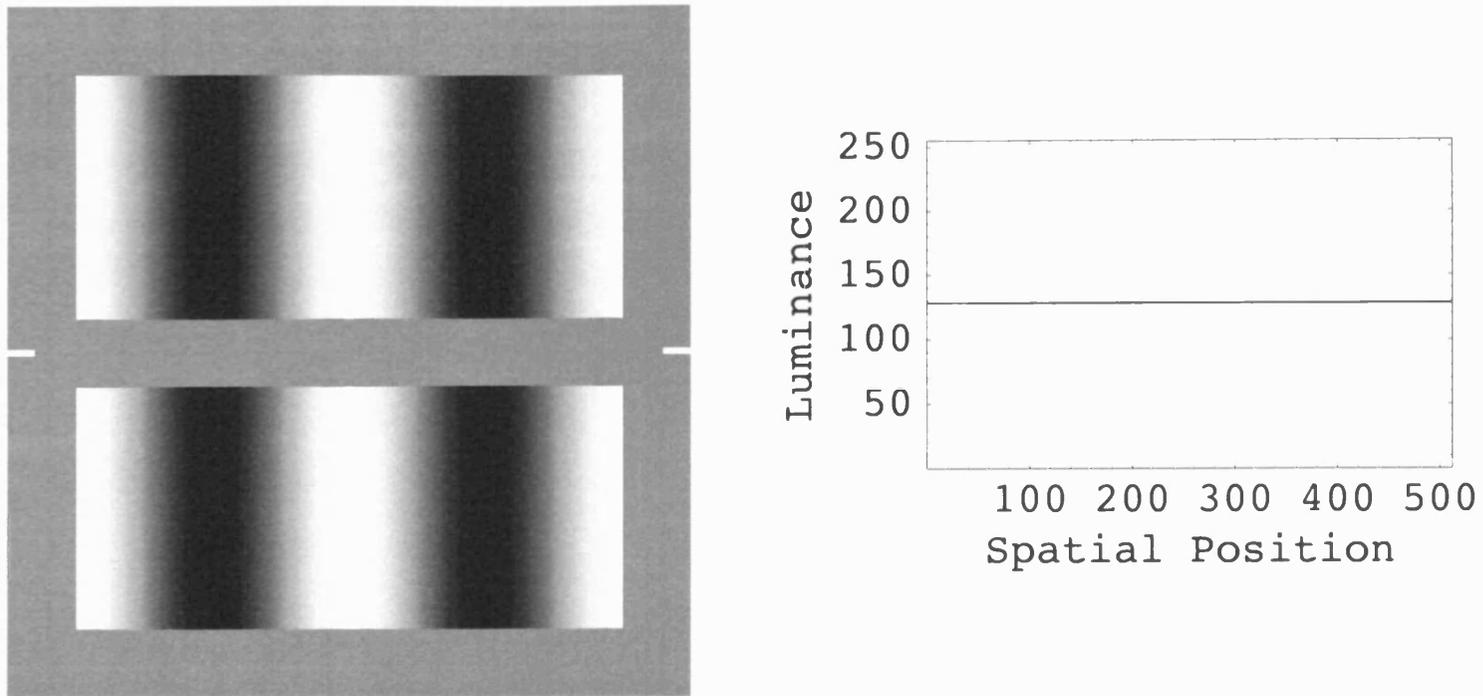


Figure 2.5: An example of the grating induction phenomenon demonstrated by McCourt & Kingdom, (1996). A grating is perceived in the centre strip, perfectly out of phase with the inducing grating (left). However, the luminance of the centre-strip is actually constant, as illustrated by the luminance cross-section (right) taken at the point indicated on the left. The strength of the effect will become stronger as the spatial frequency of the grating decreases (as the reader hold the page closer to them).

A more quantitative model of V1 that combines information from oriented filter banks at multiple scales has been developed by Blakeslee and McCourt (1999, 2001, 2004). Their oriented difference of Gaussian (ODOG) model incorporates 42 filter banks: Six filter banks oriented from -60° to 90° in increments of 30° , with each orientation sampled at seven SFs (with filters tuned to $0.1c/^\circ$ to $6.5c/^\circ$ in octave intervals, covering the visible range for humans). The output at every orientation is combined across scale through a weighted linear sum, where the weights (or gains) are determined by a power-law function of frequency (with a slope of 0.1 when linearised), consistent with the shallow low-frequency fall-off of the suprathreshold contrast sensitivity function (Georgeson & Sullivan, 1975). That is, it takes in to account the fact that the suprathreshold detection level of a high SF grating is at higher luminances than lower SF, implying that the gains on the latter channels are less than that on the former. The salient point here, is that orientation information is combined across spatial scale, consistent with the psychophysical investigations of Olzak and Wickens (1997), to give six orientation channels. Each channel is then normalised by its space-averaged root mean square (RMS) contrast, with the intention of weighting the stimulus features extracted by all orientation channels equally, and ensuring that high contrast features captured by filters at one orientation will not overwhelm lower contrast features captured by filters at other orientations. This stage of processing, necessary to the model's success, is motivated by the non-linear contrast gain responses of simple cells in V1, which can be modelled by dividing the initial half-squared response of a cell by the average half-squared responses from a group of other cells (Carandini, Heeger, & Movshon, 1997). In the context of the ODOG model (which does not explicitly implement this non-linear neural model of contrast gain control), this group of cells is defined to be all other cells tuned to the same orientation, across the entire visual field. Finally, the output from all orientation channels is combined in an un-weighted linear sum to provide the final brightness output. The ODOG model is shown to assign brightness correctly to White's effect, simultaneous brightness contrast, and grating induction, all in alignment with psychophysical data collected on these illusions (Blakeslee & McCourt, 1999). In particular, it predicts the increase in effect strength with SF for White's effect. This places the ODOG model

in a privileged place amongst current theories of brightness perception. However, it should be recognised that this model does not provide any insight as to *why* these specific neural mechanisms should be combined in this way. That is, no progress is made in understanding their computational significance.

An alternative approach to processing information at multiple scales and orientations, inspired by the known statistics of natural scenes (NS), is proposed by Dakin and Bex (2003), and from here on is referred to as the Natural Scene Statistics (NSS) model. It should be noted that the model was originally cast with rotationally symmetric RFs common to the LGN, but as the filters used are approximately equivalent to an un-weighted linear sum of oriented RFs (see Chapter 3), it can also be described in the context of V1 simple cells. The distribution of SFs, or power spectra, of NS are known to be proportional to $1/f^\alpha$, where f is frequency and α is a constant that lies within the range 0.7 to 1.5. Thus, there is an abundance of low SF information. Field (1987) showed that the response of an ensemble of filter bank channels in V1 (modelled using Log-Gabor RFs) to NS with these statistics has a distinct form; filter bank outputs are approximately constant across SF channels. The NSS model starts with the assumption that the visual system has *a priori* knowledge of the statistics of NS, and uses this information to reconstruct the image from the filter bank responses, through a weighted linear sum over SF channels. This is implemented through the adjustment of the gain, $1/f^\alpha$, on each SF channel, such that the output over all channels is approximately constant, in alignment with Fields' findings. The amplitude of each channel is expected by the visual system to be proportional to $1/f^\alpha$ (Tadmor & Tolhurst, 1994), and the modification of the gains comes through the global adjustment of α for all channels. The net effect of this process is to increase the power of low SFs and attenuate higher ones, increasing the brightness of larger patches, or 'filling-in', whilst reducing image detail. This approach is capable of accounting for White's effect (amongst many others), which challenges the authority of the ODOG model. Furthermore, this NSS model is simpler than the ODOG model. Only one set of gains are required (for SF channels), and mentioned above, a direct consequence of the properties of the Log-Gabor model of V1 RFs is that an un-weighted linear sum of filters at the same SF, but over all differing orientation tunings,

is equivalent to filtering with just one centre-surround-like filter. Thus, the model could be implemented at the level of the retina or lateral geniculate nucleus (LGN), as well as V1. The gains are also set according to a specific computational principle, and the RF model implemented encodes information optimally for all SFs and orientations (that is, there are no preferred orientations or SFs).

Local Energy Models

The possible dependence of lightness and brightness on the information contained in lines and edges has already been discussed (in the context of simultaneous brightness contrast). However, none of the approaches discussed so far has been principally concerned with a biologically plausible means of feature classification, and subsequent lightness and brightness assignment. For example, although the MIDAAS algorithm of Kingdom and Moulden (1992) is explicitly driven by lines and edges and their accurate identification, the brightness assigned to the regions they subtend is contingent on a set of heuristics correctly interpreting the filter bank responses.

Morrone and Burr (1988) have avoided the need for labelling filter responses with their underlying causes, through the idea of ‘local energy’. This approach is based on the description of lines and edges in Fourier-space. A line is an even function when the origin is centred on its midpoint, and hence its Fourier-series representation consists purely of cosine components. Similarly, an edge, when considered as a step function with the origin centred on the midpoint of the discontinuity (the mean luminance), is an odd function with a Fourier-series consisting entirely of sine components. Thus, the Fourier phase spectrum is constant for both lines and edges. It is $0/\pi$ at all frequencies for a line and $\pm \frac{\pi}{2}$ for an edge. For other choices of origin the phase spectrum will not be constant, but the distinction between lines and edges is maintained by exploiting the fact that the phase of all Fourier harmonics (referred to as ‘arrival phases’) will be identical at the point of an isolated line or edge; that is, they occur at points of the waveform where the arrival phases are maximally similar, and the average arrival phase (over all harmonics) at that point determines the

nature of the feature: values near zero correspond to a line, and values near $\frac{\pi}{2}$ correspond to an edge.

Although the visual system does not have explicit knowledge of the phase arrivals at all points of the retinal image, this information may be extracted via a system of paired filter responses. If, as the psychophysical and physiological data suggests, there exist an even- and odd-symmetric filter for every point on the retinal image (Pollen & Ronner, 1981), then the even-symmetric filters will respond maximally to the even-symmetric components of the image, and the odd-symmetric filters will respond maximally to the odd-symmetric components. These filters represent the even and odd parts of a local Fourier representation of the retinal image, and can be chosen to be orthogonal in the same way as sine and cosine functions. Thus, the phase spectrum (and hence the phase arrivals) can be determined by proxy at any point.

Local energy is defined as the magnitude of a vector in the orthogonal space defined by the pair of filter responses, and vectors in this space represent the response of the filter-pair to a specific point on the image. Similarly, the mean phase arrival is represented by the orientation of this vector. Mean phase arrivals corresponding to lines or edges, in turn correspond to extrema in local energy. Whether a line or an edge signals a transition from light to dark (or vice versa), depends on the polarity of the phase arrival (positive and negative respectively), which in turn depends on whether the local energy is at a maximum or minimum. Edges that cause a change in lightness or brightness are made to propagate this change until the next line or edge, although a mechanism for performing this operation is not described.

Biologically, the even- and odd-symmetric filters correspond to the RFs of simple cells in V1, and the local energy corresponds to the output of complex cells, the responses of which can be modelled as a non-linear function of the square of inputs from simple cells (Dayan & Abbott, 2001). Morrone and Burr (1988) describe a complete model of this system of cells, implemented with four pairs of complementary filters (modelled by Log-Gabor functions

with constant bandwidth), which span the visible SF range and are approximately equally well tuned to all SFs; each filter-pair are matched to have the same power spectrum, but differ only in phase. Local energy is calculated and analysed independently for each scale, and is capable of correctly signaling lines and edges at each scale.

Later versions of the local energy model (Burr & Morrone, 1994), extend its application to multiple orientations as well as scale, and demonstrate that high SF information structures the way the low SF information is perceived. Additionally, when phase information is sufficiently far apart the features attributed to each set of phase arrivals will not be bound together perceptually and this can lead to transparency effects. However, the model fails completely to locate any features on a pure sinusoid, as local energy is uniform everywhere. Similarly, it fails with heavily blurred images. It is suggested that such low-pass information may be encoded differently, perhaps directly from the luminance itself, and that local energy is used purely for feature detection which then influences lightness and brightness perception.

Lightness and Neural Networks

Models of lightness perception that embody explicit knowledge of the response properties of cortical cells, their interactions, and the anatomical organisation of the visual pathways from the retina through to V4, can be realised through the construction of multi-layer neural networks. A comprehensive discussion of such models is outside the scope of this review, but one such approach courtesy of Grossberg (2003), is worthy of mention for its account of a possible ‘filling-in’ mechanism for lightness. This is of particular note as nearly all theories of lightness that are dependent on lines and edges to successfully determine lightness, require that this information is somehow propagated between features. Grossberg has proposed that boundaries are determined by successive processing in the V1 inter-blob and V2 inter-stripe regions, while the appropriate filling-in is facilitated by successive processing in the V1 blob and V2 thin-stripe regions. These pathways then converge to result in the correct filling in of geometrical shapes determined by the identification of boundaries.

More specifically, boundaries that determine the perception of geometrical shapes are signalled in the V2 inter-stripe area in response to complex cells in the V1 inter-blob area tuned to orientations directed inwards towards to one another. These complex cells pool the input from complementary pairs of simple cells in the V1 inter-blob regions, which although being tuned to the same orientation and position are sensitive to different contrast polarities. It should be noted that in doing this complex cells lose their ability to respond to colour, and so in some respects the boundary is ‘invisible’ to the V1 inter-blob pathway. Networks within V2 are thought to select the boundary grouping from among many possible interpretations of a scene, and may be thought of as performing a type of statistical inference to select and complete the statistically most favoured boundary groupings of a scene (while suppressing noise and incorrect groupings). This system will build boundaries across space like in the Kanizsa square illusion, and is by default insensitive to contrast polarity. The filling-in process begins in the retina where lateral inhibitions are modelled by a centre-ON surround-OFF network of cells capable of discounting the effects of spatially varying illuminants. This results in a modified luminance profile for the retinal image, which is used in networks downstream in the blob cortical system to fill space by means of an anisotropic diffusion process, subject to the constraint of the geometrical shape determined in the complementary pathway, and thus making the boundary visible.

In a more recent development of this modular neural processing strategy, relative measures of surface lightness are mapped to an absolute lightness scale, by means of a novel ‘anchoring process’ (Hong & Grossberg, 2004; Grossberg & Hong, 2006). As mentioned in the discussion of low-dimensional and Retinex models of colour vision, measuring the lightness of surfaces relative to the brightest surface (assumed to be white) or the mean reflectance of natural scenes (assumed *a priori*), in an attempt to generate an absolute measure of lightness, does not work in all circumstances. Grossberg and Hong (2006) circumvent the problems encountered with these approaches by introducing a higher-level stage of processing (assumed to exist in V2 or V4), which measures the lightness of all surfaces relative to highest output in a large scale blurred version of a neural surface representation of the scene, which exists upstream in V1 or V2. This method of anchoring the lightness of sur-

faces to a specific reference point is shown to be more successful than the mean reflectance and highest-luminance-as-white rules. Interestingly, the *large* scale blurring has the effect of making the anchoring reference point lower than the value attributed to any *small* scale region of the image that is also the brightest. Subsequently, such regions are predicted to be super-luminous, or sources of light in the image, and in agreement with experiment.

In addition to this extra unit of higher-level processing, a new method for filling in surfaces based on boundary information is also proposed (Hong & Grossberg, 2004; Grossberg & Hong, 2006). The previous filling-in algorithm based on a diffusive processes is replaced by an approach which uses the magnitude of the response in low-SF channels as a measure of the reflectance of a surface, and performs the filling-in across the surface by means of long-range horizontal connections, which are gated by the boundary signals (created using the same processing stage as the previous incarnation of the model), and constrained in cortical distance to lie within the maximum range of possible intra-cortical connections, as determined by experiment (Angelucci et al., 2002; Stettler, Das, Bennett, & Gilbert, 2002). This has the advantage of performing the filling-in stage with fewer iterations than required with the diffuse filling-in algorithm, by two orders of magnitude.

Arranging visual information processing into sub-modules of functionally-specific neural networks, and integrating the outputs to yield visual perception, has the capability of assigning lightness correctly to simultaneous contrast, the Craik-Cornsweet-O'Brien effect, Mondrians, natural scenes, and can capture some of the temporal dynamics of the filling-in process. However, they suffer from requiring the determination of *many* free parameters - in their most recent work Grossberg and Hong (2006) require 50 free parameters to be determined or fitted to experimental data, with different phenomena requiring different sets of parameters. This greatly diminishes the explanatory power of such approaches, as the choice of parameters needs subsequent motivation. Furthermore, such models demonstrate that a lot of lightness computations require processing beyond the low-level functionality of V1, in particular because global-scale calculations, such as the anchoring process described above, are required at some stage. Processing on these scales is more commonly associated

with areas of the cortex downstream from V1 (Grossberg & Hong, 2006).

Models of Higher-Level Vision

As discussed at the beginning of this chapter, if all lightness and brightness phenomena are derived *solely* from low-level visual processes, then they ought to be unaffected by the presence of higher-level properties (or cues) present in the scene. For example, depth and form perception are not involved at this stage of visual processing, and subsequently cannot affect lightness. Thus, all images, no matter how complex in composition, are processed as if they are Mondrians, and any illusory phenomena are purely the result of such processes, and can be labelled as brightness phenomena. However, it has been shown that lightness perception is drastically altered when cues pertaining to the properties of complex scenes (such as depth and form), are present.

Some models are based on the view that brightness cannot exist independently of lightness and lightness constancy, and all such judgements are fundamentally determined by higher-level interpretations of the scene. As this latter point cannot be discounted, some of the more prominent higher-level approaches will be discussed.

The Bayesian Approach: Inferring the Most Likely Origin of the Luminance

One approach to vision, usually attributed to the works of Helmholtz (1887), is to cast it in terms of Bayesian inference. That is to say, that all visual perception is, is the result of the visual system trying to infer the most likely underlying real-world situation that generated the visual input, given prior knowledge of the statistics of natural scenes in the world (which are presumably learnt whilst the visual system is developing). Figure 2.6 demonstrates how simultaneous brightness contrast can be explained in this way. Figures 2.6b-d represent the most likely scenarios which could generate the spatial configuration in Figure 2.6a (equivalent to the simultaneous brightness contrast stimulus illustrated in Figure 1.1, left). They include: Surrounds with different illumination under the same illumination (B); surrounds of the same luminance, but one is in shadow (C); surrounds of the same luminance, but only one is illuminated (D); and, the surrounds form difference faces on a cube, one of which is in

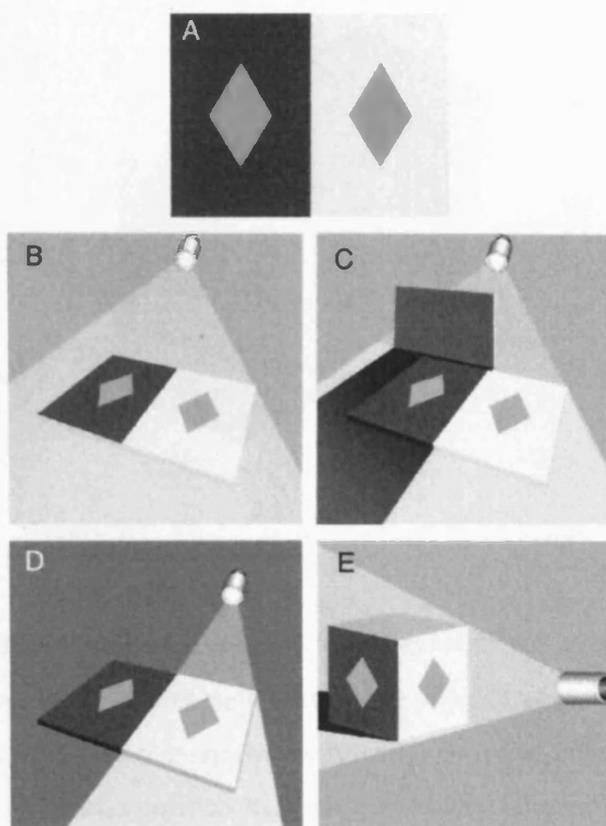


Figure 2.6: Simultaneous brightness contrast (A) and four different ways that this particular configuration of surfaces could have arisen: Surrounds with different illumination under the same illumination (B); surrounds of the same luminance, but one is in shadow (C); surrounds of the same luminance, but only one is illuminated (D); and, the surrounds form different faces on a cube, one of which is in direct illumination whilst the other isn't (E). Reprinted from Williams, McCoy & Purves (1998).

direct illumination whilst the other isn't (E). As only one of the four causes for Figure 2.6a is due to an actual change in reflectance for the surrounds, it's more likely that one of the surrounds is illuminated in a different way to the other, in which case, if the luminance generated from the two test patches is the same, then it is likely that one ought to be brighter than the other if lightness constancy is to be maintained (Williams, McCoy, & Purves, 1998).

A computational framework for inferring the origin of the luminance through calculation of the underlying reflectance, shading, and illumination intrinsic images, has been suggested

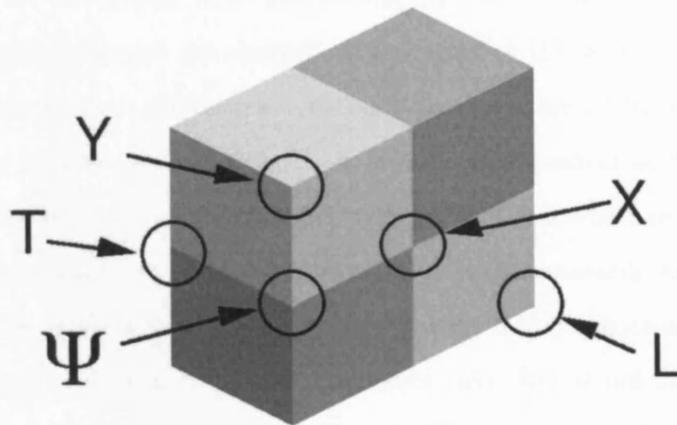


Figure 2.7: Examples of junctions in an image.

by Adelson and Pentland (1996). A cost function is defined, and a requirement made that the interpretation of the luminance preferred by the visual system must always yield the minimum cost. This is essentially equivalent to using Bayesian inference, where the cost of each interpretation is analogous to its prior probability of occurrence. Although the premises of this approach are reliable and logical, it only works for trivial configurations of surfaces in 3-D space, is computationally intensive, and the choice of cost function (or prior probability) is hard to justify, as the statistics of naturally occurring shapes and possible methods of illumination are not available for analysis. Hence, the power of such models to correctly explain lightness in complex scenes remains open to debate. However, there exists an alternative system for assigning lightness in complex scenes, which although heuristic in nature, is highly successful in explaining many of the phenomena discussed so far, and doesn't suffer from the problems faced by the construction of such computational models. The discussion of this paradigm, based on image features referred to as junctions, occupies the next part of this chapter.

Local Junction Analysis

A good source of information about the nature of edges in a global context, whether they correspond to a change in reflectance, a sharp change in shading due to a change in surface

orientation, or the positioning of an attenuating transparent strip over an object, may be determined by proxy through the analysis of ‘junctions’ in the image. Formally, a junction is defined as a place where two or more contours come together (Adelson, 2000). Junctions can be classified into various types: L, T, X, Y, and ψ , dependent on the configuration of the contours relative to one another (as betrayed by the letter which represents them). For example, two connected but differently oriented surfaces necessarily exhibit a Y-junction at the end points, while a ψ -junction signals a change in reflectance at some point along this edge (Sinha & Adelson, 1993). A reflectance edge that is not deliberately made to coincide with any corners of the underlying 3-D object exhibits a T-junction at the end points, and the intersection of two reflectance edges on a surface, as in a checker-board, is represented by an X-junction. Finally, the unconnected corner of a surface is represented by the L-junction. Examples of these junctions are illustrated in Figure 2.7. The ‘strength’ of the interpretation inferred by the junction is assumed to decrease with distance along the edges from the junction point; hence the interpretation assigned to a single edge can be affected by the existence of nearby junctions. This system facilitates, to some extent, the reversal of the mapping between a 3-D object onto a 2-D image. However, it should be noted that there does not exist any well determined ‘look-up table’ that automatically associates a junction type with a specific interpretation. For example, it is believed by some that T-junctions indicate the presence of transparency in the image, while others insist that T-junctions impart a sense of ‘belongingness’ between the two surfaces that form the stem of the T (Anderson, 2001).

The junction paradigm correctly predicts that images with an absence of Y- and ψ -junctions, which are likely to correspond to illumination edges (and hence 3-D shapes), will appear as flat (Sinha & Adelson, 1993). Furthermore, by keeping all image attributes constant, other than a slight perturbation to lines and edges, such that particular junction types are created or abolished, the appearance of a scene can be drastically changed. An analysis based on these principles can be used to provide a qualitative explanation of some of the complex phenomena described in Chapter 1. For example, the ‘Wall of Blocks’ demonstration can be explained through a change from ψ - to X-junctions (Figure 1.2, left, to Figure

1.2, right). When this occurs the interpretation of the scene is altered from one where luminance changes are attributed to reflectance and shading on a 3-D structure, to one where the luminance changes are attributed solely to reflectance changes in a 2-D image and the existence of a transparency across the image. In the latter case, the inferred transparency, combined with the similar luminance of both sets of grey diamonds, leads to an inference that the reflectance of the grey diamonds ‘under’ the dark transparency must be lighter than their counterparts with the light surrounds. Thus, this illusion is attributed to higher-level reflectance calculations.

However, junctions have failed to provide a *complete* account of White’s effect. In the standard White’s effect (as illustrated in Figure 1.1, right), T-junctions exist between the grey test patches and their coaxial bars (which form the stem of the Ts, while the top component comes from their collinear contours with adjacent bars). Anderson (2001) claims that these T-junctions signal the existence of a transparency in a scission of the image into its causes. In this paradigm the grey test patches are assumed to be the transparency, such that the bars on which they are placed can influence their luminance. Hence, if the black bar is removed then the lightness of its test patch becomes lighter, whereas if the white bar is removed then the lightness of the test patch becomes darker. So, in an attempt to remain lightness constant this process yields White’s effect⁵. Alternatively, the T-junctions may signal that the test patch and their coaxial bar ‘belong’ to one another, such that they are only interpreted within this context, with the only simultaneous brightness contrast effect that can occur and influence brightness being the one between the coaxial bar and the test patch, which also produces White’s effect (Todorovic, 1997; Gilchrist et al., 1999). However, Figure 2.8 demonstrates a circular version of White’s effect with no junctions, in which the effect is still as strong as the more traditional version (Howe, 2005). Furthermore, junctions have not been shown to explain the SF dependency of White’s effect (White, 1981; Kingdom & Moulden, 1991; Blakeslee & McCourt, 1999; Anstis, 2004). Hence, the role of junctions in accounting for this phenomenon is controversial.

⁵A salient point here, is that such models make the implicit assumption that the area of the brain which performs this higher-level interpretation, receives the image information in an unadulterated form.

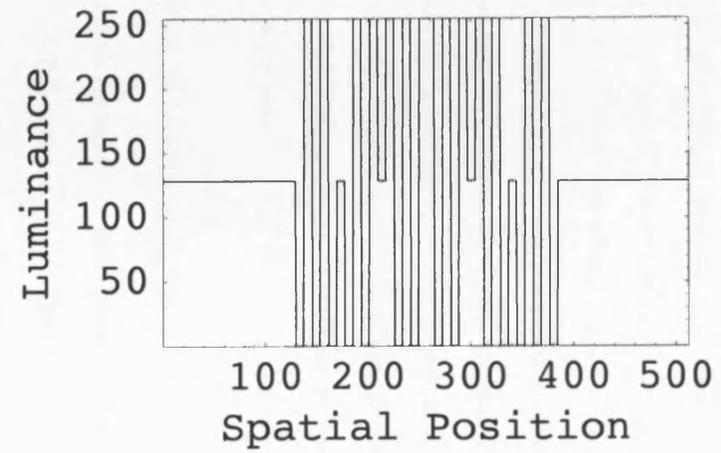
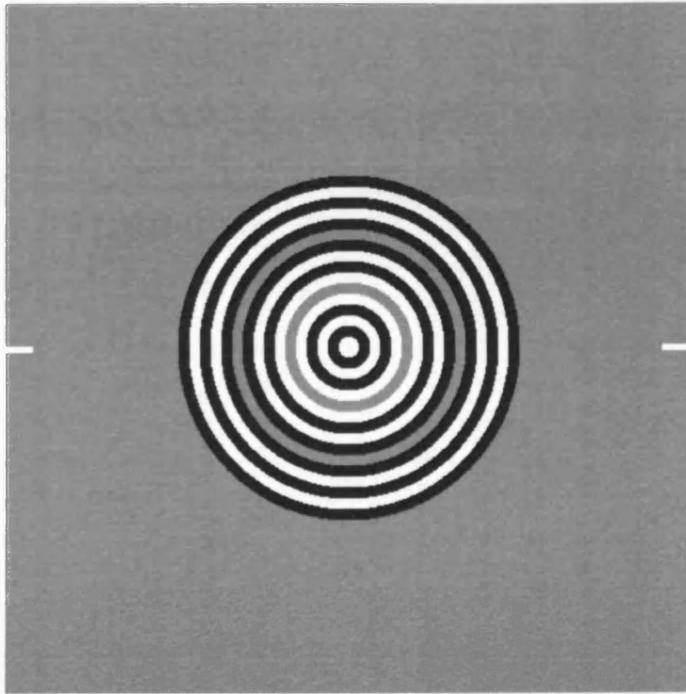


Figure 2.8: An example of the type of stimulus used by Howe to demonstrate how White's Effect is not dependent on junctions within the stimulus.

The relative success of local junction analysis in facilitating scene interpretation motivates their further study. Furthermore, it is possible that the information in junctions is actually being used by higher-level areas of the visual cortex. Physiological studies of RFs in V4 have revealed cells which respond to complex junction-like contour configurations, highly selective for size, shape, and orientation (Gallant, Connor, Rakshit, Lewis, & Essen, 1996). This is not an unreasonable hypothesis, as the importance of lines and edges in low-level vision has already been emphasised, and so it is conceivable that at higher-levels of processing, information derived from complex ensembles of lines and edges, such as junctions, may also be utilised in some way.

The Anchoring Theory of Lightness

Another heuristic approach to lightness perception, is the Anchoring Theory of Lightness (Gilchrist et al., 1999). A full account of this theory is beyond the scope of this work, but the principles on which it is based will be described.

At the heart of this theory is the idea that the lightest surface in a scene is perceived as the 'white' (the anchor), and that the lightness of all other surfaces are computed relative to the white surface (the scaling process). In this functional sense it is similar to the first incarnation of Retinex (Land & McCann, 1971), although the Anchoring Theory of Lightness is never cast as a functioning computational model that is capable of making quantitative predictions.

In addition to the anchoring and scaling processes, is the notion of 'frameworks' for interpretation. More specifically, as well as a global anchor and scaling, there may exist an independent local anchor(s) and scaling(s) derived from a given surface's 'belonging' to a given region. The final lightness attributed to a surface is a trade-off between local and global lightness assignments. For example, in the simultaneous brightness contrast stimulus (Figure 1.1, left), the local frameworks are the light surround and the test patch within it (left), and the dark surround and the test patch within it (right). In the framework on

the left the surround is anchored to white and the test patch is considered darker. For the framework on the right, the test patch is anchored to white while the surround is considered darker. Globally, the surround on the left is the white. As the test patch on the right cannot be the global white it is scaled to be just below, and as the test patch on the left is not as dark as the surround on the right, it is scaled to be lighter. Thus, the test patch on the right is perceived to be brighter than that on the left, correctly predicting the existence of simultaneous brightness contrast.

A similar interpretation is afforded to White's effect (Figure 1.2, right). The test patch on the dark bar is said to belong to the framework of the dark bars, and the test patch on the white bar is said to belong to the framework of the white bars. In the former framework the test patch is the local white, while in the latter framework the bars are the local white. The white bars are also the global white, so the test patch on the dark bar has to be darker, while the test patch on the white bars is not as dark as the black bars, and has to be lighter. Hence, the test patch on the dark bars is lighter than that on the white bars, producing White's effect. Like local junction analysis, it is not clear how this paradigm can account for the SF dependency of the effect.

Summary

Low-level models of vision are based on the underlying physiology of the primary visual stream, and within this context lightness and brightness phenomena are artefacts of the processes of encoding and decoding (or reconstructing) the original image. These models are successful at attributing brightness in simpler visual illusions such as simultaneous brightness contrast, and in some cases White's effect.

Low-level physics-based models of colour vision, based on the adaptational mechanisms of the retina, suffer from the problem of requiring a reference surface or some sort of anchor in the visual world, in order to be able to infer the illuminant and attribute lightness with the aim of retaining lightness constancy (Maloney, 1999). Furthermore, there exists only one biological model of how the primary visual stream may achieve such calculations:

Retinex theory (Land & McCann, 1971; Land, 1983). The Retinex theory of colour vision makes relative measures with regards to lightness; the lightness of a surface is attributed relative to the lightness of all other surfaces. As a consequence of this, the theory also require a reference surface or anchor, against which all relative measures are implicitly made. Furthermore, Retinex theory is not capable of accounting for visual phenomena such as simultaneous brightness contrast, without additional modification (Rudd & Arrington, 2001).

Simple theories based on modelling the response of RGCs through their RF properties can be successful in predicting the existence of simultaneous brightness contrast (Cornsweet, 1970; Goldstein, 2003), but are not cast in a meaningful computational context - they are models of the underlying physiology and are not aiming to achieve a particular calculation in the way that Retinex and the physics-based models of colour vision are. Thus, models constructed in these terms are models of brightness and not lightness. These theories also suffer from the need for some kind of ‘filling-in’ process to propagate brightness signals between the responses of cells at the edges of the surface (where they are instigated).

However, RF-based models which avoid the need for an explicit ‘filling-in’ process, can be constructed by filtering the image at multiple scales, and possibly at multiple orientations for each scale (using the RF properties of cortical neurons), and integrating this information into a single reconstruction of the original image (Kingdom & Moulden, 1992; Blakeslee & McCourt, 1999; Dakin & Bex, 2003). These approaches have also been successful in correctly assigning brightness in phenomena such as grating induction, as well as the canonical simultaneous brightness contrast effect. White’s effect, however, remains particularly hard to account for, and only models which implement some kind of re-weighting or re-normalisation procedure across separate SF and/or orientation channels, before integrating information into the final reconstruction have been successful in accounting for both simultaneous brightness contrast *and* White’s effect (Blakeslee & McCourt, 1999; Dakin & Bex, 2003).

Arranging visual information processing into sub-modules of functionally-specific neural

networks, and integrating the outputs into to yield visual perception, has the capability of assigning lightness correctly in a variety of circumstances, and for a variety of visual phenomena (Grossberg, 2003; Hong & Grossberg, 2004; Grossberg & Hong, 2006). However, the explanatory power of such approaches is weakened by the large number of free-parameters that arise from the computational complexity of these models, and which subsequently need to be determined. Furthermore, in order to achieve their results modules of higher-level processing downstream from the low-level functionality of V1 are necessary.

Higher-level models of vision, which cast vision as the problem of inferring the source of luminance through the interpretation of the scene, have been shown to be successful in providing accounts for simultaneous brightness contrast, White's effect, and more complex phenomena such as Adelson's 'Wall of Blocks' demonstration, which low-level models have thus far not addressed.

However, these models fail at providing quantitative predictions. The Bayesian, or pure statistical inference approach, has proven successful at explaining lightness in a variety of both simple and complex phenomena (Purves et al., 1999), and could be built into a computation model. However, the approach is impeded when it comes to determining and justifying the choice of prior probability distributions, as the statistics of naturally occurring shapes and possible methods of illumination are not available for analysis. It is also 'computationally expensive', requiring large amounts of processing time (Adelson & Pentland, 1996). As such, no fully functioning computational models of lightness employing this approach are known to exist.

Assigning lightness through the analysis of junctions in the scene has proven successful in correctly attributing brightness in complex scenes such as Adelson's 'Wall of Blocks' demonstration, and simpler configurations such as White's effect (Adelson, 1993; Sinha & Adelson, 1993; Anderson, 2001). Junctions appear to be a good proxy for interpreting cues as to the precise configuration of the scene. However, this paradigm is not capable of making quantitative predictions. Furthermore, this approach fails at correctly assigning lightness

to the circular variant of White's effect, which is devoid of any junctions (Howe, 2005).

The Anchoring Theory of Lightness has proven exceptionally successful at attributing lightness in many complex phenomena, including simultaneous brightness contrast and White's effect (Gilchrist et al., 1999). It achieves its success through a combination of a global analysis of the scene, together with a cascade of increasingly local analyses, by means of a set of heuristic rules. Furthermore, it attributes lightness relative to the the white (or anchor) in the scene. The main problem with this approach, is that which plagues all other higher-level models of vision - it is not capable of making quantitative predictions and has yet to be cast as a biologically plausible computational model of visual processing. Additionally, this approach is not capable of offering an explanation as to why the magnitude of White's effect increases in parallel with the SF of the inducing grating in the stimulus.

It is possible to interpret the above in such a way that reinforces the views about vision set-out at the beginning of this chapter - that low-level vision *alone* can account for 'simpler' visual phenomena such as simultaneous brightness contrast and White's effect, and that they fail only when explicit cues pertaining to the exact configuration of the scene are present in the image. It is not believed that such cues genuinely exist in simultaneous brightness contrast and White's effect. Thus, higher-level, more complex theories are not required to account for these types of phenomena.

Chapter 3: An Analysis of Current Neural Models

White's effect and simultaneous brightness contrast are of particular importance in the study of vision, as accounts for these phenomena come from *both* theories of low- and high-level vision, as described in the previous chapter. Hence, implicit in the study of these phenomenon is the exploration of where low-level vision ends, and high-level vision begins; where visual information is coded and subsequently decoded to infer the nature of the source. So far, it has not been possible to resolve this moot point.

Although low-level neural models have been shown to be competent at explaining simultaneous brightness contrast, White's effect has proven particularly tricky, and to the knowledge of the author, only two theories capable of accounting for both simultaneous brightness contrast *and* White's effect are known to exist (Blakeslee & McCourt, 1999; Dakin & Bex, 2003). As this work is primarily concerned with ideas that place such illusions in low-level vision, a deeper understanding of how low-level vision alone can concurrently account for phenomena such as simultaneous brightness contrast *and* White's effect is necessary, if these views are to be upheld. Hence, there is a need for a more detailed understanding and evaluation of exactly how and why current models of brightness can account for White's effect - are the mechanism(s) responsible for the effect biologically plausible, and are they part of any meaningful computational paradigm? Therefore, a greater insight into which of the precise components of these models are fundamental in explaining these effects, is required. A suitable entry point into this line of enquiry begins with a detailed analysis and comparative study of the only two low-level models capable of accounting for simultaneous brightness contrast and White's effect: the ODOG model proposed by Blakeslee and McCourt (1999), and the NSS approach suggested by Dakin and Bex (2003).

In this study, both models are assembled computationally from the specifications detailed in their respective publications, and the brightness profile attributed by each model to the same variation on White's effect is compared to psychophysical data (to be discussed below). Additionally, an examination into the robustness of these brightness profiles to the

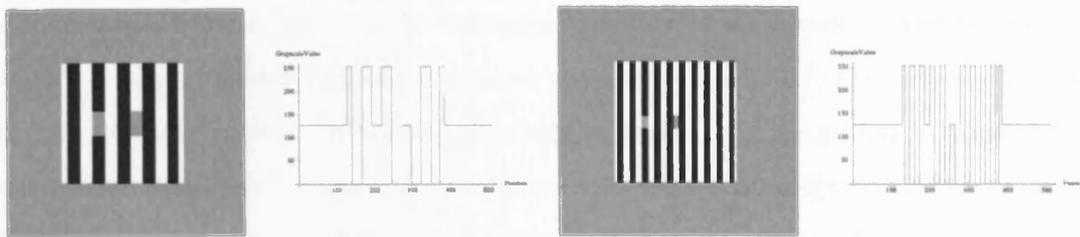


Figure 3.1: Variations of White's effect used in this study, with cross-sectional luminance profiles at the mid-point. The frequency of the inducing grating is $0.5c/^\circ$ (left) and $1.0c/^\circ$ (right).

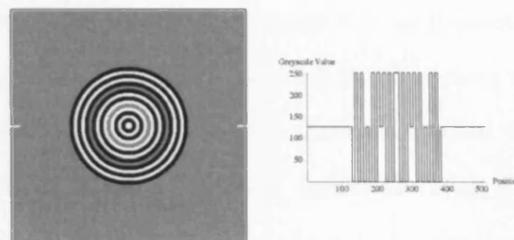


Figure 3.2: A circular variation of White's effect stimulus with an the inducing grating frequency of $1.6c/^\circ$ (left), and a cross-section of the luminance profile at the mid-point (right).

removal of key components of each model is carried out in an attempt to isolate, understand, and evaluate the key factors that may be responsible for brightness judgements in low-level vision.

White's Effect: Test Stimuli and Psychophysical Data

An example of White's effect is given in Figure 3.1 (left). This particular variant was originally used by Blakeslee and McCourt (1999) in their computational and empirical study of the effect, and is similar to the original first presented by White (1979), the only significant difference being that the grey test patches in the image are in-line with one another as opposed to vertically displaced. As already mentioned in this work, the test patch on the black bar should appear lighter than the test patch on the white bar, as if the test

patches have assimilated some of the luminance from the adjacent bars, in contradiction to what is predicted by simultaneous brightness contrast. White (1981) also showed that the magnitude of the effect is dependent on the spatial frequency of the grating, so long as the aspect ratio of the test patches remains constant (Kingdom & Moulden, 1991). An example of this dependency on SF should be apparent in Figure 3.1 (right) where the SF of the inducing grating has been doubled relative to the variant in Figure 3.1 (left), from $0.5c/^\circ$ to $1.0c/^\circ$ (a more complete discussion of the effect's dependency on SF will be postponed until Chapter 5). It should also be noted at this stage, that the exact spatial configuration of the stimulus need not match the one in Figure 3.1, as illustrated in Figure 3.2, which demonstrates the effect when the grating and test patch regions take the form of circular rings (removing the junctions from the original version - a point which will be returned to later on this chapter) (Howe, 2005). However, unlike the spatial configuration, the luminance of the test patches relative to the that of the inducing grating is of critical importance for the existence of the effect. Spehar, Gilchrist, and Arend (1995) have shown that the effect will not persist when the luminance of the test patches are above that of the light bar (when the luminance of the light bar is below that of the test patches), and when the luminance of the test patches is below that of the dark bar (when the luminance of the dark bar is greater than that of the test patches). Examples of such variants of White's effect can be seen in Figures 5.32 and 5.33, and will also be discussed in greater detail in Chapter 5.

The variant of White's effect used for the central part of this study is that depicted in Figure 3.1 (left). The White's configuration part of the stimulus spans 10° of visual angle (represented by 256 pixels in this case), both vertically and horizontally, with the frequency of the inducing grating at $0.5c/^\circ$. The height of each test patch is twice its width (1° in this case), and the luminance of each test patch is set at the mid-point between the luminance of the light and dark bars (a grey-scale level of 128). The full stimulus is 20° in width and length, and is constructed by presenting the White's configuration on a background with a luminance set at the mean luminance of the image (also a grey-scale level of 128). This configuration suppresses any erroneous effects that may occur as a result of using large spatial filters. This specific variant of White's effect was chosen for both its simplicity and

because psychophysical data has been acquired using this stimulus (Blakeslee & McCourt, 1999), and hence the predicted brightness profiles for this stimulus can be directly linked with experimental data. More specifically, Blakeslee and McCourt (1999) report that the test patch on the black bar is found with a brightness of 54% of the white in the scene, and that the test patch on the white bar is found with a brightness of 45.5% of the white in the scene. Thus, the overall effect strength (defined as the brightness of the test patch on the white bar, less that of the test patch on the black bar), is calculated to be 8.5%.

In addition to the above, two other variants of White's effect will be tested with the models. The variant in Figure 3.1 (right), with an inducing grating at double the frequency ($1.0c/^\circ$), but with the same aspect ratio of that in Figure 3.1 (left), will be used, and ought to predict that the effect strength increases, as discussed above (White, 1981). Although this effect has already been demonstrated for the ODOG model (Blakeslee & McCourt, 1999), using gratings with lower SFs than $0.5c/^\circ$, this has not yet been demonstrated with the NSS model (although one would expect the model's predictions to be consistent, given that it has a relatively scale-invariant approach to processing SF information). Furthermore, when deconstructing the models in order to understand and evaluate their underlying mechanisms, using a stimulus with a different SFs will serve as a check that any conclusions derived from this analysis are applicable across all SFs that the models respond to. The circular variant of White's effect as depicted in Figure 3.2, with an inducing grating of $1.6c/^\circ$, will also be used. The motivation for testing it with the models comes from the fact that it has not yet been tested with either of them, and because this particular variant has a different spatial configuration - the test patches (or rings) only have boundaries with their coaxial rings, and not with any bar that they are placed within. Thus, if the models successfully predict the existence of White's effect in this configuration, then it will demonstrate that they can account for the effect in more general terms, and not just for a particular configuration.

The Natural Scene Statistics Model

The model of brightness described by Dakin and Bex (2003) is inspired by the work of Field (1987), who developed an efficient image-coding scheme based on a Log-Gabor model of V1 RFs. Efficient image-coding in this context, relates to the ability of each filter bank to represent the greatest amount of information,⁶ at a particular SF and orientation, with the fewest possible filters, and for visual information at all SFs and orientations to be represented by as few filter banks as possible⁷. With this coding scheme, it was found that the variance (or energy) in each filter bank's response to natural scenes is approximately constant and flat across all SF channels (Field, 1987). More specifically, a plot of log-energy vs. log-SF yields a straight line with a mean slope of -0.04 (Dakin & Bex, 2003). This is a direct result of the fact that the SFs of natural scenes have a distribution of $1/f^\alpha$, where f is frequency and α is a constant that is always found to lie in the range of 0.7 - 1.5.

Based on these findings, a simple model of brightness coding is proposed, and is illustrated schematically in Figure 3.3⁸. It is conjectured that brightness arises out of the need to be able to reconstruct the source image from the information within the responses of each SF channel, and that the visual system has *a priori* knowledge of the statistics of natural scenes, expecting all candidate reconstructions to give rise to the observed constant response across SF channels (presumably as a mechanism for validating reconstructions). If the latter is not possible with an unweighted linear combination across SF channels, then it can be achieved by introducing a multiplicative gain of $1/f^\alpha$ across every channel, and choosing α to result in a slope of -0.04 on a plot of log-energy vs. log-SF. If the viewed image has a perfect $1/f^\alpha$ distribution of frequencies, then α is close to 0 and the correct reconstruction is approximately that of the unweighted linear combination across SF channels. If the image has a frequency distribution that deviates from that of natural statistics, then α is chosen such that low SFs are adequately amplified and high SFs adequately attenuated, until the desired response

⁶in this case measured by the variance in filter responses to an image.

⁷hence the responses from all filters become maximally un-correlated.

⁸It should be noted, that the original version was cast in terms of circularly symmetric RFs, which could be used in the LGN, but in this work it has been interpreted in terms of oriented RFs, which could be implemented in V1. However, the two versions are approximately equivalent, as the sum over orientations for a particular SF is approximately equal to the circularly symmetric RF.

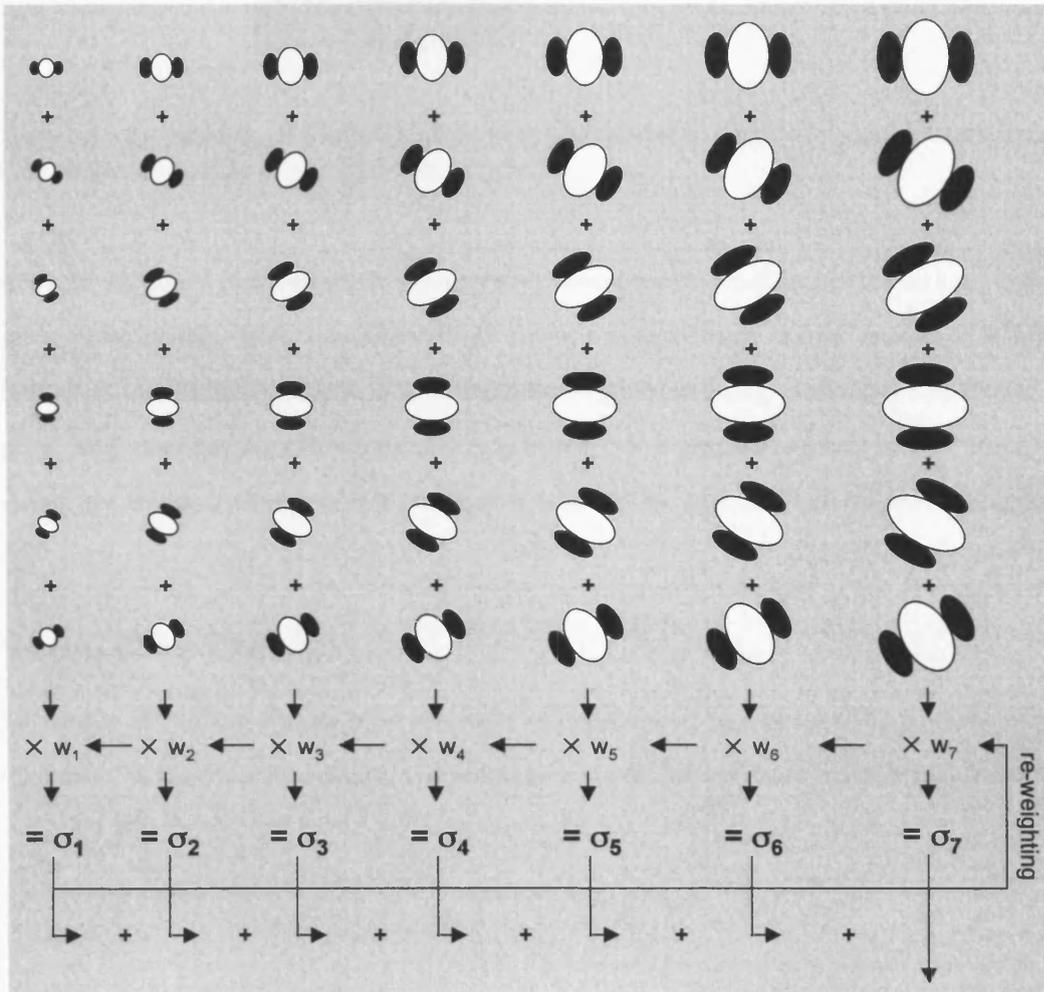


Figure 3.3: A schematic representation of the NSS model: The output from banks of Log-Gabor filters are summed across orientation to produce SF channels (the σ 's), which are then re-weighted with a $1/SF^\alpha$ gain, until the linearly combined output satisfies the expectations of the visual cortex.

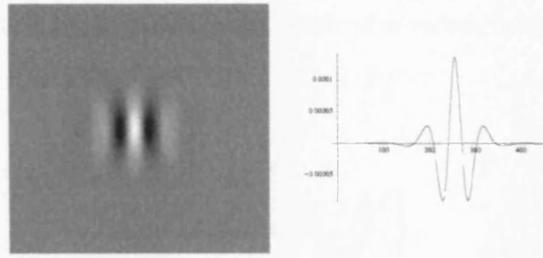


Figure 3.4: An example of a Log-Gabor filter in the spatial domain (left), and a cross-section of the response profile at the mid-point (right).

across SF channels is achieved. It is suggested that a reconstruction of this sort on images that deviate heavily from the statistics of natural images, such as the variant of White's effect used in this study, can lead to anomalous brightness being attributed to parts of an image, and that the resulting amplification of low SF information leads to the 'filling-in' in what are traditionally thought to have been edge-driven lightness and brightness illusions.

The Log-Gabor Code

The details of the Log-Gabor code as used and constructed to replicate the NSS model are as follows: In the Fourier domain, the response of each RF (or filter) to SFs is defined as,

$$G(f) = \exp \left\{ - \frac{[\ln \frac{f}{f_{peak}}]^2}{2[\ln \frac{\sigma_{SF}}{f_{peak}}]^2} \right\}, \quad (3.1)$$

where f is frequency (and represents the radial component of a vector in Fourier-space), f_{peak} is the peak response frequency, and σ_{SF} is the relative spread in response frequency. These filters are uniquely defined in the Fourier domain, and no analytical expression exists for Log-Gabor filters in the spatial domain. Hence, all filtering is carried out in the Fourier domain, and the numerical inverse-Fourier-transform is used to produce the resultant image in the spatial domain. An example of a Log-Gabor filter in the spatial domain is shown in Figure 3.4. Similarly, in the Fourier domain the response sensitivity of each filter to a

SF pattern oriented at an angle θ (the polar angle of a vector in the 2-D Fourier-space), is given by,

$$G(\theta) = \exp \left\{ - \frac{[\theta - \bar{\theta}]^2}{2\sigma_\theta^2} \right\}, \quad (3.2)$$

where $\bar{\theta}$ is the peak response angle, and σ_θ is the spread in angle response sensitivity. The final filter is given by,

$$G(f, \theta) = G(f)G(\theta). \quad (3.3)$$

This separable polar distribution of SF and orientation is unique to Log-Gabor filters. Another advantage of Log-Gabor filters over conventional Gabor filters, is that on a log-SF axis the response curve is a perfect Gaussian distribution and does not over represent low SFs, unlike the conventional Gabor model of V1 RFs.

In order to effectively (and indeed efficiently) cover the Fourier domain, the bandwidth of the SF channels⁹, was kept to 1 octave, and the peak frequency varied from 2 cycles per-image to 128 cycles per-image¹⁰ in octave intervals (equivalent to $0.1c/^\circ$ to $6.4c/^\circ$). To achieve this, it is necessary that,

$$\sigma_{SF} = f_{peak} \exp \left\{ - \frac{1}{2 \log_2[e] \sqrt{2 \ln[2]}} \right\}, \quad (3.4)$$

$$\ln[\sigma_{SF}/f_{peak}] = \text{constant}, \quad (3.5)$$

⁹bandwidth is defined as $\log_2(A - B)$, where A and B are the SFs at which the RF yields half of the maximum response.

¹⁰filters selective for frequencies below 1 cycle per-image were avoided as they were too small to adequately sample in the Fourier domain, and filters principally selective for frequencies above 128 cycles per-image were avoided as they were too small to adequately represent in the spatial domain.

for all peak frequencies. The resultant overlap between filters at differing peak SF tunings leads to an approximately equal amplification for all SFs in this range.

Similarly, for SFs at all orientations to receive equal amplification with filters tuned to 6 separate angles (from 0 to 2π in intervals of $\frac{\pi}{6}$), it is necessary that the filters have an orientation bandwidth of $\frac{\pi}{3}$, such that,

$$\sigma_{\theta} = \frac{\pi}{6\sqrt{2\ln 2}}. \quad (3.6)$$

Each SF channel is constructed from an unweighted linear combination of all filter banks with the same SF selectivity, but differing orientation, and the energy in each channel is then determined by discretely sampling the response-envelope of each SF channel. The response envelope, $V(x, y)$, is defined as,

$$V(x, y) = \sqrt{E(x, y)^2 + O(x, y)^2}, \quad (3.7)$$

where $E(x, y)$ is the output in the spatial domain from the even Log-Gabor filter, and $O(x, y)$ is the output from the odd Log-Gabor filter (it should be noted that all reconstruction were made using the even set of filters). In practice, the even and odd Log-Gabor filters are equivalent to the real and imaginary components of the Log-Gabor filters, as defined in Fourier-space. This gives a phase-independent measure of the response of any given point in a SF channel. The discrete sampling reflects the minimum number of RFs required by the visual system at each spatial scale, and as such the number of sampling points is proportional to the size of the RFs in each channel.

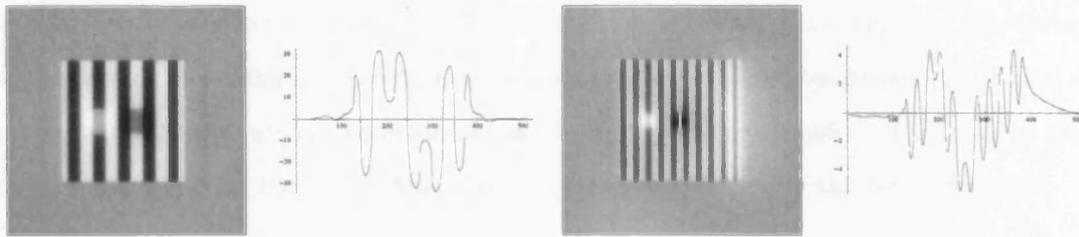


Figure 3.5: The raw NSS model output for the $0.5c/^\circ$ (left) and $1.0c/^\circ$ (right) White's effect stimuli, and the predicted brightness cross-sections at their mid-points. The output is in model-units, defined as the pixel-value increments in the reconstruction of the image.

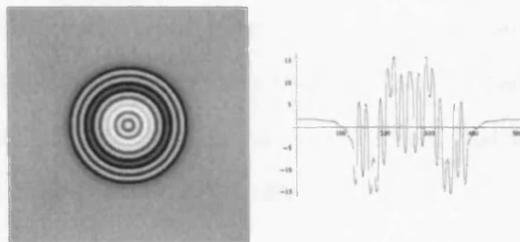


Figure 3.6: The NSS model output for the $1.6c/^\circ$ circular variant of White's effect, and the predicted brightness cross-section at the mid-point.

Results

The model correctly predicts the brightness for all variations on White's effect, as illustrated in Figures 3.5 and 3.6. For the $0.5c/^\circ$ canonical stimulus (Figure 3.5, left), the mean predicted brightness of the test patch on the left is 58% of the brightness of white, and 40% of the brightness of white for the test patch on the right, yielding an effect strength of 18%, compared to the empirically determined effect strength of 8.5%. For the $1.0c/^\circ$ stimulus (Figure 3.6, right), the mean predicted brightness of the test patch on the left is 65% of the brightness of white, and 30% of the brightness of white for the test patch on the right, yielding an effect strength of 35%, which is much greater in magnitude than the variant with an inducing grating at a lower SF, as observed psychophysically (White, 1981). The NSS model's predictions for the circular variant of White's effect is also in agreement with what is observed psychophysically (Howe, 2005), in that the brightness of the test ring between the dark bars is darker than the test ring between the light bars (with an

average predicted effect strength of 44% the brightness of white in the scene), as illustrated in Figure 3.6. One point of concern, however, is that the predicted brightness of the bars in the inducing grating are not constant across the reconstructed stimuli¹¹. This may be due to error derived from the $0.1c/^\circ$ filters, whose size exceeded that of the entire stimulus.

Figure 3.7 illustrates the output from each SF channel to the $0.5c/^\circ$ stimulus. It is clear that it is the lowest two SFs ($0.1c/^\circ$ and $0.2c/^\circ$), which generate White's effect, while all other channels predict a simultaneous brightness contrast effect¹². Similarly, Figure 3.8 illustrates the output from each set of filter banks at every orientation (combined across SF preference). It can be seen that all filters that are not oriented parallel to the inducing grating contribute to White's effect, while the set of filters that are in parallel contribute towards a simultaneous brightness contrast effect.

We can explain why this stimulus produces such channel outputs. The average luminance of the grating (without the grey test patches) is the same as the background. Thus, to the larger spatial filters the display is effectively uniform, and their response are negligible. Now consider what happens when a single grey test patch is added to a black bar, as illustrated in Figure 3.9. To a large spatial filter this is seen like a spike in luminance (or a delta function which is 'on' at the centre of the stimulus), and the resulting output is similar to that when convolving the filters with an image comprised of a genuine spike at the centre of a background with uniform luminance (yielding a map of the filters response profile). Figure 3.10 illustrates filter bank output for the stimulus in Figure 3.9, and it is clear to see this behaviour working for the larger filters, but disappearing when the size of the filters approaches that of the test patch. Similarly, if the test patch were to be introduced onto a white bar, then the same response, but opposite in polarity, will occur. Thus, by amplifying the output from low SF selective filters, this response is amplified above that of the high SF selective filters, yielding White's effect. This effect is not apparent when the filters are

¹¹a similar effect is also observed in Figure 3d of (Dakin & Bex, 2003), which illustrates the NSS model output for another version of White's effect.

¹²The same trend was found for the $1.0c/^\circ$ stimulus, but in this case it was the lowest three SF channels that demonstrated White's effect.

oriented at $\frac{\pi}{2}$ to the vertical (parallel to the grating), as at this orientation the response from the higher SF selective filters produces a simultaneous brightness contrast effect, which is far stronger than the output from low SFs.

It is possible to conclude that according to the NSS model, the amplification of low SFs are responsible for the perceived lightness of the test patches in White's effect.

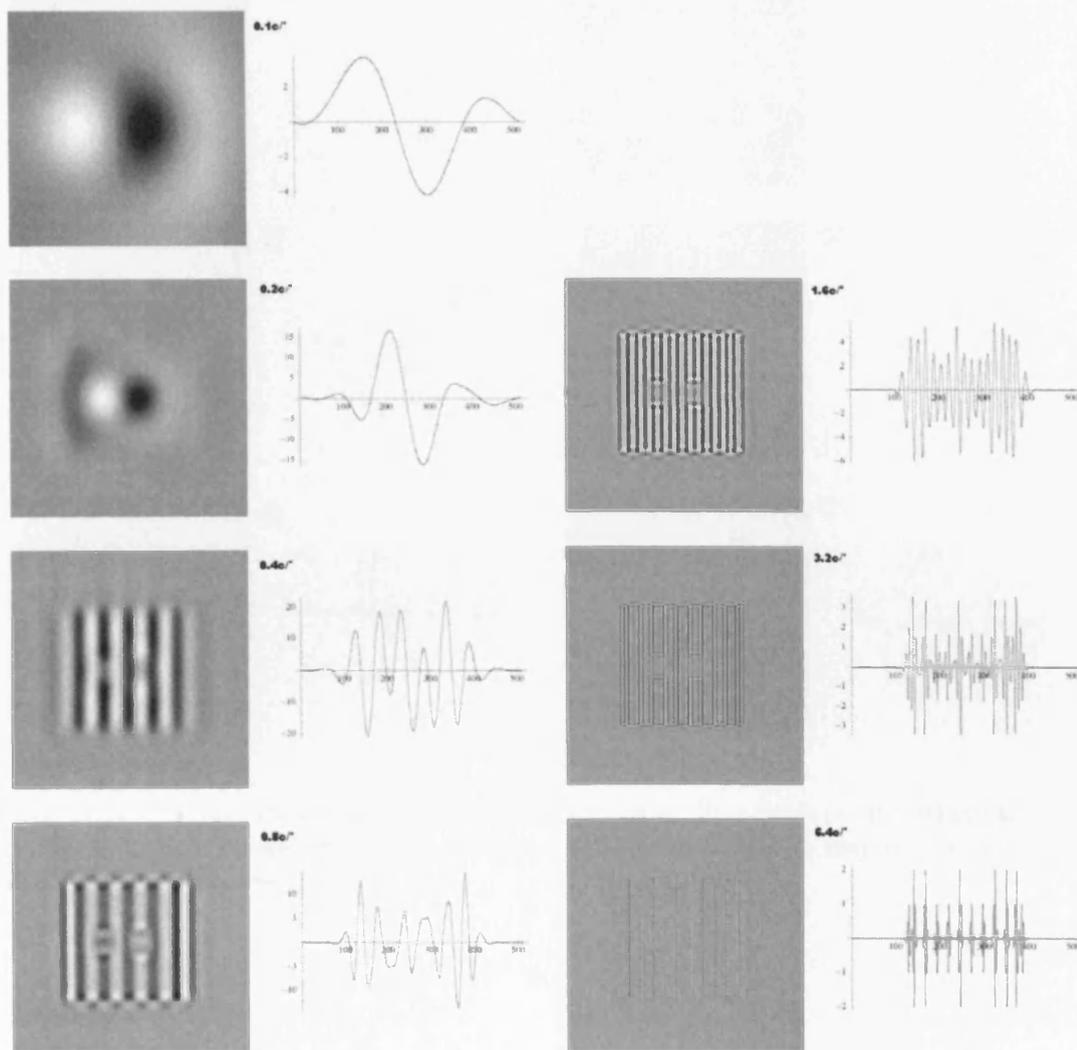


Figure 3.7: The output for each SF channel in the NSS model, with cross-sectional profile, in response to the $0.5c/^\circ$ White's effect stimulus.

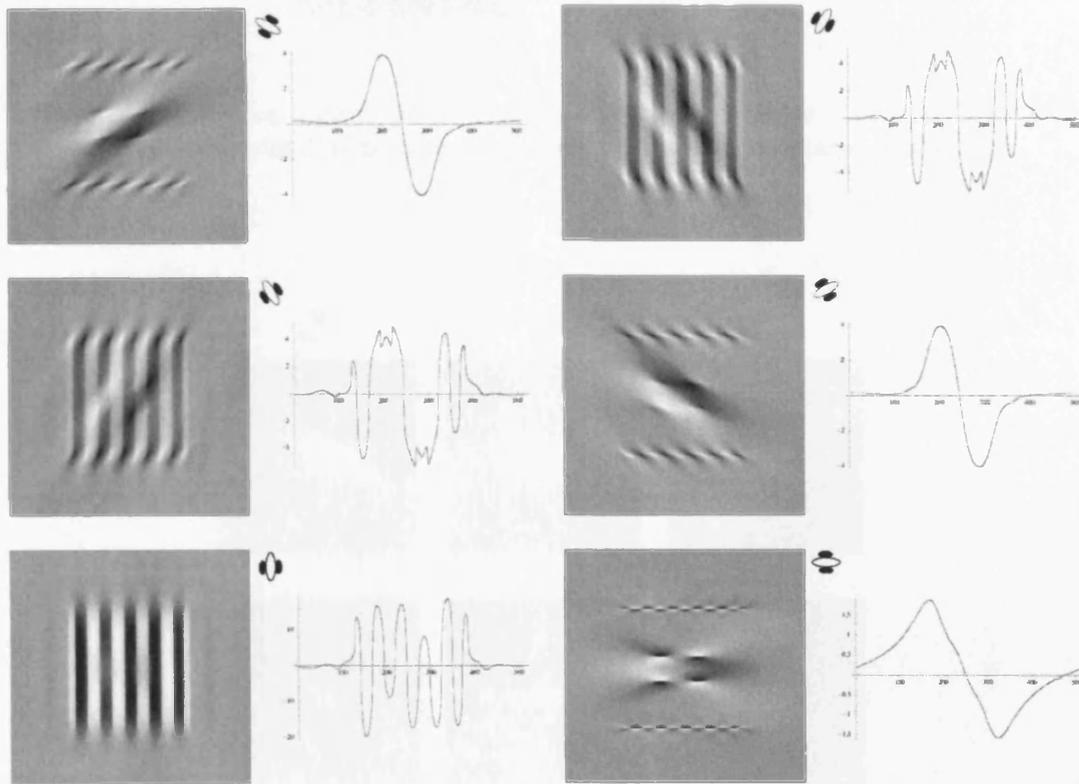


Figure 3.8: The output across SF for each orientation of filter bank (or hypothetical orientation channels), in the NSS model, with cross-sectional profile, in response to the $0.5c/^\circ$ White's effect stimulus.

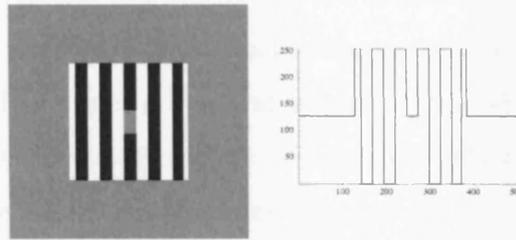


Figure 3.9: Introducing a grey test patch onto a grating is analogous to a spike (or delta function) on a background of uniform luminance, for filters with a large spatial scale.

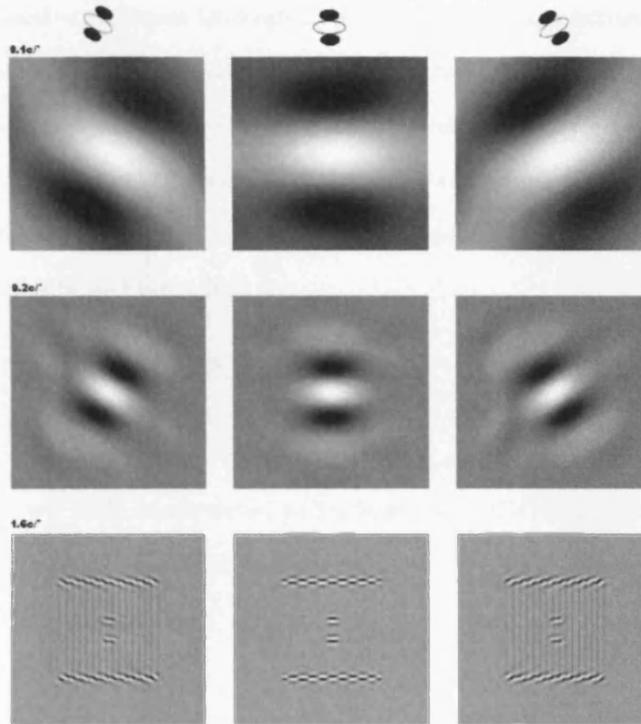


Figure 3.10: The response of $0.1c/^\circ$, $0.2c/^\circ$, and $1.6c/^\circ$ filters (at $-\frac{\pi}{3}$, 0 , and $\frac{\pi}{3}$ to the vertical) to the 'delta function' stimulus in Figure 3.8.

The Oriented Difference of Gaussians Model

The ODOG model of Blakeslee and McCourt (1999) is similar to the NSS model, in that it casts brightness as a problem of reconstructing the original image from the information contained within a set of filter bank outputs, at multiple scales and orientations. However, it differs in three fundamental ways. Firstly, V1 RFs are modelled using anisotropic difference-of-Gaussian (DoG) functions. Secondly, gains are applied individually to each filter bank selective for a particular SF, and are chosen to be in accordance with the suprathreshold detection level of gratings, which is at higher luminances for high SF gratings than lower SFs. This implies that the gains on the former channels are greater than that of the former (Georgeson & Sullivan, 1975), and this leads to SF dependent gains that are inversely proportional to those implemented in the NSS model. Thirdly, the filter banks are arranged into orientation selective channels through a weighted linear sum across SF selectivity, that are then normalised by their space-averaged RMS contrast. The latter non-linear processing step has the effect of weighting the stimulus features extracted by all orientation channels, such that high contrast features captured by a channel at one orientation will not overwhelm lower contrast features represented by channels at other orientations. The ODOG model is illustrated schematically in Figure 3.11.

The DoG Code

The details of the DoG code constructed to replicate the ODOG model are as follows: The DoG filters are defined in the spatial domain by,

$$DoG(x, y) = \frac{1}{2\pi} \left(\frac{1}{\sigma_c^2} \exp \left\{ -\frac{x^2 + y^2}{2\sigma_c^2} \right\} - \frac{1}{4\sigma_c^2} \exp \left\{ -\frac{x^2}{2\sigma_c^2} - \frac{y^2}{8\sigma_c^2} \right\} \right), \quad (3.8)$$

where σ_c represents the spatial constant for the two Gaussians used to construct the filter. With this configuration, the resultant filters are non-isotropic and oriented as a result of the variance in the y -direction in the second Gaussian, being twice that of its extent in the x -direction. An example of such a filter is illustrated in Figure 3.12.

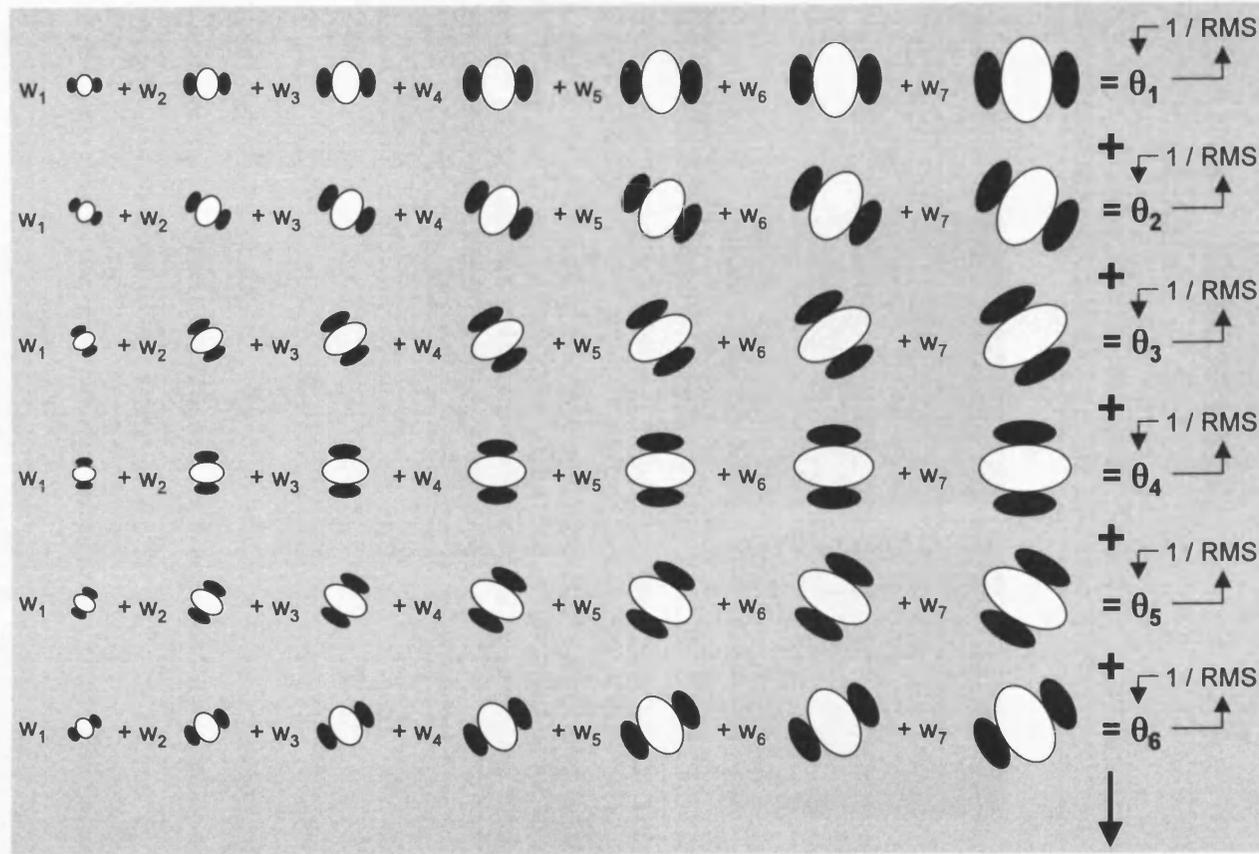


Figure 3.11: A schematic representation of the ODOG model: The output from banks of DoG filters are combined in a weighted linear sum across SF producing orientation channels (the θ 's), which are then re-normalised by their RMS output, before being integrated to produce the predicted brightness response.

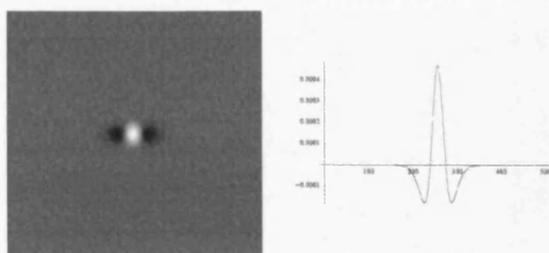


Figure 3.12: An example of an anisotropic DoG filter in the spatial domain (left), and a cross-section of the response profile at the mid-point (right).

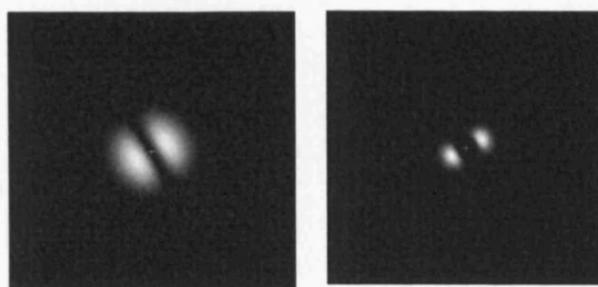


Figure 3.13: A DoG filter (left) and Log-Gabor filter (right) in the Fourier domain. Both filters are selective for SFs of $1.6c/^\circ$ oriented at an angle of $\frac{\pi}{6}$ to the horizontal. The cross-hair represents the origin of the Fourier domain. The DoG filter is far more distributed over SFs in the x - and y -directions, and hence is less selective for this orientation when compared to the Log-Gabor filter. Consequently, using these DoG filters at 6 orientations will over-represent orientation information at SFs.

The spatial constants are selected such that the peak response frequencies in the Fourier domain cover the range of $0.1c/^\circ$ - $6.4c/^\circ$ in octave intervals, in parallel with the NSS model. For each spatial scale there are filters at six orientations separated by an angle of $\frac{\pi}{6}$. The bandwidth of each filter is ~ 1.9 octaves for SF and $\sim \frac{\pi}{2}$ radians for orientation (both approximately double the corresponding values used with the Log-Gabor code).

Unlike the Log-Gabor code, the DoG code has not been constructed to be efficient in any way, and as a result of the larger SF bandwidth of the filters, lower SF are naturally over-represented (as illustrated in Figure 3.17, left). That is, lower SFs are amplified more than

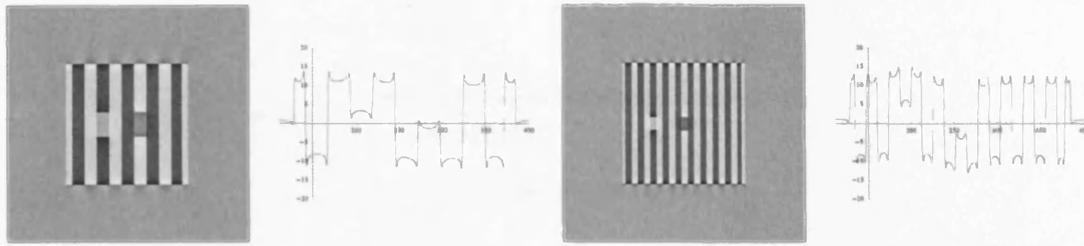


Figure 3.14: The raw ODOG model output for the $0.5c/^\circ$ White's effect stimulus (left), and the predicted brightness cross-section at the mid-point (right). The output is in model-units, defined as the pixel-value increments in the reconstruction of the image.

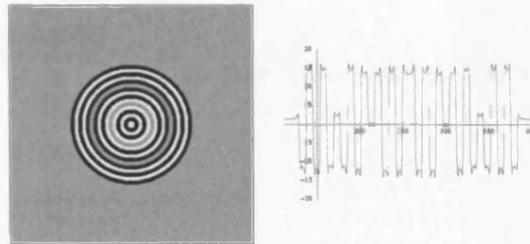


Figure 3.15: The ODOG model output for the $1.6c/^\circ$ circular variant of the White's effect stimulus (left), and the predicted brightness cross-section at the mid-point (right). The ODOG model clearly fails to provide an account of this effect.

higher SFs. Similarly, orientation information is over-represented at all SFs, due to the broad orientation tuning of these filters. These properties of the DoG code can be seen in Figure 3.13 where DoG filters are compared to the Log-Gabor filters in the Fourier domain.

Results

Figure 3.14 shows that the model correctly predicts the perceived lightness for canonical variants of Whites effect, for both $0.5c/^\circ$ and $1.0c/^\circ$ inducing gratings, and as claimed by the authors. More specifically, the effect strength for the $0.5c/^\circ$ is predicted to be 17% of the brightness of white (compared to the empirically derived effect strength of 8.5% determined by Blakeslee and McCourt (1999)), and the effect strength for the $1.0c/^\circ$ is predicted to be 39% of the brightness of white, consistent with data that demonstrates that the effect strength increases with the SF of the inducing grating (White, 1981; Blakeslee & McCourt,

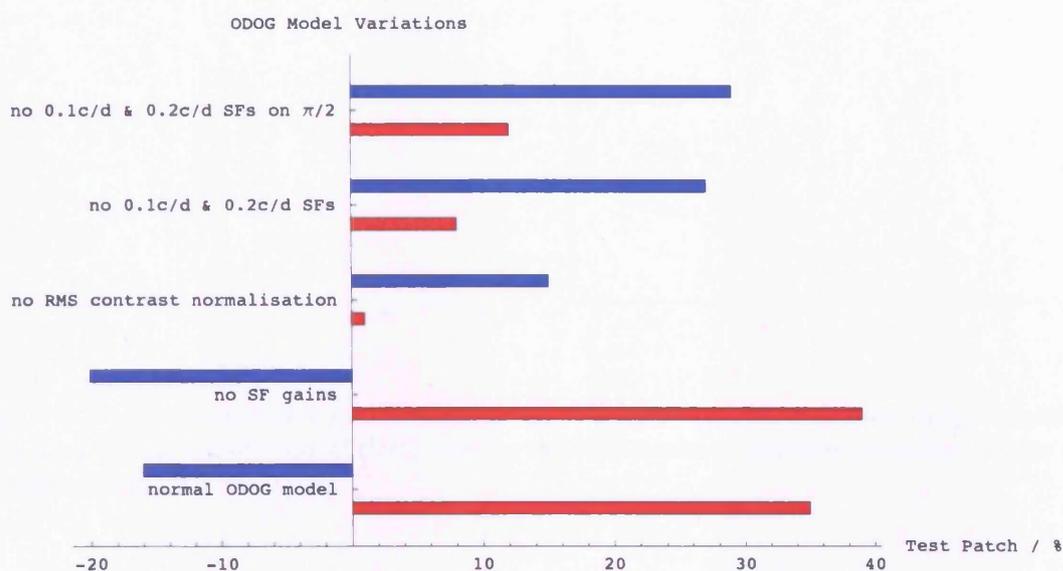


Figure 3.16: A summary of the predicted brightness of the test patches in the $0.5c/^\circ$ White's effect stimulus, as predicted by variations of the ODOG model, each with a different component of the model knocked-out. The bars represent the predicted brightness of the test patches as a percentage of the mean predicted brightness in the entire output (that is, 50% of the predicted brightness of white in the scene). Negative values correspond to values beneath the mean luminance of the output. The blue bars represent the values assigned to the test patch on the right hand side of the stimulus (on the white coaxial bar), and the red bars represent the values assigned to the test patch on the left hand side of the stimulus (on the black coaxial bar).

1999). However, Figure 3.15 shows that the model does not predict the existence of White's effect for the circular configuration, which goes against the psychophysical data collected with such stimuli (Howe, 2005).

The output of the model in response to the $0.5c/^\circ$ White's effect stimulus, but with key elements individually knocked-out, are summarised in Figure 3.16. Without the SF dependent gains a strengthening of Whites effect was observed. However, the output of the model without the RMS contrast normalisation resulted in a loss of Whites effect, and a weak simultaneous brightness contrast effect was observed instead. To examine the role of low SFs, the output of the model without the $0.1c/^\circ$ and $0.2c/^\circ$ filters was calculated. It was

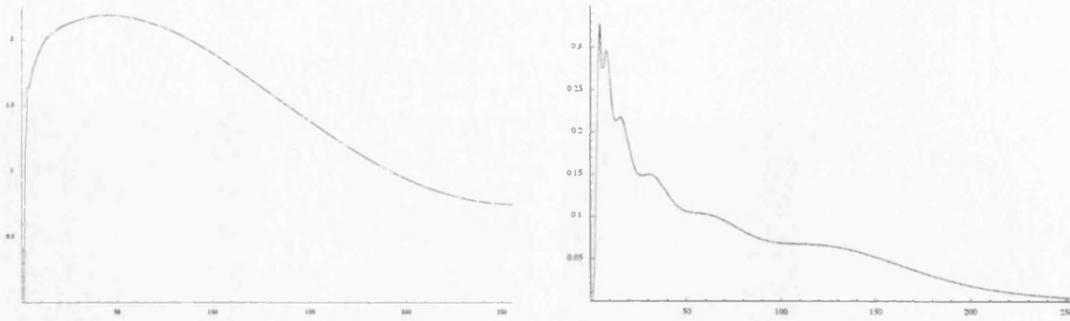


Figure 3.17: The cumulative gain applied to all SFs (measured in cycles per image for the stimulus), by the ODOG model (left), and for the NSS model in response to the $0.5c/^\circ$ White's stimulus (right).

found that White's effect was lost and once again replaced with a simultaneous brightness contrast effect. These results are in alignment with the operation of the NSS model, and are not surprising when one considers the cumulative gain over SF for the two models, as illustrated in Figure 3.17. Both models amplify low SFs relative to high SFs.

Figure 3.18 illustrates the output from each orientation channel, and it can be seen that only the channel oriented perpendicular to the inducing grating (at $\frac{\pi}{2}$ to the vertical) contributes to generating White's effect. It is also this channel that receives the greatest amplification from the divisive normalisation stage, as can be seen from the RMS contrast values for each channel as illustrated in Figure 3.19.

Based on the above results, the effect of knocking out the $0.1c/^\circ$ and $0.2c/^\circ$ filters in only the $\frac{\pi}{2}$ channel was examined, and found to be enough to convert White's effect to a simultaneous brightness contrast effect¹³. Thus, it is possible to conclude that it is the low SFs in the $\frac{\pi}{2}$ oriented channel that generate the effect, as they 'fill-in' the simultaneous brightness contrast effect arising from the boundaries of the coaxial bar and the test patch (and hence this account is analogous to the Todorovic (1997) account, which uses T-junctions to discount the simultaneous brightness contrast arising from the boundaries of the test patch

¹³for the $1.0c/^\circ$ White's stimulus it required the $0.1c/^\circ$, $0.2c/^\circ$, and $0.4c/^\circ$ filters to be knocked-out for simultaneous brightness contrast to be generated.

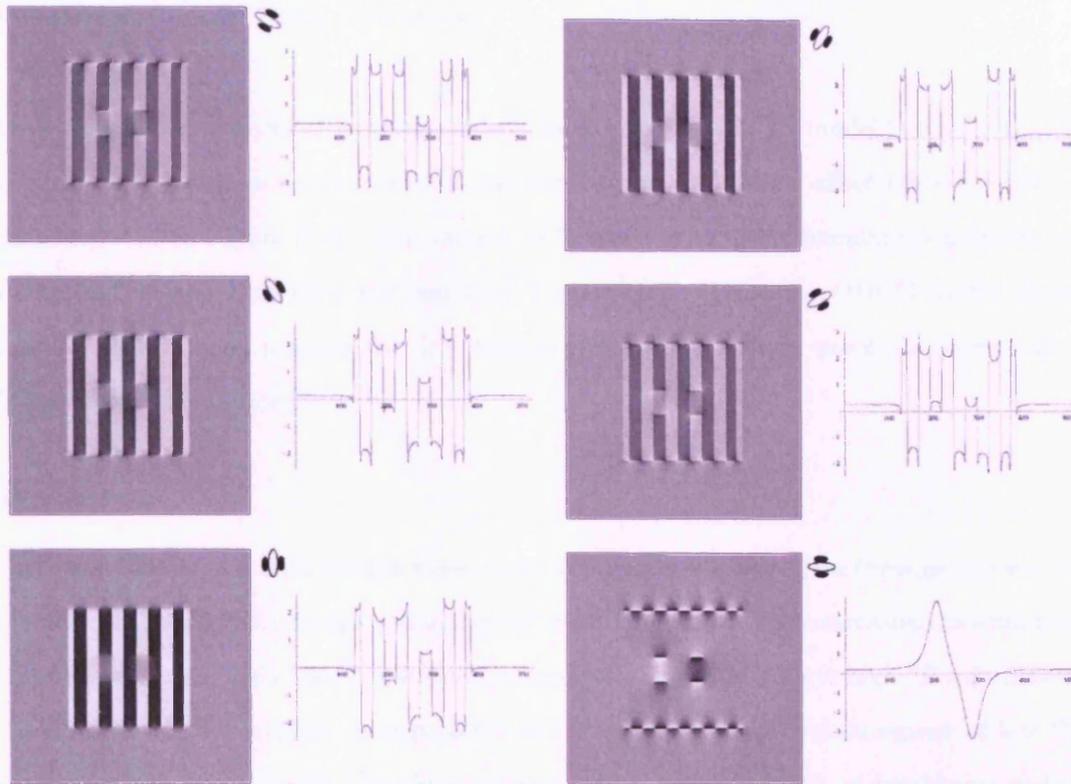


Figure 3.18: The output across SF for each orientation channel in the ODOG model, with cross-sectional profile, in response to the $0.5c/^\circ$ White's effect stimulus.

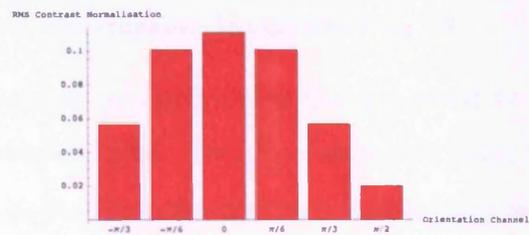


Figure 3.19: The ODOG RMS contrast normalisation values for the $0.5c/^\circ$ White's stimulus.

and the adjacent bars). However, as the output from this channel is relatively small, it receives a natural amplification through the renormalisation procedure, and subsequently this channel's output comes to dominate.

Given the above analysis, it becomes clearer as to why the ODOG model is not be capable of providing the same account for the circular variant of White's effect (Howe, 2005), as demonstrated in Figure 3.15 - this variant of White's effect lacks boundaries between test patch and coaxial bars (and the resulting T-junctions). Thus, the ODOG model clearly fails at providing an account for this version of White's effect in much the same way as T-junction analysis does.

Discussion

Both the ODOG and NSS models were shown to predict White's effect through the use of a similar computational principle, regardless of the difference in their constituent mechanisms. That is, in both models low SF information receives a boost relative to high SF information. As a result of this analysis, it is possible to infer that it is this enhancement of low SFs that generates White's effect in the context of these neural models of brightness coding. However, the exact nature and purpose of low SF representation in V1 is a controversial issue, and the biological plausibility of the low SF amplification mechanisms in the ODOG and NSS models is a moot point, for reasons to be discussed.

The Spatial Extent of Information Integration in V1

The existence in V1 of simple cells with Gabor-like RFs tuned for the low SFs required by these models, is a controversial issue. The RFs of these cells would have to span a very large extent of the visual field, requiring the existence of long-range intra-cortical connections. Figure 3.21 demonstrates this by plotting the normalised response profile for log-Gabor RFs tuned to $0.1c/^\circ$ and $0.2c/^\circ$, as used in the NSS model - the 1024×1024 pixel plots represent $40^\circ \times 40^\circ$ of visual angle, and it is clear to see that they will respond to stimulation as far away as 20° and 10° from their respective centers. However, recent physiological studies have measured the spatial extent of V1 RFs in Macaque and found the mean radius of the

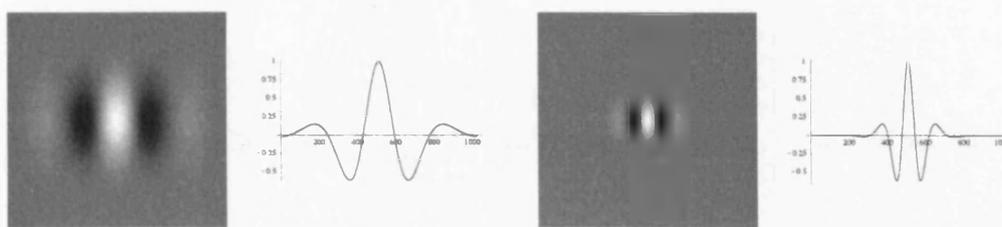


Figure 3.20: Log-Gabor RFs tuned to $0.1c/^\circ$ (left) and $0.2c/^\circ$ (right). Their peak responses have been normalised to 1.

classical RF to be between 0.4° and 1.5° (depending on eccentricity within the parafoveal representation), with some radii as large as 3° (Sceniak, Hawken, & Shapley, 2001; Cavanaugh, Wyeth, & Movshon, 2002). Both of these studies aimed to map the entire spatial extent of V1 receptive fields by measuring a cell's response to circular patches of drifting grating, of various diameters, at the optimum SF and orientation - a technique regarded as more sensitive than the traditional method of determining the minimum response field (found by these authors to lead to RF size estimates half as large).

In addition to measuring the extent of the classical receptive field (CRF), the extent of the suppressive surround (SS) was also determined in both these studies. This latter part of a cell's RF, located outside and adjacent to the CRF, has been shown to modulate the initial response of the CRF, although it is unresponsive to independent stimulation (Allman, Miezin, & McGuinness, 1985; DeAngelis, Freeman, & Ohzawa, 1994; Levitt & Lund, 1997). Hence, a single cell can respond to information over a larger area than the limits of the CRF would imply; the mean radius of the SS was found to be between 1.1° and 3.5° (depending on eccentricity within the parafoveal representation), with some radii as large as 5° . Although both the ODOG and NSS models do not explicitly model the effect of the SS, some important results from Sceniak et al. (2001) and Cavanaugh et al. (2002) will be summarized to add context to the discussion of information integration and the size of RFs in V1.

A parsimonious, yet competent account of the interaction between the CRF and SS can be

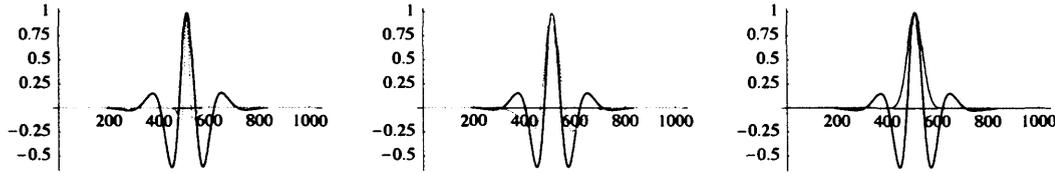


Figure 3.21: Concurrent plots of a $0.2c/^\circ$ tuned log-Gabor RF profile (black lines) against the profiles of the DoG model of CRF and SS interaction, with various empirically derived parameters: the ‘mean cell’ of the population (left); the largest possible cell with both CRF and SS (centre); and the largest possible cell with CRF only (right).

achieved by using a difference of Gaussians function to model the combined RF of cells in V1 (DeAngelis et al., 1994; Sceniak et al., 2001). In this framework the response of the CRF is modeled by a positive Gaussian with spatial extent a , and gain K_a , and the SS is represented by a negative Gaussian centered on the CRF, with spatial extent b , and gain K_b . The overall response is the sum of the two individual responses. The underlying RF is expressed formally as,

$$DoG(r) = K_a \exp \left\{ - \left(\frac{2r}{a} \right)^2 \right\} - K_b \exp \left\{ - \left(\frac{2r}{b} \right)^2 \right\}. \quad (3.9)$$

Both Sceniak et al. (2001) and Cavanaugh et al. (2002) successfully fitted this function to the responses of a large population of V1 cells (138 and 315 cells respectively), with the mean error under 10%. The various RFs determined by the parameters in the model were characterized by defining the suppression index (SI) as,

$$SI = \frac{K_b b}{K_a a}, \quad (3.10)$$

where SI was found to lie in the range 0-1: An SI of 1 implies that the strength of the SS equals that of the excitatory CRF, and a SI of 0 that the strength of the SS is negligible relative to that of the excitatory CRF. For the population of cells used by Sceniak et al. (2001), the mean SI was found to be 0.62 with the mean value of a equal to 1.0° (some cells were found with values as large as 3.3°), and the mean value of b equal to 2.2° (some cells were found with values as large as 8.0°). These measurements are consistent with those

found independently by Cavanaugh et al. (2002). From this information it is possible to establish that the mean value of $K_b = 0.28K_a$. The 1-D spatial response profile for such a ‘mean cell’ is plotted in Figure 3.21 (left) concurrently with a $0.2c/^\circ$ tuned log-Gabor RF, for comparison. Similarly, the largest possible CRF extent in this paradigm is constructed by assuming that $a = 3.3^\circ$ and $K_b = 0$ (when $SI = 0$), and the largest possible RF with both CRF and SS is constructed by assuming $a = 3.3^\circ$, $b = 8.0^\circ$ (when $SI = 1$). These RF response profiles are also plotted concurrently with a $0.2c/^\circ$ tuned log-Gabor RF in Figures 3.21 (right) and 3.21 (centre), respectively. It can be seen that the $0.2c/^\circ$ tuned log-Gabor RF has systematically larger spatial spread than the empirically motivated DoG model of CRF and SS interaction. If these results are taken as an indication of RF sizes and the extent of spatial summation in V1, then it is possible to exclude the $0.1c/^\circ$ Gabor-like RFs from being biologically plausible, and it becomes hard to justify the existence of RFs tuned to $0.2c/^\circ$ without involving the SS.

These results are given additional weight from recent studies that have measured the physical extent of horizontal and lateral intra-cortical connections in V1. Angelucci et al. (2002) repeated the above single-unit recording experiments with circular patches of drifting grating as stimuli, fitted the responses to the DoG model of CRF-SS interaction to derive the spatial parameters (in equation 3.9), and then by anatomical staining and tracing, the actual physical spread of intra-cortical connections from cells in the same locations were derived (assuming normality within the population of cells), and compared. The sample mean of the intra-cortical connection distances was estimated to be $1.24^\circ \pm 0.15^\circ$ (for the area of the cortex representing 2° - 8° of eccentricity in visual space), and was found to be significantly correlated with the spatial spread of the CRF (a in equation 3.9, and measured to be 1.2°), but not correlated to the spatial spread of the SS (measured to be 5.4°). However, the extent of the SS was found to be consistent with the regions of visual space conveyed by feedback connections from the extrastriate cortex. These results are also supported by Stettler et al. (2002), who used an alternative method of anatomical staining and tracing to identify the physical spread of intra-cortical connections and the extent of feedback connections from the extrastriate cortex.

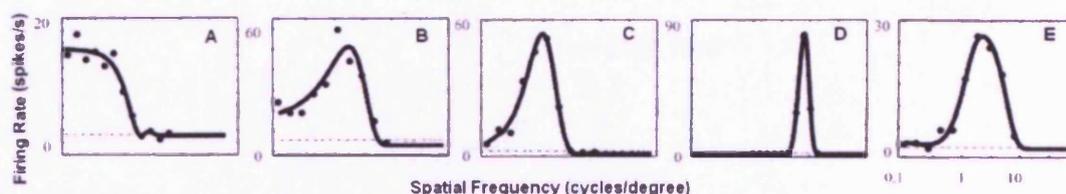


Figure 3.22: Examples of SF tuning curves found in a sample of V1 simple cells, reproduced from Xing, Shapley, & Hawken (2004).

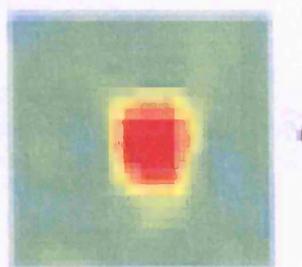


Figure 3.23: An example of a 'blob' type RF in V1. The closer the colour of the pixels in the image are to red, then the greater the response to luminance in that region of the RF. Reproduced from Ringach (2001).

Given these suggested limits on the distances that information can be communicated between neurons in V1, a picture emerges in which simple cells with multi-region Gabor-like RFs cannot be responsible for representing low SFs in V1. Furthermore, any global analysis and renormalisation in V1 is also constrained by these limits to be relatively local in extent. Both the ODOG and NSS models employ global re-normalisation processes that are intrinsically non-local. Thus, it is hard to think of the ODOG and NSS models as being represented by the biology of V1.

The Coding of Luminance in V1

To code low SFs, information need not be integrated from across large extents of the visual field, and cells which respond to luminance and low SFs are often found in the cytochrome oxidase (CO) regions of V1 (Silverman, Grosf, De Valois, & Elfar, 1989). Recent studies

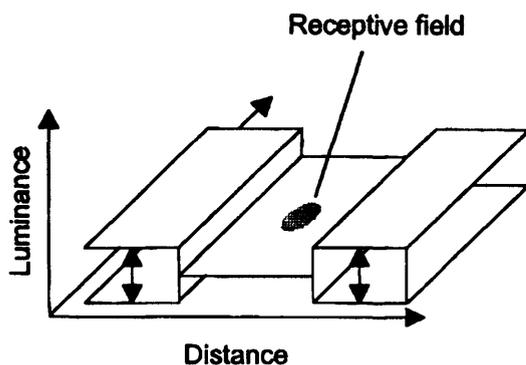


Figure 3.24: An example of the stimulus configuration used to test the response of a cortical neuron to a temporally modulated surrounding luminance level. The receptive field represents the CRF and SS. Reproduced from Rossi & Paradiso (1999).

utilizing reverse-correlation techniques for mapping RFs have identified cells in V1 that are low-pass in SF (Xing, Ringach, Shapley, & Hawken, 2004). Figure 3.22 illustrates five example tuning curves representative of the different types of cells in V1, as found by Xing et al. (2004). The tuning curve in Figure 3.22a clearly shows a cell with a low-pass response to SF. Furthermore, it was found that this cell was poorly tuned to orientation. These details are consistent with the ‘blob’ like RFs found in an earlier reverse-correlation study by Ringach (2001), an example of which is illustrated in Figure 3.23. Figure 3.22b illustrates a cell that is tuned to SF ($> 1c/^\circ$), but which also has a response to very low SF and that was found to be sharply tuned for orientation. It should be noted that cells with these characteristics have also been found in the optic tract and LGN, as well as V1 (Rossi & Paradiso, 1999). Figures 3.22c-e show SF tuning curves for cells that are highly tuned for both SF and orientation, but which do not have a large responses to low SFs, and are characteristic of purely contrast responsive cells. Thus, it appears that both luminance *and* contrast information is present in V1.

However, the role that these cells play in the representation of surface luminance remains unclear. In V1, and to a lesser extent in LGN, a set of luminance sensitive cells have been found with responses that can be influenced by the presence of a contrasting surround with

a temporally modulated luminance level (Rossi & Paradiso, 1999; Kinoshita & Komatsu, 2001). An example of such a stimulus configuration is illustrated in Figure 3.26. The contrasting surround is arranged to be around the surface within which the cell's CRF and SS reside during stimulation. Hence, the cell's response should be independent from the luminance of the surround. Perceptually, the stimulus should give rise to a temporally modulated SBC effect (this configuration is equivalent to one half of the canonical SBC illusion, as depicted in Figure 1.1, left, with the surround luminance temporally modulated). What has been found in such experiments, is that there exists a group of such cells whose responses are consistent with the perceived brightness in these stimulus configurations, and a group whose responses correlate with the absolute magnitude of the test patch's luminance. Similarly, Roe, Lu, and Hung (2005) have found neurons in the CO 'thin stripe' regions of V2, whose responses are consistent with the apparent brightness of the Craik-Cornsweet-O'Brien illusion (see Chapter 5 for a detailed discussion, and Figure 5.43 for an example of this phenomenon). Thus it appears that there exist cells which code actual luminance, and cells which represent luminance. A theory underpinning the existence of both will be outlined in the next section.

Low Spatial Frequencies and the Fourier Representation of Images in V1

The computational principle motivating the incorporation of spatially-large, low SF tuned Gabor-like filters into the ODOG and NSS models, is that without them there would be no 'filling-in' of the surfaces bound by edges (which would be signaled by cells tuned to higher SFs), and hence the reconstruction of the retinal image would be incomplete. Such filter-derived models are performing an image reconstruction based on something that resembles a local-Fourier analysis of the original image; the filters provide a measure of the terms in a Fourier series representation of a local region of the image. However, the local mean (or D.C) level is not adequately represented by these filters, and in essence the low SF filters are compensating for this deficit.

In theory, the ODOG and NSS models could be modified to incorporate the filter output from cells that are low-pass in SF, possibly with blob-like receptive fields replacing the low

SF cells that are large in spatial extent (the blobs need not be large to sample low SF and D.C.). In practice, this would introduce an additional low-pass SF channel in place of those tuned to $0.1c/^\circ$ and $0.2c/^\circ$, but as this is essentially providing a measure of the local D.C. component it is not clear how any meaningful SF or orientation dependent gains could be applied, if at all, and hence whether White's effect would emerge. Essentially, it would not naturally fit into the current incarnations of the ODOG and NSS models.

An alternative view is presented, based on the properties of the luminance responsive cells discussed thus far. It is conjectured that the cells which appear to consistently respond to absolute surface luminance, are those which appear to leak local D.C information in their response (Figure 3.22b), and that they do not play an important role in surface representation, as their response to a surface would be hard to disambiguate from that to a sub-optimal grating; that is, their leaky behaviour is not part of any significant functional role. However, it is believed that the blob-like cells *are* involved in surface representation, but that they are not part of an explicit feed-forward luminance code - if arrays of such cells did exist, and the retinal image is explicitly represented in the cortex, then there would be little reason for the existence of contrast coding cells upstream from V1.

This view is supported by work which proposes that contrast coding cells exist as a direct result of the optic tract having a limited channel capacity, which is physically incapable of transmitting the required amount of luminance information for an explicit representation of the retinal image (Barlow, 1961, 1981). The amount of information which can be transmitted through the optic nerve has been estimated to be two orders of magnitude less than the amount of information arriving at the retina (Kelly, 1962; Nirenberg, Carcieri, Jacobs, & Latham, 2001; Z. Li, 2006). By coding contrast, however, the necessary information is efficiently coded (Buchsbaum & Gottschalk, 1983; Z. Li, 2006). Further motivation for the existence of contrast coding cells is that they make for good 'feature detectors', responding well to bars and edges, the fundamental constituents of images (Marr, 1982). That is, they are optimised for performing image processing tasks.

Instead, it is proposed here, that the blob-like cells represent the luminance as inferred from the responses of contrast coding cells, and that their resultant output represents inferred surface luminance. In the context of the experiments described above, this theory is implicitly asserting that the cells found to responsive to absolute surface luminance are those that leak D.C., and those which respond in accordance with perceived brightness are the V1 blobs. This view forms the basis for an alternative model of luminance coding developed in the next chapter.

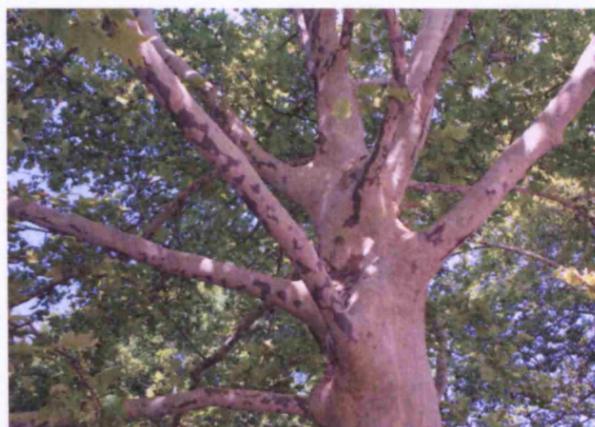


Figure 4.1: An example of a natural scene with information present at multiple scales.

Chapter 4: The Implicit Coding of Luminance in V1

In this chapter, an alternative approach to modelling the functionality of V1 is presented. The model deviates from the local-Fourier analysis techniques deployed by the NSS and ODOG models, and consequently large spatial filters tuned to low SFs are not required. Instead, the problem of recovering the local mean (or D.C.) level of luminance from the output of contrast-coding filters (tuned to the SFs typically found in V1), is cast mathematically in the Scale-Space paradigm, and then solved. The model implicitly codes the luminance of a given surface, and as such is referred to as the Implicit Luminance Coding (ILC) model.

Scale-Space, Local Jets, and the Known Functionality of V1

The Scale-Space theory of image processing has been developed by a segment of the computer vision community, with the aim of providing a multi-scale representation for images (Koenderink, 1984; Lindeberg, 1996; Haar Romeny, 2003). The need to describe an image at multiple scales arises from the fact that some information is more relevant at smaller scales, and some at larger scales: for example, discussing the concept of a solar system on the scale of individual particles yields little meaningful information, and it is more relevant to think about it at the scale of individual planets. Similarly, in the natural scene depicted in Figure 4.1, there is meaningful information at the scale of individual leaves as well as at

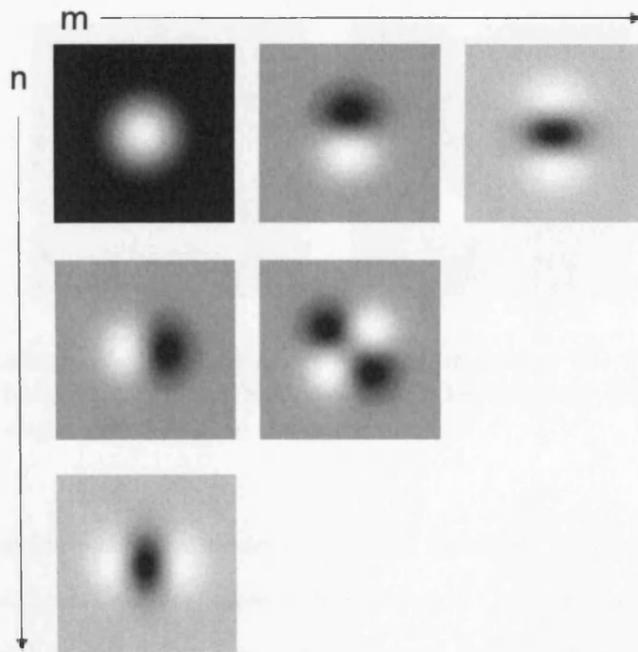


Figure 4.2: An example of a set of derivative of Gaussian filters, up to second-order in x - and y -differentiation.

the scale of the tree's trunk. When analysing this image, choosing one scale for describing both sets of objects will result in a loss of information relative to independent descriptions at both the scale of leaves and tree's trunk.

One crucial requirement for such a multi-scale image representation, is that image features at coarse-scales should constitute simplifications of corresponding structures at finer-scales; they should not be accidental phenomena created by the method for suppressing fine-scale structures. Scale-Space theory achieves such a multi-scale representation by convolution of the image with a Gaussian filter and its derivatives, as given by,

$$g_{n,m}(x, y; \sigma) = \frac{\partial^n}{\partial x^n} \frac{\partial^m}{\partial y^m} \frac{1}{2\pi\sigma^2} \exp\left\{\frac{-(x^2 + y^2)}{2\sigma^2}\right\}, \quad (4.1)$$

where the notion of scale is represented by the standard-deviation, σ , of the Gaussian function. Figure 4.2 depicts a set of such Gaussian Derivative (GD) filters up to second order

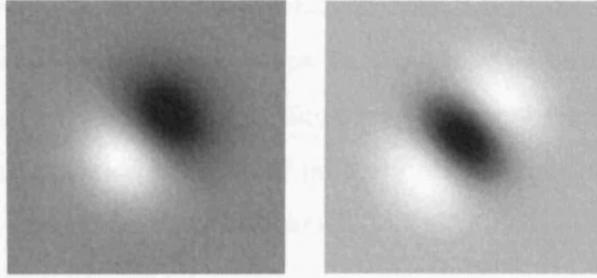


Figure 4.3: An example of first (left) and second order (right) GD filters oriented to an angle of $\frac{\pi}{4}$ to the horizontal. These filters are created by ‘steering’ GD filters of the same order, oriented to angles of 0 and $\frac{\pi}{2}$ to the horizontal.

in x - and y -differentiation. As a model for the RF structure of V1 simple cells, GD filters have proven a successful alternative to the DoG and Gabor models, and accurately fit empirically-derived RFs with fewer parameters (Young, Lesperance, & Weson Meyer, 2001; Young & Lesperance, 2001).

The Representation of Orientation Information

One of the interesting properties of the GD filters, is that it is possible to construct GD filters oriented (or ‘steered’) to an angle away from the horizontal and vertical, through a weighted linear combination of filters oriented to the horizontal and vertical (Freeman & Adelson, 1991). First and second order oriented filters are constructed as follows,

$$g_1^\theta = g_{1,0} \cos \theta + g_{0,1} \sin \theta \quad (4.2)$$

$$g_2^\theta = g_{2,0} \cos^2 \theta - 2g_{1,1} \cos \theta \sin \theta + g_{0,2} \sin^2 \theta, \quad (4.3)$$

where g_1^θ and g_2^θ represent first- and second-order filters oriented at angle θ to the horizontal. Figure 4.3 illustrates examples of $g_1^{\frac{\pi}{4}}$ and $g_2^{\frac{\pi}{4}}$ filters. An advantage of steering filters, is that fewer filters are required to represent information at all orientations; only a subset of the set of filters tuned to all orientations, are required to represent all orientation information.

Furthermore, the second-order $g_{1,1}$ filter, whose structure is harder to reconcile with the known RF properties of cells in V1, can now be thought of as being produced from the output of g_2^0 and $g_{2,0}$ (by re-arrangement of Equation 4.3), whose structures are common to V1. This latter point is critical for another property of GD filters, which make extensive use of the $g_{1,1}$ filter, and will be described shortly.

The Local-Jet: From Filter Output to Image

In accord with the ODOG and NSS models, which use a weighted linear sum of filter outputs to reconstruct the original image, the output from GD filters can be combined to reconstruct the input image. In Scale-Space, this proceeds by exploiting the fact that the convolution of an image with a GD filter is equivalent to using the appropriate derivative operator, on the output of the original image convoluted with a pure Gaussian filter. That is,

$$\begin{aligned} I(x, y) \otimes g_{n,m}(x, y; \sigma) &= \frac{\partial^n}{\partial x^n} \frac{\partial^m}{\partial y^m} [I(x, y) \otimes g_{0,0}(x, y; \sigma)] \\ &= \frac{\partial^n}{\partial x^n} \frac{\partial^m}{\partial y^m} \bar{I}(x, y; \sigma), \end{aligned}$$

where $I(x, y)$ is the original image and $\bar{I}(x, y; \sigma)$ is a blurred version of the image (through convolution with a Gaussian filter of scale σ). Given the above relation, it is then possible to construct a 2-D Taylor expansion at a point (x, y) on the blurred image (using the output of GD filters at that point), which can then be used to approximate the blurred image at a point $(x + \Delta x, y + \Delta y)$ in the neighbourhood around the original sampling point, where,

$$\bar{I}(x + \Delta x, y + \Delta y; \sigma) = \sum_{\alpha=0}^{\infty} \sum_{\beta=0}^{\infty} \frac{(\Delta x)^\alpha (\Delta y)^\beta}{\alpha! \beta!} \left[\frac{\partial^\alpha}{\partial x'^\alpha} \frac{\partial^\beta}{\partial y'^\beta} \bar{I}(x', y'; \sigma) \right]_{x'=x, y'=y},$$

and is the definition of a local-jet (Florack, Haar Romeny, Viergever, & Koenderink, 1996). Truncating this series expansion of the blurred image to include only terms up to second-

order, yields an approximation that can be computed using the filters in Figure 4.2, which are typical of the RFs of cells in V1,

$$\begin{aligned} \bar{I}(x + \Delta x, y + \Delta y; \sigma) \simeq & g_{0,0} + (\Delta x)g_{1,0} + (\Delta y)g_{0,1} \\ & + \frac{1}{2}(\Delta x)^2 g_{2,0} + (\Delta x)(\Delta y)g_{1,1} + \frac{1}{2}(\Delta y)^2 g_{0,2}. \end{aligned} \quad (4.4)$$

The first term in (4.4), $g_{0,0}$, represents a measure of the mean luminance in the region about the sampling point, and plays an analogous role to the large spatial filters of the NSS and ODOG models, which attempt to sample the local mean (or D.C.) component of a local-Fourier representation of the image. However, as the $g_{0,0}$ filters respond to regions of uniform luminance (i.e. 0 c/°), they do not have to be large to represent low SF information adequately. In fact, there is no size constraint at all, and the filters fulfil their purpose at every scale of interest.

It should be noted, that in terms of neural processing, the local-jet is equivalent to the weighted linear sum of filter outputs, where the weights (or gains) are determined by the coefficients of the 2-D Taylor expansion. Such a calculation is easily executed by a neuron which integrates the information from the output of cells with the necessary RF properties, with varying degrees of synaptic connection strength representing the required gains. This process would proceed in the same way as that for calculating the outputs of V1 cells from the weighted linear sum of LGN cell outputs (Dayan & Abbott, 2001). Thus, in this context the way in which the local-jet reconstructs the image is analogous to the way in which the image is reconstructed by the ODOG and NSS models.

Local Jets and the Role of ‘Blob’ Cells in V1

Chapter 3 concluded by proposing that the responses of cells with the single-lobed or ‘blob’ like RF structures that have been found in V1, are consistent with that of inferred luminance in a representation of surfaces. That is, that the responses of such cells are not derived from



Figure 4.4: The test image for the ILC model, referred to as “Lena”. The image is 256×256 pixels large and has a $1/f^\alpha$ distribution of power as a function of SF (with $\alpha = -0.04$), in keeping with all images classified as natural scenes.

a feed-forward luminance code from the retina to V1, but are inferred from the responses of contrast coding cells. Formulating these ideas in the context of Scale-Space, the responses of the $g_{0,0}$ filters required to reconstruct the original image using local-jets, must be derived from the outputs of higher-order GD filters, in character with the way that the $g_{1,1}$ filters have no clear analog in V1, but can be recovered from the information in other filters that do (by means of steering filters).

Given the above discussion, a new model of low-level brightness is proposed, based on the following ideas:

- An approximation to the retinal image is reconstructed using the outputs of contrast coding cells of the type most commonly found in V1. These are modelled using the first- and second-order GD filters of Scale-Space;
- the cells in V1 need not sample the retinal image continuously. RFs need only overlap to the extent that information is neither obviously lost or redundant, in keeping with the ideas on coding efficiency described by Field (1987);
- a blurred version of the retinal image is used as the desired approximation to the retinal image. This is reconstructed in-between the sampling points of GD filters by using an approximation of the second-order local jet (*without* the $g_{0,0}$ component), to ‘fill-in’ local brightness values in square regions around each sampling point. Each square region is referred to as a Local Brightness Tile (LBT);

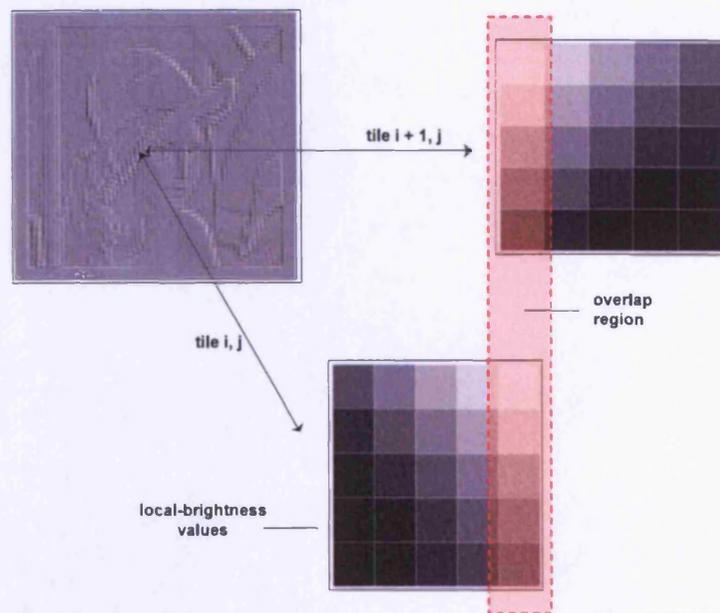


Figure 4.5: An example of Lena reconstructed using local jets (with a scale of $\sigma = 2$ pixels), arranged in a lattice structure, and two adjacent local brightness regions (or tiles). The first- and second order GDs are sampled at the centre of each LBT. The two tiles overlap (represent mutual information) in the shaded region.

- the $g_{0,0}$ mean-luminance component associated with each LBT is either inferred or recovered through global analysis of the image. This value is then added to every point on the LBT. The resulting set of 'shifted' LBTs yield the final brightness image. Thus, low SF information in the image is reconstructed as opposed to explicitly sampled. The set of $g_{0,0}$ components constitute the global luminance structure of the image;
- the image is reconstructed independently at multiple scales, which are subsequently combined in an un-weighted linear sum to yield the final predicted brightness image.

Recovering Global Luminance Structure at a Single Scale: An Analytical Solution

Figure 4.5 depicts a reconstruction of the test image depicted in Figure 4.4, performed using local jets without the $g_{0,0}$ component (and with a scale of $\sigma = 2$ pixels) to construct LBTs, which are arranged in a lattice structure. Also shown are two adjacent LBTs. The first- and second order GDs used in each local jet are sampled at the centre of each LBT. It is

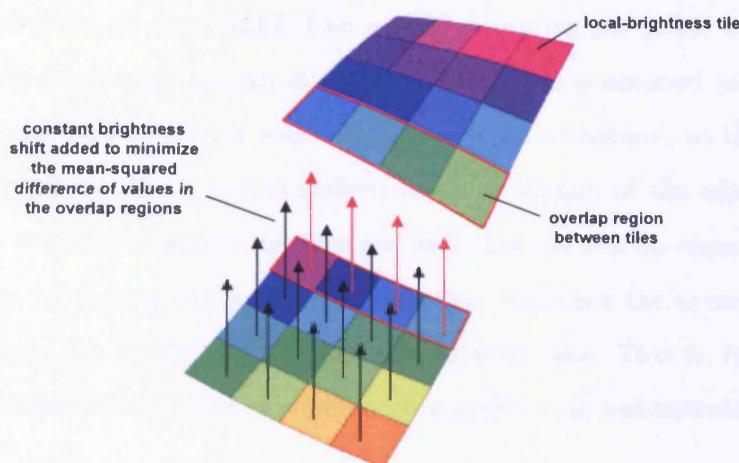


Figure 4.6: An example of how two LBTs are brought into alignment, implicitly coding the mean luminance level of the local jets.

clear that the resulting output contains nothing more than a representation of the lines and edges in the original image (at a scale of $\sigma = 2$ pixels). Formally, the relationship between tile (i, j) constructed *without* the zero-order component in the local jet, $\hat{I}_{i,j}$, and the desired tile constructed *with* the zero-order component in the local jet, $\bar{I}_{i,j}$, is given by,

$$\bar{I}_{i,j} = \hat{I}_{i,j} + z_{i,j}, \quad (4.5)$$

where $z_{i,j}$ is a constant, and $z_{i,j} = g_{0,0}$ is the correct solution to (4.5). The set of $z_{i,j}$'s can be correctly recovered from the information held just within the LBTs, the $\hat{I}_{i,j}$'s. Consider the blurred version of the image, \bar{I} , as a continuous function on \mathbb{R}^2 (put aside the fact that it is intrinsically discrete and composed of pixels). When \bar{I} is reconstructed using local jets without the zero-order term, discontinuities are introduced into the function, at the boundaries between adjacent LBTs. This is as a direct consequence, of the fact that the value at the origin of every LBT (the sampling point), is 0 by default¹⁴ regardless of the fact that the true image \bar{I} may be continuously increasing or decreasing; that is, the reconstruction ignores the global luminance structure which varies by an additive constant

¹⁴which can be easily seen by setting $\Delta x = \Delta y = g_{0,0} = 0$ in (4.4)

throughout each point on every LBT. One way of recovering the global luminance structure is to enforce continuity between all adjacent tiles. This is achieved as follows: firstly, by structuring the tiles in such a way that adjacent tiles ‘overlap’, so that information at the boundary of one tile is also contained at the boundary of the adjacent tile. Figure 4.6 depicts two such ‘overlapping’ tiles and indicates the overlap region of each tile in red¹⁵. Secondly, by finding the $z_{i,j}$ for each tile that minimises the squared difference of the mean values in the overlap regions, between adjacent tiles. That is, finding the additive shift that makes adjacent tiles as continuous as possible, as demonstrated in Figure 4.6.

The squared difference of the mean values in the overlap regions of tiles $\hat{I}_{i,j}$ and $\hat{I}_{i+1,j}$ in Figure 4.6, is given by,

$$\begin{aligned} S_{i+1,j} &= \left[\left(\frac{1}{n} \sum_{k=1}^n \hat{I}_{i,j}^{B_k} \right) - \left(\frac{1}{n} \sum_{k=1}^n \hat{I}_{i+1,j}^{B_k} \right) \right]^2 \\ &= [M_{i,j} - M_{i+1,j}]^2, \end{aligned} \tag{4.6}$$

where there are n pixels in the boundary, $\hat{I}_{i,j}^{B_k}$ and $\hat{I}_{i+1,j}^{B_k}$ represent the local brightness values at the boundaries, and $M_{i,j}$ and $M_{i+1,j}$ the mean local brightness values at the boundaries. When an additive constant is added to tile $\hat{I}_{i,j}$, (4.6) is then modified to,

$$S_{i+1,j} = [(M_{i,j} + z_{i,j}) - M_{i+1,j}]^2. \tag{4.7}$$

¹⁵It should be noted that in the final image reconstruction, after all of the $z_{i,j}$ have been calculated and added to the LBTs, only one of the two overlap regions at every tile boundary are kept, and the other is discarded. The convention followed in this work is to discard all overlap regions above and to the right.

The optimal $z_{i,j}$ is then calculated by solving $\frac{dS_{i+1,j}}{dz_{i,j}} = 0$ for $z_{i,j}$, which yields the following solution,

$$z_{i,j} = M_{i+1,j} - M_{i,j}, \quad (4.8)$$

the difference in the mean boundary values. It should be noted, that minimising the mean squared difference¹⁶ (as opposed to the squared difference of the means), yields an identical solution to (4.8).

For a LBT surrounded by four adjacent tiles, the optimal $z_{i,j}$ is determined by minimising the sum of the squared difference of the means, for all four adjacent tiles. This sum defines an error function,

$$E(z_{i,j}) = S_{i,j+1} + S_{i+1,j} + S_{i,j-1} + S_{i-1,j}, \quad (4.9)$$

which is minimised by the solution to $\frac{dE(z_{i,j})}{dz_{ij}} = 0$, given by,

$$z_{i,j} = \frac{1}{4} [(M_{i,j+1} - M_{i,j}) + (M_{i+1,j} - M_{i,j}) + (M_{i,j-1} - M_{i,j}) + (M_{i-1,j} - M_{i,j})], \quad (4.10)$$

the mean of the differences of the mean boundary values. Both of these examples of tile alignment have only considered the additive constant applied to one tile to bring it into alignment with its neighbours, and have neglected the fact that each neighbouring tile needs to have an additive constant added to it in order to align the entire set of tiles (and not just one to all the others). That is, *all* $z_{i,j}$ have to be calculated simultaneously. In order to step up to this level of complexity, a new error function for the entire set of tiles

¹⁶defined as, $S_{i+1,j}^z = \frac{1}{n} \sum_{k=1}^n [(\hat{I}_{i,j}^{B_k} + z_{i,j}) - \hat{I}_{i+1,j}^{B_k}]^2$.

has to first be defined,

$$E(\mathbf{z}) = \sum_{i=1}^N \sum_{j=1}^M \sum_{p=-1}^1 \sum_{q=-1}^1 S_{i+p,j+q}, \quad (4.11)$$

where,

$$S_{i+p,j+q} = [(M_{i,j} + z_{i,j}) - (M_{i+p,j+q} - z_{i+p,j+q})]^2, \quad (4.12)$$

and where N and M are the dimensions of the array of LBTs.

If a tile is at the boundary of the image then,

$$M_{boundary} = z_{boundary} = 0, \quad (4.13)$$

such that $S_{i+p,j+q} = [M_{i,j} + z_{i,j}]^2$. This is equivalent to aligning all boundary tiles to the same ‘virtual plane’, and is a requirement for an analytical solution to the problem, as will become apparent¹⁷.

The optimal set of $z_{i,j}$ ’s are now derived from the solution to $\frac{dE(\mathbf{z})}{dz} = 0$, which yield a system of $N \times M$ simultaneous linear equations,

$$\frac{dE(\mathbf{z})}{dz_{i,j}} = \sum_{p=-1}^1 \sum_{q=-1}^1 [(z_{i+p,j+q} - z_{i,j}) + (M_{i+p,j+q} - M_{i,j})] = 0, \quad (4.14)$$

¹⁷Physicists may recognise these as the Dirichlet boundary conditions that arise in the solution to some well known differential equations.

with $N \times M$ unknowns. Equation (4.14) can then be re-written as,

$$\mathbf{w}_{i,j}^T \cdot \mathbf{z} = q_{i,j},$$

where $q_{i,j} = -\sum_{p=-1}^1 \sum_{q=-1}^1 (M_{i+p,j+q} - M_{i,j})$, and $\mathbf{w}_{i,j}$ is vector of length $N \times M$ and contains constants that weight the set of additive constants, \mathbf{z} , such that,

$$\mathbf{w}_{i,j}^T \cdot \mathbf{z} = \sum_{p=-1}^1 \sum_{q=-1}^1 (z_{i+p,j+q} - z_{i,j}).$$

The entire set of $\mathbf{w}_{i,j}$ can then be assembled into a sparse matrix, \mathbf{W} , with dimensions $(N \times M) \times (N \times M)$, where each row represents one of the $\mathbf{w}_{i,j}^T$, such that the entire set of simultaneous linear equations defined in (4.14) can be written as,

$$\mathbf{W}\mathbf{z} = \mathbf{q}, \tag{4.15}$$

where \mathbf{q} is a vector of length $N \times M$ containing all of the $q_{i,j}$. The solution of (4.15) now reduces to finding \mathbf{W}^{-1} . Provided that the boundary conditions in (4.13) are satisfied, then $\det[\mathbf{W}] \neq 0$, and \mathbf{W}^{-1} always exists, such that,

$$\mathbf{z} = \mathbf{W}^{-1}\mathbf{q}, \tag{4.16}$$

where \mathbf{W}^{-1} is not sparse, but fully populated. It should be noted that the optimal shifts, \mathbf{z} , calculated using equation (4.16), do not constitute a unique solution to $\frac{dE(\mathbf{z})}{d\mathbf{z}} = 0$. The full set of solutions are given by $\hat{\mathbf{z}} = \mathbf{z} + c$, where c is an arbitrary constant. However, this does not change the validity of the solutions, as c is added to all $z_{i,j}$ simultaneously, which preserves the state of continuity between adjacent tiles. That is, the minimum configuration of tile alignment arrived by minimising $E(\mathbf{z})$, is invariant under a translation of \mathbf{z} , and the

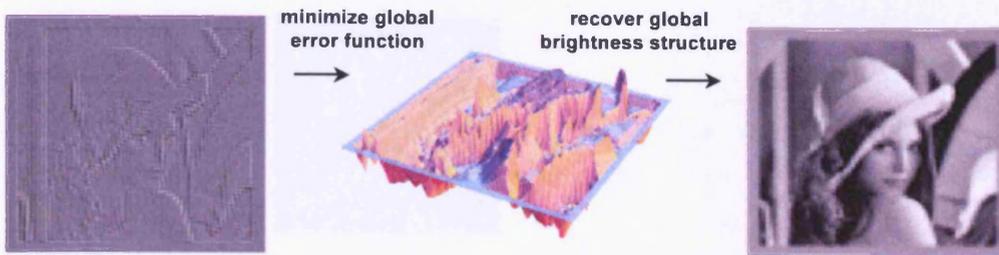


Figure 4.7: An example of how the global luminance structure can be analytically recovered.

complete set of solutions lie on a 1-D manifold (a line), in the $(N \times M)$ -dimensional space of additive shifts.

Figure 4.7 illustrates how this process can be used to recover the global luminance structure, that is, estimate the $g_{0,0}$ components of the local jets used to fully reconstruct the image. Figure 4.8 illustrates a map and plot of the differences between a reconstruction made using the actual $g_{0,0}$ components (sampled from the image), and those calculated using the algorithm described above. It is clear to see from the concurrent plot of the reconstruction cross-sections that the two are almost identical. This is confirmed by the root-mean-squared error (RMSE) between the two reconstructions, which was calculated to be 0.014 model-units¹⁸, where the standard deviation of model output values for the reconstruction made using the actual $g_{0,0}$ components was calculated as 0.42 model-units. Thus, it is possible to accurately recover the local-mean (or D.C.) luminance level without directly sampling it, or using filters with a large spatial extent as an approximation.

The error in the reconstruction of the global luminance structure is a direct result of the error incurred by the local jets when estimating the form of $\hat{I}_{i,j}$ with derivative information from a single point, and only using GD filters up to second-order. Figure 4.9 illustrates a map and plot of the differences between a reconstruction made using the actual $g_{0,0}$ components and those calculated using the algorithm described above, for the same size of LBT as before, but with GD filters set at double the scale ($\sigma = 4$); analogous to decreasing the size

¹⁸where model-units are defined as the pixel-values increments in the reconstructed image

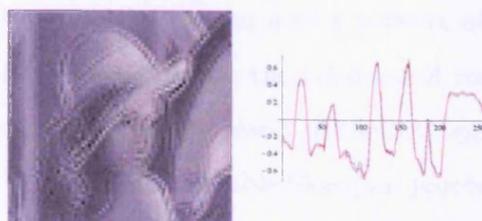


Figure 4.8: (left) A map of the differences between reconstructions made using the actual $g_{0,0}$ components and that calculated using the analytical ILC algorithm. (right) A plot of the reconstruction made using the actual $g_{0,0}$ components (black) and those derived from the analytical ILC model (red), for a cross-section at the midpoint of the figure on the left.

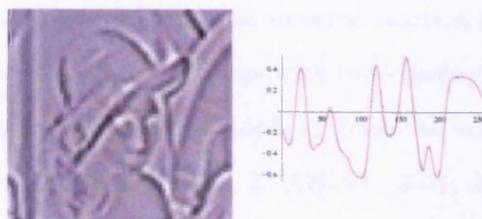


Figure 4.9: (left) A map of the differences between reconstructions made using the actual $g_{0,0}$ components and that calculated using the analytical ILC algorithm. The tiles are the same size as that used in Figure 4.8, but the scale of the GD filters used in the local jets is doubled ($\sigma = 4$), such that there is less error in the reconstruction. (right) A plot of the reconstruction made using the actual $g_{0,0}$ components (black) and those derived from the analytical ILC model (red), for a cross-section at the midpoint of the figure on the left.

of the LBTs. As the function becomes smoother and more predictable, the reconstruction error decreases: the RMSE between the two reconstructions was re-calculated to be 0.0033 model-units.

A Biologically Plausible Implementation of the ILC Model

Despite the success of the analytical ILC algorithm in reconstructing the global luminance structure, its mathematical formulation is incompatible with the distributed processing properties of neurons (MacKay, 2003). That is, it is asking too much of a single neuron to be able to invert a large matrix analytically. Furthermore, the information required to compute the global brightness structure comes from across the entire visual field¹⁹. Thus,

¹⁹by inspection of (4.16) it can be seen that *every* optimal $z_{i,j}$ is a weighted linear sum of *all* the q_j 's, and hence the $M_{i,j}$, which come from across the entire visual field.

the ILC algorithm would require the existence of a network of long-range inter-neurons to transmit this information - depending on the existence of such a system of cells in V1 would be controversial, as discussed in Chapter 3. If the ILC approach to recovering global luminance structure is to be considered a viable biological process, then it needs to be cast in the computational context of neurons.

One way in which this problem can be addressed, is by relaxing the criterion for an analytically exact solution and proceeding to minimise the error function with a numerical approximation to the analytical solution. Minimising an error function by gradually converging to an optimal solution is a process synonymous with the training phase of neural networks, and in-particular with the functionality of Hopfield networks, which can be constructed to solve similar optimisation problems (Dayan & Abbott, 2001; MacKay, 2003). The construction of a fully operational spiking neural network that implements the ILC algorithm is beyond the remit of this work. Instead, what will be described is a biologically-plausible implementation of the ILC algorithm, based on a numerical method for minimising the error function. Furthermore, it will be shown that by considering only local interactions between neighbouring LBTs, a globally optimal solution can still be recovered.

Perhaps the simplest algorithm for minimising a function, is the method of Gradient Descent (GRADD). By taking steps proportional to the negative of the gradient of the function at the current point, the GRADD algorithm approaches the local minimum of the function (Press, Teukolsky, Vetterling, & Flannery, 2002). For the error function defined in (4.11), each step proceeds as,

$$\mathbf{z}^{n+1} = \mathbf{z}^n - \gamma \nabla E(\mathbf{z})|_{\mathbf{z}=\mathbf{z}^n}, \quad (4.17)$$

where n is the number of iterations, γ is a constant and determines the size of each ‘step’ across the function, and \mathbf{z}^0 are the initial conditions that must be defined before the algorithm begins (and which may affect the outcome).

If $\mathbf{z}^0 = 0$ are the chosen initial conditions and $\gamma = \frac{1}{4}$, then the components of (4.17) after the first step become,

$$\begin{aligned}
 z_{11}^1 &= -\frac{1}{4} \sum_{p=-1}^1 \sum_{q=-1}^1 (M_{1+p,1+q} - M_{1,1}) \\
 z_{12}^1 &= -\frac{1}{4} \sum_{p=-1}^1 \sum_{q=-1}^1 (M_{1+p,2+q} - M_{1,2}) \\
 &\vdots = \vdots \\
 z_{i,j}^1 &= -\frac{1}{4} \sum_{p=-1}^1 \sum_{q=-1}^1 (M_{i+p,j+q} - M_{i,j}). \tag{4.18}
 \end{aligned}$$

Hence, the best the algorithm can do without *a priori* knowledge of what the initial conditions should be, is to shift each tile by the mean difference of the mean boundary values as in (4.10). If these shifts are immediately added to their LBTs, and as opposed to progressing to step $n = 2$ of (4.17), the mean difference of the mean boundary values is recalculated, and step (4.18) repeated, then an alternative to (4.17) is constructed. This algorithm is a modification of (4.17) such that,

$$\mathbf{z}^{n+1} = \mathbf{z}^n - \gamma \nabla E(\mathbf{z}; n)|_{\mathbf{z}=0}, \tag{4.19}$$

where,

$$z_{i,j}^{n+1} = z_{i,j}^n - \frac{1}{4} \sum_{p=-1}^1 \sum_{q=-1}^1 (M_{i+p,j+q}^n - M_{i,j}^n). \tag{4.20}$$

This is designed to be the simplest and most biologically plausible implementation of the GRADD algorithm: if the shift required for each LBT is represented by a single cell, then the only inputs that this cell would require are its boundary differences with neighbouring cells. This process could also proceed in real time, which would not be the case if each tile

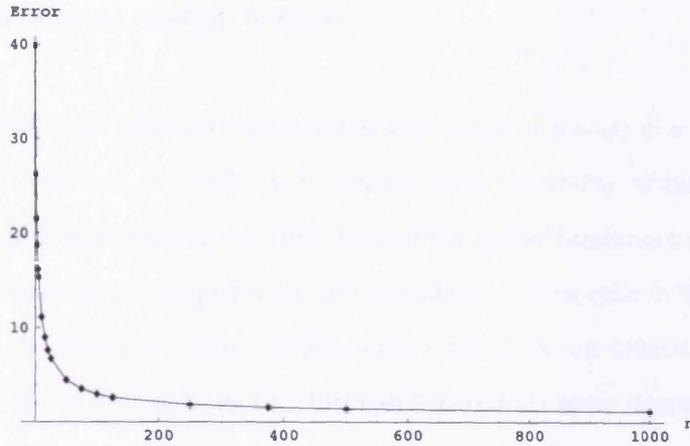


Figure 4.10: A plot of the RMSE per-tile after each successive iteration of (4.19), for the test image reconstructed with local jets of $\sigma = 2$, as in Figure 4.5.



Figure 4.11: The reconstruction of the test image at a scale of $\sigma = 2$ using the biologically plausible implementation of the ILC algorithm. From left to right: the reconstruction terminated after 0, 8, 64, and 512 iterations of (4.19).

has to receive input about the shifts from neighbouring tiles, before its own shift could be updated (which would be the consequence of a direct application of (4.17) when $n > 1$).

Figure 4.10 shows a plot the RMSE alignment error per-tile after all of the shifts have been added to their corresponding LBTs,²⁰ for the test image reconstructed with local jets of $\sigma = 2$ as in Figure 4.4. The plot clearly shows the rate at which this biologically-plausible implementation of the ILC algorithm converges. The RMSE alignment error per-tile for the analytical version of the ILC algorithm is calculated as 0.62 model-units, and after 1000 iterations the biologically plausible ILC algorithm arrives at 0.72 model-units; within close

²⁰RMSE alignment error per-tile is defined as the sum of $S_{i+1,j}^z = \frac{1}{n} \sum_{k=1}^n \left[\left(\hat{f}_{i,j}^{B_k} + z_{i,j} \right) - \hat{f}_{i+1,j}^{B_k} \right]^2$, for all tile boundaries in the image, divided by the number of tiles representing the image.

proximity of the analytical solution to (4.16).

Figure 4.11 illustrates the reconstruction of the test image at a scale of $\sigma = 2$ at intervals of 0, 8, 64, and 512 iterations of (4.19), and demonstrates the ability of the biological implementation of the ILC algorithm to not only recover the global luminance structure, but also to accurately reconstruct an image from a set of contrast coding cells in V1 that sample the image discretely. It should be noted at this stage, that it is not crucial that the RFs are *perfect* GDs - small deviations from this form will introduce some degree of error into the LBTs, and the minimum possible alignment error, but will ultimately yield similar results via the reconstruction process.

Towards a Complete Model of Luminance Coding in V1

Before the biological ILC algorithm can be used to model low-level visual processes, the following issues have to be considered: the role of multiple scale and SF channels; the scales at which the visual system recovers the global luminance structure; the algorithm's performance under noisy conditions; and the minimum number of iterations which the algorithm needs to perform in order to achieve an accurate reconstruction.

Multiple Scale Channels

Motivation for the inclusion of multiple scales in the model comes from the principles of Scale-Space: that there is a need to represent image information at multiple scales as different types of information are present at different scales. Furthermore, there are the physiological studies which have found cells tuned to many different SFs (De Valois, Albrecht, & Thorell, 1982), and evidence from psychophysical experiments which demonstrate that sensitivity to a particular range of SFs can be altered by adaptation to SFs in that range (Goldstein, 2003). Thus, the existence of multiple scales is a reasonable assumption.

The issue of 'how many channels and at what scale', is constrained from physiology and the available computing power for running the model. Following on from the discussion in

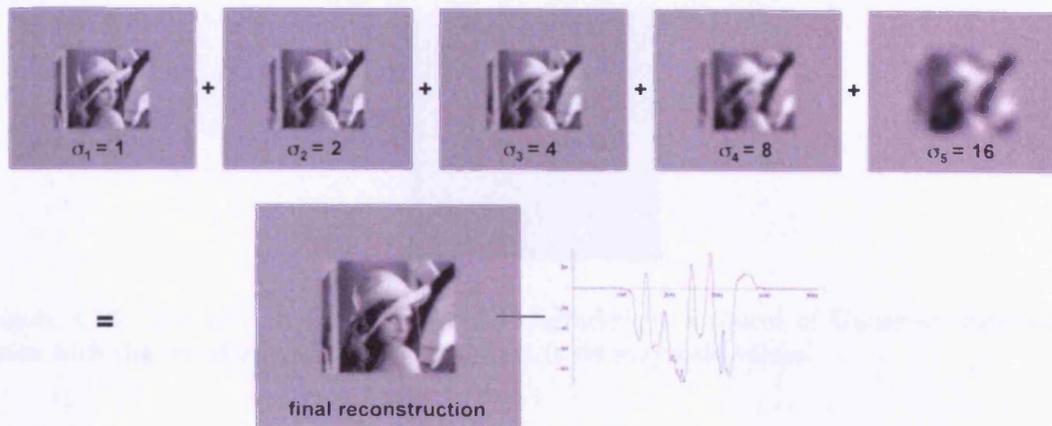


Figure 4.12: The final multi-scale reconstruction of the test image is the result of an un-weighted linear sum of the reconstructions at every scale.

Chapter 3 regarding the range of SFs represented in V1, early studies seldom found many simple cells in the foveal representation of V1 tuned to SFs below $0.5c/^\circ$. The mean SF tuning was found to be between $3.0c/^\circ$ and $4.6c/^\circ$, and cells tuned to SFs as high as $15c/^\circ$ were also discovered (De Valois et al., 1982; Foster, Gaska, Nagler, & Pollen, 1985; Silverman et al., 1989).

The final incarnation of the biological ILC algorithm has been configured to use images with dimensions of 512×512 pixels to represent 10° of visual angle. With this configuration, GD filters with a scale of $\sigma = 16$ (or 0.313°) are tuned to SFs of $0.6c/^\circ$ for first-order derivatives and $0.8c/^\circ$ for second order derivatives. Similarly, GD filters with a scale of $\sigma = 1$ (or 0.020°) are tuned to SFs of $9.6c/^\circ$ for first-order derivatives and $12.8c/^\circ$ for second order derivatives. Filters with $\sigma < 1$ are not used to avoid discretisation problems. The half-width of each square LBT, Δx , is set to σ for each scale, such that the distance between adjacent sampling points is 2σ , and the Gaussian envelopes of filters at adjacent sampling points sum to approximately one across the entire image. This forces all pixels to contribute an approximately equal amount to the total output of all filters²¹. Thus, information from the retinal image will not be over or under-represented in the filter output. Furthermore,

²¹That is, if a 2-D Gaussian distributions with scale σ are arranged on a lattice with a spacing of 2σ , then the sum over all distributions will be approximately equal to one across the entire image. Each point on the image will then contribute an approximately equal amount to the responses of the GD filters



Figure 4.13: The test image in Figure 4.4 degraded by a source of Gaussian distributed noise with the standard-error of the noise set to 64 grayscale values.

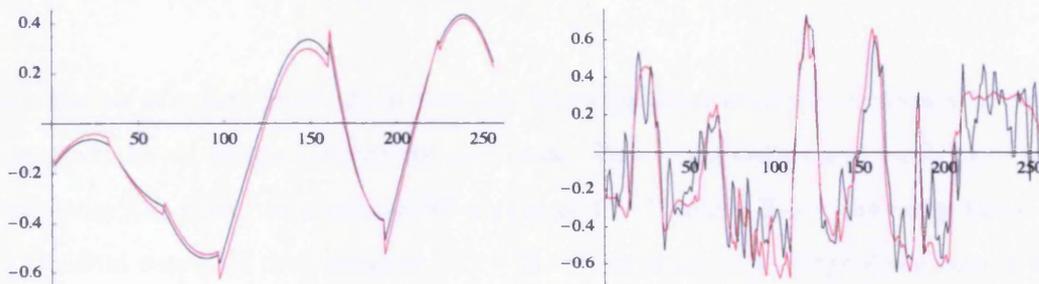


Figure 4.14: (left) A plot of the reconstruction made using the analytical ILC algorithm for the original image (red) and the noise degraded image (black), for a cross-section at the midpoint of the $\sigma = 16$ (0.31°) scale channel's output. The RMSE between the two reconstructions was calculated to be 0.019 model-units. (right) A plot of the reconstruction made using the analytical ILC algorithm for the original image (red) and the noise degraded image (black), for a cross-section at the midpoint of the $\sigma = 1$ (0.02°) scale channel's output. The RMSE between the reconstructions was calculated to be 0.1 per-tile.

LBTs with $\Delta x = \sigma$ will not incur a pathological amount of error in their approximation of the blurred image, as demonstrated in Figure (4.8). This configuration yields 256 tiles for a $\sigma = 16$ (0.313°) channel and 65,536 tiles for a $\sigma = 1$ (0.020°) channel. If the size of the image used were larger, such that filters tuned to higher SFs could be used, then the number of tiles involved would result in the model taking a prohibitory long time to run. However, seeing as the vast majority of simple cells in V1 are tuned to SFs between $0.5c/^\circ$ and $8.0c/^\circ$ (De Valois et al., 1982; Silverman et al., 1989), it is felt that the model configured as described above, is capable of accurately representing the scales present in V1. Between $\sigma = 16$ (0.313°) and $\sigma = 1$ (0.020°) channels, the other scale channels which are included are $\sigma = 8$ (0.156°), $\sigma = 4$ (0.078°), $\sigma = 2$ (0.039°). The principle reason for assembling

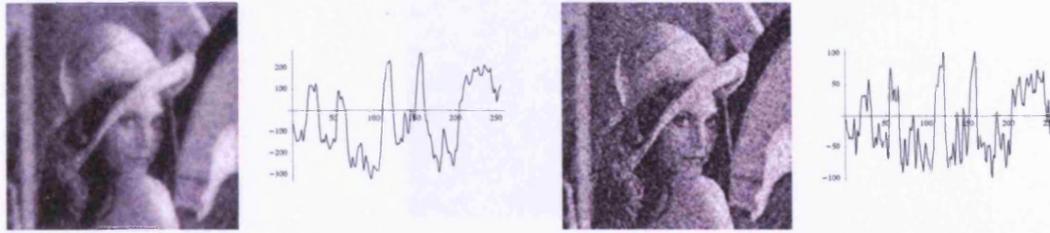


Figure 4.15: (left) Multiple-scale reconstruction of the noisy test image with a cross-sectional profile taken at the mid-point. (right) Reconstruction of the test image at a scale of $\sigma = 1$ (or 0.020°) with a cross-sectional profile taken at the mid-point. The multiple-scale reconstruction has the effect of averaging out the effects of the noise in the test image.

scale channels in octave intervals, is that this forces the number of tiles necessary to cover the image to be an integer number for each scale. This configuration also yields first- and second order GD filters with a mean SF tuning of $4.3c/^\circ$, which lies in the range found in V1. The final output of the biological ILC is the result of an un-weighted linear sum of the reconstructions at every scale. Figure 4.12 demonstrates the model's output for the test image. Although the sum over scales has no obvious significance, as it does for the NSS and ODOG models, where it is motivated by each channel representing a component in a local Fourier-series representation of the image, its necessity will become more obvious after the issue of 'robustness to image noise' is discussed.

Robustness to Image Noise

Figure 4.13 illustrates a version of the test image in Figure 4.4 degraded by Gaussian distributed noise with a standard-error of 64 grey-scale values. This may be considered a crude model of the noise present in the primary visual stream from the PRCs in the retina up to the set of simple cells whose RFs are modelled as GD filters. Figure 4.14 shows concurrent plots of the reconstructions made by the analytic ILC algorithm for both the degraded and original test images, for the largest and smallest scales. It is apparent that the large scale is little effected by the presence of noise whereas the small scale is heavily perturbed. This is reflected in the RMSE between the reconstructions, calculated as 0.019 for the large scale and 0.1 for the small scale (almost an order of magnitude greater). Furthermore, Figure 4.15 illustrates the output for the analytical multi-scale reconstruction of the degraded test

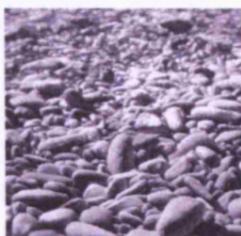


Figure 4.16: Natural scene image used in the analysis of algorithm convergence.

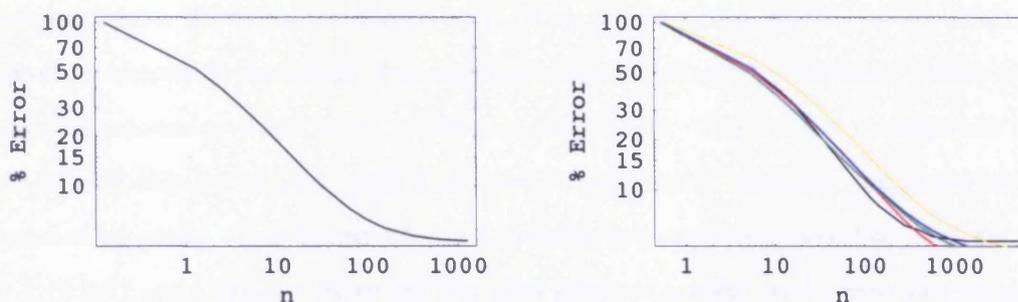


Figure 4.17: (left) A plot of the average RMSE per-tile, expressed as a percentage of the total RMSE per-tile after each successive iteration. (right) A plot of the average RMSE per-tile, expressed as a percentage of the RMSE per-tile of the analytical reconstructions, after each successive iteration, for each individual scale channel: (yellow) $\sigma = 1$ (0.020°), (blue) $\sigma = 2$ (0.039°), (green) $\sigma = 4$ (0.078°), (red) $\sigma = 8$ (0.156°), (black) $\sigma = 16$ (0.313°).

image (left) and the analytical reconstruction for just the smallest scale (right). It is clear to see that the multi-scale reconstruction has a considerably less granular appearance as a consequence of the strong smoothing properties of the large scale channel, while retaining a high-level of acuity. Hence, it can be concluded that the larger scales, in the context of a multi-scale reconstruction, play an important role in providing a robust calculation of the global luminance structure.

Determining the Number of Iterations

The RMSE per-tile, over all scales, and expressed as a percentage of the total RMSE per-tile before the algorithm began, was calculated for three natural scene test images (Figure 4.1, Figure 4.4, and Figure 4.16), and averaged. Figure 4.17 (left) illustrates how this quantity evolves through time (with the number of iterations), and Figure 4.17 (right) illustrates how this quantity evolves when it is calculated for each individual scale channel. It is clear

that the algorithm does not need to progress further than about 300 iterations before an adequate reconstruction is achieved in terms of error. However, as opposed to selecting a fixed number of iterations based on a subjective judgement, the following concept was used instead: the analytical solution given by equation (4.16), is a linear combination of *all* mean local brightness values at the boundary. That is, it takes information from *all* tiles. Motivated by this fact, it was decided that each tile must have at least one opportunity to ‘interact’ with the tile most distant to it. The two most distant tiles at any scale are in opposite corners of the image. For these two tiles to be able to interact, it takes twice as many iterations as there are tiles between the corners. This is as a consequence of a tile not being able to ‘interact’ with a tile that is not directly adjacent, and hence having to interact by proxy, via adjacent tiles, which takes two steps as opposed to one. For the $\sigma = 1$ (0.02°) scale channel there are 362 tiles between corners, and hence 724 iterations are required for these two to have the opportunity to interact. Although the larger scales require fewer iterations in this paradigm, it makes little sense to terminate the algorithm for one scale, while another scale is still being processed, and hence 724 iterations will be used for all channels.

By inspection of Figure 4.17, it is apparent that all scale channels will have a RMSE per-tile of $\lesssim 10\%$. However small this figure may appear, it should be acknowledged that there are a total of 87,296 tiles in the model, and thus in relative terms the total amount of error in the system is still significant.

The Spatial Frequency Content of the Biological ILC Model’s Output

Now that a complete model of luminance coding in V1 has been proposed, it remains to test the quality of its image reconstruction. This will be performed by analysing the SF content of the model’s reconstructions for white noise, a natural scene with a $\frac{1}{f^\alpha}$ distribution of SFs (Figure 4.4), and a $0.4c/^\circ$ fundamental frequency square wave grating. Results will be compared with the SF content of the original input and with that from a multi-scale blurring at the same spatial scales as the model (equivalent to including the actual $g_{0,0}$ components in the reconstruction, or convoluting the input images with a 2-D version of

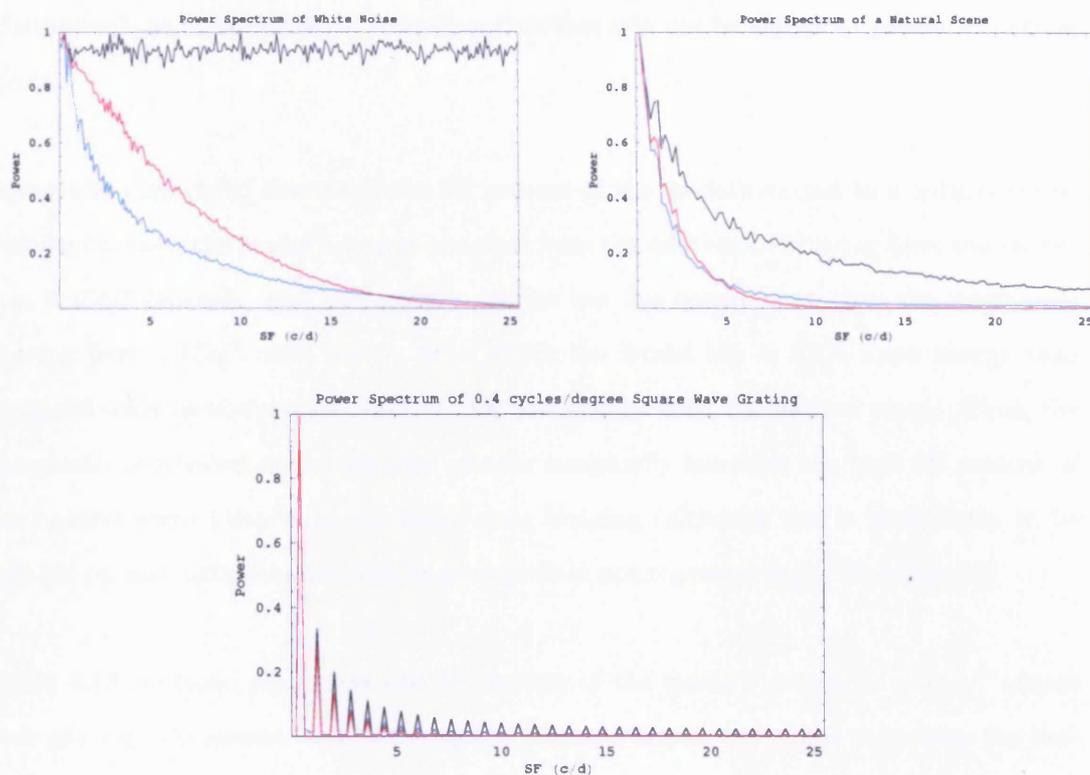


Figure 4.18: The SF content of the model's output (blue lines) compared to that of the input image (black lines) and that from a multi-scale blurring at the same spatial scales as the model (red lines), for white noise (top left), a natural scene (top right), and a $0.4c/^\circ$ square wave grating (bottom).

the filter depicted in Figure 4.25). The comparison with the multi-scale blurring output has been included, so that differences in SF representation can be compared against the pure blurring properties of the model, and thus highlight any that arise out of the biological tile alignment process.

Figure 4.18 (top left) illustrates the SF content of the model's output to white noise. It can be seen that both the model's output and that from the multi-scale blurring have less energy than white noise from $0.45c/^\circ$ onwards, and that the model's output has less energy than the multi-scale blurring up until $17.5c/^\circ$, after which the model has marginally more energy (by $\approx 1\%$). Thus, the biologically motivated reconstruction process, which will be shown to generate contrast effects, does not represent SFs as efficiently (in terms of raw

information), as the multi-scale reconstruction that will not be shown to generate contrast effects.

Figure 4.18 (top right) illustrates the SF content of the model's output to a natural scene. Once again, both the model's output and that from the multi-scale blurring have less energy from $0.45c/^\circ$ onwards, and the model's output has less energy than the multi-scale blurring from $0.77c/^\circ$ until $4.5c/^\circ$, after which the model has $\approx 2.5\%$ more energy than the multi-scale reconstruction, and $\approx 15\%$ less energy than the natural scene. Thus, the biologically motivated reconstruction process marginally increases the high SF content of this natural scene relative to the multi-scale blurring (although this is more likely to be high SF reconstruction noise), but in general does not represent high SFs efficiently.

Figure 4.18 (bottom) illustrates the SF content of the model's output to a $0.4c/^\circ$ square wave grating. Consistent with the outputs discussed above, the model represents the lowest SFs ($0.4c/^\circ$ and $1.2c/^\circ$) almost perfectly with respect to the original image and the multi-scale blurring, and onwards from $1.2c/^\circ$, both the model's output and that from the multi-scale blurring encode the higher SFs inefficiently, with the model having a marginally higher energy than the output from the multi-scale blurring (again, this is more likely to be high SF reconstruction noise).

To summarise, the model represents low SFs ($\lesssim 1c/^\circ$) with reasonable accuracy, but higher SFs are considerably suppressed. This is a result of the intrinsic multi-scale blurring properties of the model, which can be seen by comparing the model's output to that from a multi-scale blurring with no necessary reconstruction. However, the model represents higher SFs ($\gtrsim 5c/^\circ$) marginally better than the pure multi-scale blurring output, by $\approx 1\% - 3\%$. This is likely to arise out of the remaining tile-alignment error inherent in the model's reconstructions at multiple scales, for which it has been shown that there is more error in the higher spatial-scale channels. This is of particular relevance, as later on in this chapter it will be shown that this error is responsible for generating contrast effects.

Modelling Lightness and Brightness Illusions

The multi-scale biological ILC model configured as outlined above, is considered a model of brightness based on the known physiology of V1. It has already been shown to accurately reconstruct natural scene images. The next part of this work returns to the discussion of lightness and brightness illusions, and is concerned with using the model to provide accounts for these phenomena. Although the results of this process are covered in greater detail in Chapter 5, simultaneous brightness contrast and White's effect will be briefly discussed here as an example of how the constituent parts of the model come together to generate these effects.

From Model Output to Predicted Brightness

To be able to relate the model's output to psychophysical reality in a clear and consistent manner, it is necessary to transform the output into psychophysically measurable units. A simple way by which model units can be suitably transformed is to scale them relative to the model's response to black and white in a scene, such that the brightness of black relative to white is 0%, and that of white relative to white 100%. All other brightness levels can now be expressed as a percentage of the brightness of the white in a scene, which can be determined psychophysically. This can be achieved by the following linear and affine transformation on model units (MU),

$$brightness = \frac{MU + Abs(B)}{B + W} \times 100,$$

where B is the model's minimum response to the black in the scene, and W the model's maximum response to white. Only areas which are far from the influence of edges will be used to determine B and W, and in the event that there is no suitable area, a region of black or white will be added to the top-right corner of the scene, within the mean-luminance grey padding and away from the influence of the stimulus. The expected value of B across all scenes used to test the model (in Chapter 5), was estimated to be -449 model units with a

standard deviation of 16 model units. The expected value of W was estimated to be 453 model units with a standard deviation of 18 model units. In theory, both B and W should not be random variables, but the existence of a deviation in their values arises out of the fact that it is impossible to determine the responses to black and white in complete isolation from rest of the scene.

The Minimum Discriminable Response

Now that it is possible to map the model's output into a psychophysically measurable space, it is pertinent to ask what is the just noticeable difference (JND) in the framework derived above. A model for JNDs in human perception was first proposed by Weber (1978), who stated that a change in any perceptual dimension, P , is described by the following differential equation,

$$dP = k \frac{dS}{S},$$

where S is the magnitude of the stimulus and k a constant of proportionality. Thus, for a unit change in perception, or JND,

$$dS_{JND} = kS. \tag{4.21}$$

In the context of visual perception and psychophysics, S represents the mean luminance of an image, and k has been empirically determined, both in human and primates, to lie within the range of 0.1-0.14 (Cornsweet & Pinsker, 1965; Huang, MacEnvoy, & Paradiso, 2002). The vast majority of stimuli reported in Chapter 5 have been constructed with a mean-luminance of 50%, hence the model will predict a significant change in brightness for regions with a 5%-7% difference in predicted brightness. Stimuli with a mean luminance above or below 50% will have their predicted brightness interpreted accordingly.

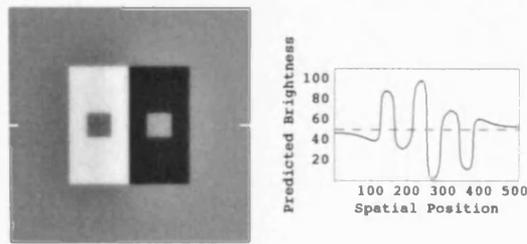


Figure 4.19: The model's output for the simultaneous brightness contrast stimulus originally shown in Figure 1.1 (left), and its cross-sectional predicted brightness profile taken as indicated by the position of the markers (right). The model correctly predicts that the test patch on the left is darker than that on the right.

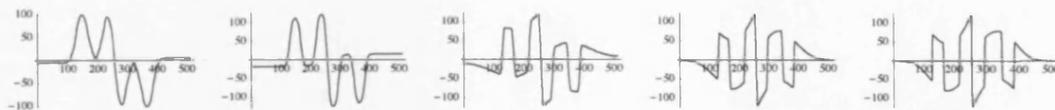


Figure 4.20: The cross-sectional predicted brightness profile of each scale used in the reconstruction above. The furthest left is the output for the largest channel ($\sigma = 16$ or 0.31°), and the furthest right is the output for the smallest channel ($\sigma = 1$ or 0.02°).

SBC and White's Effect: A Tale of Contrast and Assimilation

Figure 4.19 illustrates the output of the model for the simultaneous brightness contrast configuration (Figure 1.1, left). It is clear to see that the brightness of the test patch on the left is correctly predicted to be darker than the one on the right, and the difference between the mean predicted brightness of the test patches is 19.5%, which is greater than two JNDs. Figure 4.19 illustrates the raw output from each scale used in the reconstruction. It is obvious that the effect is derived predominantly from the output of the smaller scales, while the largest scale predicts some degree of assimilation.

Figure 4.21 illustrates the output of the model for a White's effect display (Figure 1.1, right). It can be seen that the test patch on the left is correctly predicted to be darker than the one on the right (at all points except at the very centres of the test patches), and the difference between the mean predicted brightness of the test patches is 6.2%, approximately one JND. Figure 4.22 illustrates the raw output from each scale used in the reconstruction.

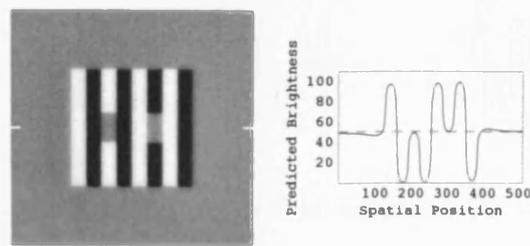


Figure 4.21: The model's output for the White's effect display originally shown in Figure 1.1 (left), and its cross-sectional predicted brightness profile taken as indicated by the position of the markers (right). The model correctly predicts that the test patch on the left is darker than that on the right.

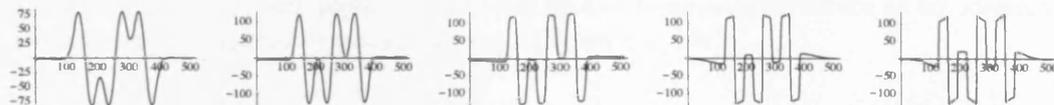


Figure 4.22: The cross-sectional predicted brightness profile of each scale used in the reconstruction above. The furthest left is the output for the largest channel ($\sigma = 16$ or 0.31°), and the furthest right is the output for the smallest channel ($\sigma = 1$ or 0.02°).

In contrast to the output for the simultaneous brightness contrast display, the effect is generated by the larger scales, while the output from the smaller scales predict simultaneous brightness contrast.

These results lead to the following hypothesis as to how the model correctly predicts the existence of both these phenomena: simultaneous brightness contrast (or induction effects) are generated by reconstruction error in the global luminance structure of the smaller scales. White's effect (or assimilation effects), are generated by the blurring inherent in the reconstructions at larger scales.

The latter proposition can be verified by examining the *actual*²² zero-order components of local jets used to reconstruct the White's effect display (Figure 4.23). It is clear to see that the effect is 'hard coded' into the global luminance structure of the largest scales (as a

²²that is, as sampled by zero-order GD filters.

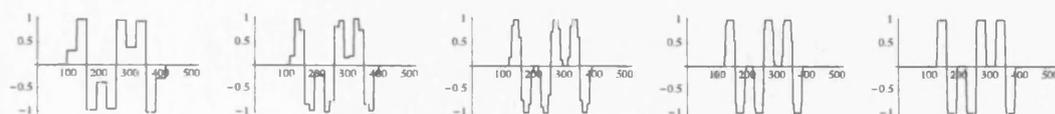


Figure 4.23: Cross-sectional plots of the *actual* global luminance structure at all scales, for the White's effect display (Figure 1.1, right).



Figure 4.24: Cross-sectional plots of the *actual* global luminance structure at all scales, for the simultaneous brightness contrast display (Figure 1.1, left).

consequence of large scale blurring), and it is this which drives the effect. In contrast, the actual global luminance structures for the simultaneous brightness contrast display do not betray the source of the effect, but again shows evidence of assimilation at larger scales. Thus, contrast effects are generated by the error incurred by the biological ILC algorithm terminating before it converges to the optimal solution. As a direct consequence of this, contrast effects will eventually vanish as the number of iterations in the model approach infinity, whereas assimilation effects will remain.

There exists an asymmetry between the assimilation and contrast mechanisms in the model: assimilation is confined to short distances, whereas contrast can occur unconstrained by distance. Figure 4.25 illustrates how the magnitude of the blurring effect varies across space. The plot was constructed by summing Gaussian envelopes at each scale used in the model. It can be seen from the plot, that the effect is very strong over short distances, but falls off rapidly: the effect diminishes by a factor of two after $\sim 0.07^\circ$. The contrast effects are not constrained by any formal aspect of the model, only by the degree of error in the reconstruction, which is a function of the number of iterations. It is manifest most obviously at edges, as it is in these regions that the shape of the LBTs will have the highest degree of extrinsic curvature. As a direct consequence, adjacent tiles in this region require

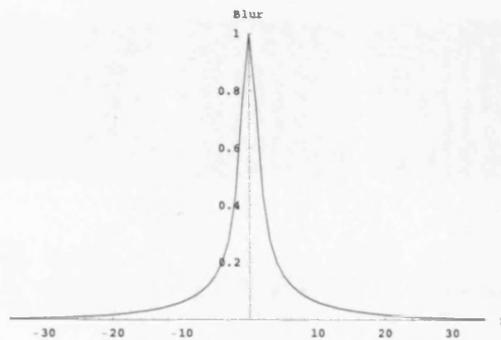


Figure 4.25: The strength of the blurring effect as a consequence of the model reconstructing a smoothed version of the image at every scale. The plot is constructed by summing Gaussian envelopes at each scale.

more alignment and therefore more iterations, before optimal error alignment is reached. As optimal alignment error is never reached (due to the algorithm stopping at 724 iterations), the effect persists to the final brightness output.

White's Effect and the Choice of Spatial Scales

As the model's prediction for White's effect is based on the assimilation derived from the intrinsic blurring of the reconstruction process, the magnitude of the effect is critically dependent on the scales chosen to comprise the final model. Figure 4.26 demonstrates the effect of successively removing the lowest spatial scale channel from the reconstruction, for the $0.8c/^\circ$ White's effect stimulus discussed previously and illustrated in Figure 4.21. It is clear to see that if only the lowest ($\sigma = 16$, or 0.31°) spatial scale channel is removed, then the effect is nulled, and then as further low spatial scale channels are removed, then the effect is reversed to SBC. Thus, the model's predictions for White's effect (and subsequently all assimilation effects), are fundamentally connected to the choice of spatial scale channels. If only one lower spatial scale channel were added, then the magnitude of all assimilation effects would be greater (although at the detriment to the magnitude of all predicted contrast effects), and vice-versa.

This is of particular relevance, as later in Chapter 5 (Figure 5.29) it will shown that the

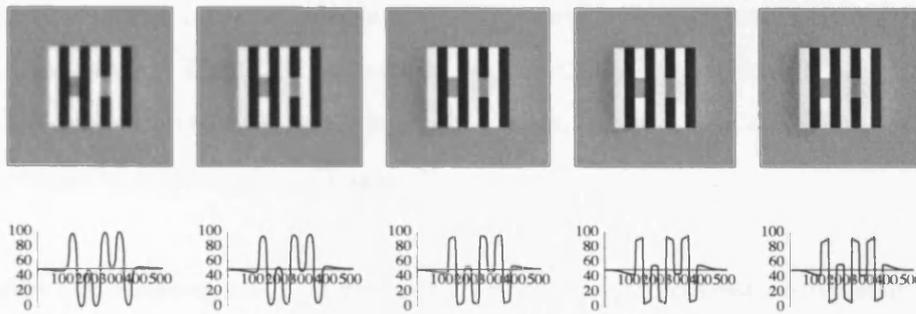


Figure 4.26: The model's predictions for White's effect ($0.8c/\circ$ inducing grating), as progressively more spatial scale channels are removed from the reconstruction. No spatial scales removed (furthest left), then the next lowest spatial scale channel is removed from each successive reconstruction, until the highest spatial scale channel remains (furthest right).

effect at this inducing grating SF is smaller than has been reported by some studies. This is a direct result of the choice of SF channels in the Biological ILC model, and in contrast to the ODOG and NSS models, which predict larger effects for this White's effect stimulus, by virtue of their particular choice of SF channels (as discussed in Chapter 3). However, for the reasons discussed previously in this chapter, relating to the known size of RFs in V1 and matters of modelling elegance, and because in the next chapter the model will be shown to yield a good fit to many different configurations of White's effect (and other assimilation phenomenon), the choice of spatial scale channels in the current incarnation of the model are deemed a good representation of the independent spatial scale channels used in V1.

Comparing the Biological ILC Model to the ODOG and NSS Models

The biologically plausible incarnation of the ILC model described in this chapter differs from the ODOG and NSS models in a number of ways. Perhaps most significantly, it is not based on a local Fourier representation of the image - the computational context within which the model operates, is defined by the conjecture that the visual system does not have access to the local mean luminance (or D.C.) within the image, and must imply this from the response of contrast coding cells. The model explicitly predicts that this information, once calculated, will be present in the blob-like cells of V1, and hence predicts that

the response of these cells will *always* always correlate with perceived luminance, and not absolute luminance. This is in contrast to the ODOG and NSS models, which use RFs $< 0.3c/^\circ$ to estimate the local mean luminance levels in the image, the existence of which is controversial (as discussed in Chapter 3).

As a direct consequence, the ILC model utilises the simple contrast coding cells found in V1 to measure the 1st- and 2nd-order derivatives of a blurred version of the original image. This information is then used to reconstruct an approximation of the image in a square region around each sampling point (or LBT), using a truncated Taylor expansion without the 0th order component, which is calculated by aligning overlapping regions of reconstructed image by means of an error minimisation scheme. As the derivatives of the blurred image do not need to be sampled continuously to generate each tile, the biological ILC model does not need to assume that RFs for all the necessary SF tunings are present continuously across the visual field, which is an implicit assumption of the ODOG and NSS models. Furthermore, it does not make use of contrast coding cells which are tuned to SFs below $0.6c/^\circ$, keeping the model within the realms of biological plausibility. Instead, low SFs are present in the reconstructions of the blurred image.

The ODOG and NSS models make use of image re-normalisation and SF channel re-weighting schemes, respectively. Both of these non-linear responses are based upon the statistics from across the entire set of filter responses, and neither model proposes a mechanism for how this information is calculated and communicated across the visual cortex - it is not feasible for one side of V1 to be immediately connected to the other as a consequence of the limited range of intra-cortical connections (as discussed in Chapter 3). The biologically plausible incarnation of the ILC model proposes an explicit mechanism for how information can be passed from one region of V1 to another, without relying on long-range intra-cortical connections. This was motivated as a means for minimising global tile alignment error from the information present in local tiles.

Furthermore, the biological ILC model accounts for both SBC and White's effect through

two separate mechanisms of contrast and assimilation, in contrast to the ODOG and NSS models, which achieve accounts for these effects through a single re-normalisation and SF channel re-weighting scheme. The model predicts that contrast is a long-range effect, whereas assimilation is a short-range effect in comparison.

Finally, one way in which the biological ILC model differs from the ODOG and NSS models, is that a *complete* reconstruction of the image is carried out for every spatial scale in the model, and then the set of reconstruction are assembled in an un-weighted linear sum to produce the final reconstruction. This is in contrast to the ODOG and NSS models, which only produce the one reconstruction out of the information present in separate SF channels, and are thus more computationally 'cheap', requiring less biological energy to run.

Chapter 5: Results & Discussion

The purpose of this chapter is to compare predictions of the model to various lightness and brightness phenomena with the aim of confirming the theory. The primary purpose of this task is to validate the model as being indicative of how the visual system is processing information. That is, if the model and the visual system both fail at reproducing the original image, and in the same places, then it may be a direct consequence of sharing similar mechanisms. Secondly, as the model is based on the physiology of V1 (and possibly V2), then all phenomena that it can account for could be attributed to processing within these areas of the cortex. This will facilitate the differentiation between the many high- and low-level principles that these phenomena are usually attributed to.

The chapter is divided into three sections according to the type of phenomenon under discussion: static lightness and brightness illusions (of contrast and assimilation); perceptual filling-in effects; and the temporal dynamics of these phenomena. Static lightness and brightness illusions are characterised as phenomena that arise as a result of the specific configuration of surfaces in a scene, and as discussed in Chapter 2 have been attributed to both mechanisms of low- and higher-level vision. In attempting to explain this class of phenomenon with the model, an attempt is made to attribute them to just low-level visual mechanisms. Perceptual filling-effects are characterised by illusions that are thought to be purely driven by the presence of edges in the scene, and hence their existence attributable to the specific mechanism(s) responsible for filling-in the brightness of a surface, based on this edge-derived information. As the model explicitly contains a mechanism for filling-in, then it ought to account for these effects as well. The temporal dynamics of the filling-in process have only been modelled once before (Grossberg & Todorovic, 1988), and are beyond the remit of most models of brightness. By applying the model in this area an extra dimension of validity for the model may be obtained. Within each class, as many different examples of the phenomena as known to the author were tested, to prove that the model can account for the class of effect, and not just one particular configuration.

Within each section, the predictions of the model are made in the context of the psychophysical experiments that are used to investigate the phenomena; the stimuli processed by the model are identical to those used in the corresponding psychophysical investigations, and where possible, all parametric variations on the stimuli have also been processed by the model. The parametric studies have been incorporated into the modelling for two reasons: it allows for a more rigorous test of the model's account of any given phenomenon; and where successful, it then provides a detailed explanation of how and why the phenomenon occurs in its many guises.

The link between psychophysics and the output of the model are made by comparing the brightness of designated test patches within the stimuli: as estimated by participants in an experiment, and as predicted by the model. Psychophysics which investigate brightness phenomena usually require the participant to perform one of two tasks: either adjusting the luminance of a test patch until it matches in brightness with a control test patch, thus directly measuring the difference between the luminance and brightness of the test patch (for example, see Rudd and Zeemach (2005)); or, by identifying the brightness of a test patch within the stimulus, with one in an array of test patches of varying luminance, whose absolute luminance has been predetermined together with that of the the test patch under consideration, hence indirectly measuring the difference between luminance and brightness (for example, see Spehar, Clifford, and Agostini (2002)). These measurements are linked with the predictions of the model by defining the predicted brightness of a test patch to be the mean model output in the test patch region. This makes an explicit assumption about the information used to make brightness judgements. It is the brightness of the test patch as a whole that is important, and not that of any sub-region(s), which are often found to be lighter and darker, both psychophysically and computationally (Blakeslee & McCourt, 1999). However, as the vast majority of psychophysics measures the brightness of entire test patches, it is felt that this is the best way to proceed in linking experiment to theory.

Static Lightness and Brightness Illusions

Static lightness and brightness illusions are defined as those illusions deriving from stimuli comprised of an ensemble of regions (or surfaces) of varying luminance, and which are present from the onset of a stationary stimulus, and remain present regardless of eye saccades and the duration of presentation.

These illusions can be divided into three groups: those in which a test patch's brightness moves in the opposite direction to the luminance of its surround, increasing the perceived contrast at its edges (and referred to as *simultaneous brightness contrast* effects); those in which simultaneous brightness contrast effects are amplified or suppressed depending on the interpretations that can be ascribed to the origin of the luminance under consideration, usually as a consequence of manipulating the geometrical configuration of the stimulus, with different variants of the same stimulus corresponding to different perceptions of the scene as a whole (*lightness effects*); and finally, those in which the brightness of a test patch is assimilated with the overall luminance of its surround (and referred to as *assimilation effects*).

Simultaneous Brightness Contrast Effects

Figure 5.1 demonstrates the canonical simultaneous brightness contrast (SBC) effect (initially described in Chapter 1). The brightness of the test patch on the left is found to be darker than that on the right, despite the luminance of both being equal to that of the grey background (which is set to be equal to the mean luminance of the inset stimulus). Hence, an increase in contrast is perceived at the boundaries of the test patches and their surrounds (Helmholtz, 1887; Hering, 1964; Goldstein, 2003). The model's output for this SBC stimulus is illustrated in Figure 5.2 (and was also discussed in Chapter 4, in terms of the ILC model's contrast mechanism). The mean brightness for the left test patch is calculated at 40.3% of the brightness of white in the scene, and for the right test patch 59.8%, clearly predicting the phenomenon's existence. In addition to the difference in test patch brightness, the model also makes a more subtle prediction: that each test patch's surround

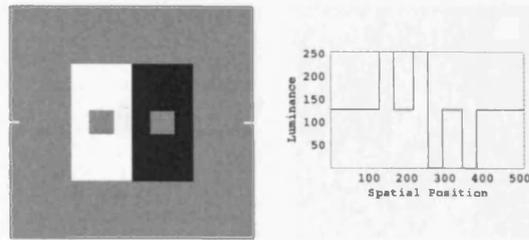


Figure 5.1: An example of the simultaneous brightness contrast stimulus used (left), and its cross-sectional luminance profile taken as indicated by the position of the markers (right). The width of the two grey test patches are 1° , the size of the inset stimulus 5° , and the entire stimulus 10° .

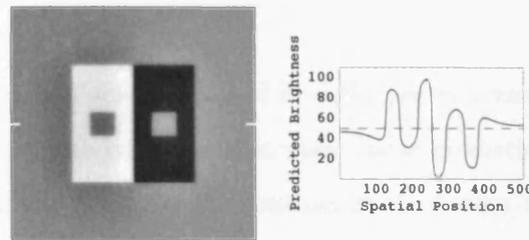


Figure 5.2: The model's output for the simultaneous brightness contrast stimulus shown above (left), and its cross-sectional predicted brightness profile taken as indicated by the position of the markers (right).

exerts an influence on the grey background. This is easily observed on the predicted brightness cross-section in Figure 5.2, where on either side of the inset stimulus the predicted brightness of the background is either suppressed or amplified relative to mean brightness (above the 5% – 7% JND threshold), depending on the original luminance of the surround. This supplementary SBC effect is also, albeit to lesser extents, predicted by the ODOG (Blakeslee & McCourt, 1999) and NSS models (Dakin & Bex, 2003), but to the author's knowledge this has yet to be verified experimentally. Furthermore, it should be noted that the parts of the surrounds closest to the centre of the stimulus are predicted to have a significantly amplified brightness relative to their corresponding parts on the opposite side of the test patch, a prediction which is in direct conflict with the anchoring theory of lightness (X. Li & Gilchrist, 1999), which would predict that the highest luminance in the stimulus (in this case the entire white surround) should always be seen as the global white. This

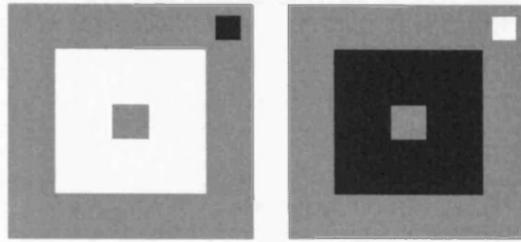


Figure 5.3: An example of the SBC stimuli used to test the influence of the test patch and surround areas on the magnitude of the effect. These are essentially the same as those illustrated in Figure 5.1, but with each test patch and surround pair constituting a separate stimulus. All surrounds were fixed at 6° in height and width, while the size of the test patches were varied.

issue will be clarified when discussing Rudd's brightness matching experiments (Rudd & Zeemach, 2005). It should also be noted that these latter predictions preserve the contrast relationships in the original stimulus; the contrast at the border between background and surround is approximately half that of the contrast at the boundary between the surrounds.

In addition to the above, the size of the test patches were also varied parametrically in an attempt to predict the behaviour of the effect as the areas of the surrounds were altered with respect to that of the test patches. In a brightness matching experiment where the stimuli were similar to those presented in Figure 5.3, Yund and Armington (1975) varied the sizes of the test patches and the surrounds from 1° - 20° , in an attempt to establish how the spatial scale of the stimuli influenced the effect. It was shown that the effect has an approximately linear dependency on the area of the surround, relative to the test patch. Using the stimuli configurations demonstrated in Figure 5.3, the width of the test patches were varied from 0.375° - 3° in octave intervals, and the models output calculated. It was not possible to use larger test patches due to the size of the overall stimulus exceeding what can be processed by the current incarnation of the model. Figure 5.4 plots the predictions for the strength of the effect (the predicted brightness of the test patch on the dark surround less that on the bright surround), against the area ratio of the surround to the test patch, together with the appropriate part of the empirical trend-line derived by Yund and Armington (1975).

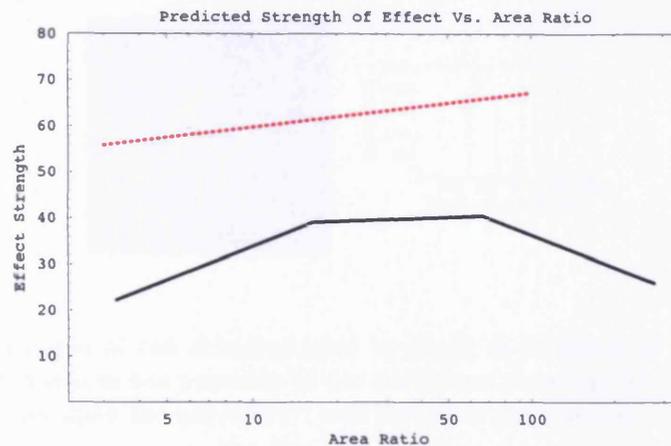


Figure 5.4: A plot of the SBC effect strength versus the area ratio of the surround to the test patch. The dark solid line represents the model's output, and the lighter dashed line, the trend-line from the same experiment by Yund & Armington (1975).

It can immediately be seen that the magnitude of the effect strength is found to be larger than what is predicted, and that the strength of the effect is not predicted to be linear as the trend-line would suggest.

However, for area ratios > 100 (or a test patch width of 0.75°), the model predicts that the brightness of the test patch should start to decrease, as the assimilation mechanism begins to have an affect. As all the data collected on this effect has refrained from using stimuli with area ratios this large, this prediction of the model cannot yet be verified.

Given the nature and polarity of SBC, it is reasonable to question whether it is a universal effect of a region's surrounding luminance configuration, or whether it is only manifest under certain conditions. In particular, in order for SBC to occur, is the luminance of the test patch constrained to lie below that of the highest luminance in the stimulus? If the test patches have the highest luminance in the stimulus, that is, if it's luminance is incremental with respect to that of the surrounds as in Figure 5.5, then the anchoring theory of lightness would predict that their brightness would remain unchanged by the surrounds, as the test patches would both be interpreted as white in both global and local anchoring frameworks,

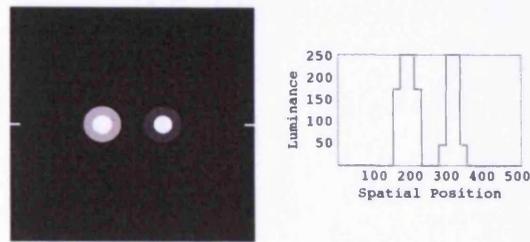


Figure 5.5: An example of the stimulus used by Rudd & Zeemach to test the anchoring theory of lightness when in the presence of the conditions necessary for SBC to occur. The diameter of the inner discs are set at 0.7° , and the outer discs at 1.4° .

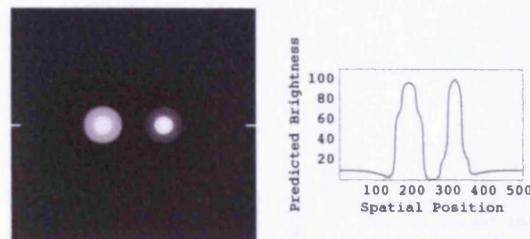


Figure 5.6: The model's output for the stimulus shown above. The predicted brightness of the test patch on the right is marginally higher than that on the left, and subsequently violates the anchoring theory of lightness.

and hence should both appear *identically* as white. Rudd and Zeemach (2005) tested this explicitly using the canonical brightness matching paradigm and the stimulus illustrated in Figure 5.5. It was found that the anchoring theory of lightness was not upheld in this situation; the brightness of the test patches were found to be modulated by the luminance of their surrounds in the same fashion as in the canonical SBC effect. Given that the model already appears to violate the anchoring theory of lightness, it was tested on the stimulus depicted in Figure 5.5, and the result can be seen in Figure 5.5. Although only a small and suppressed SBC effect, the predicted brightness of the test patch on the right is higher than that of the left, with a net difference of 2.1% (it should be noted that the JND for the stimulus is 0.2% – 0.3% as the mean luminance is 2%). Thus, the model is in agreement with the findings of Rudd and Zeemach (2005), who found the greatest magnitude of the effect to be $\approx 8\%$, for the optimal stimulus configuration.

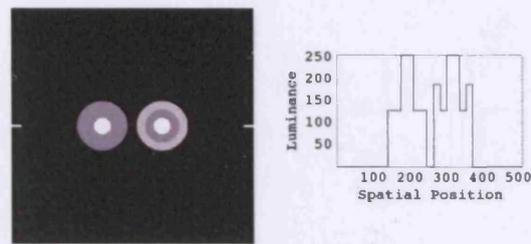


Figure 5.7: An example of the stimulus used by Rudd & Zeemach to test the effect of distant edges and surrounds on SBC. The diameter of the inner discs are set at 0.7° , the outer discs at 2.1° , with the inner surround on the right at 1.4° . The luminance of the test patches' immediate surrounds are identical.

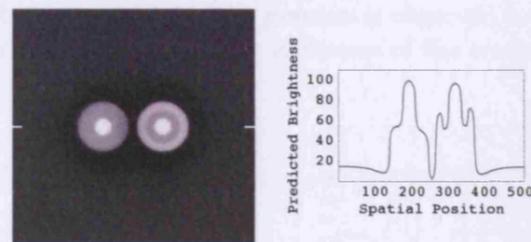


Figure 5.8: The model's output for the stimulus shown above. The predicted brightness of the test patch on the right is marginally lower than that on the left, indicating that distant edges and surrounds can also induce SBC.

An interesting artifact of the transformation from model units to percentage of the white in the scene, can be seen in Figure 5.6 in the regions away from the stimuli that ought to be black, but are in fact brighter. In the regions immediately around the discs a SBC effect occurs to darken the adjacent areas, and assign values that are 'blacker than black'. It is clear from this that modifications of the model to place a threshold on the lowest possible value might need to be introduced, unless these effects are to be considered as actual reality, which is not the opinion of the author

Using the same paradigm, Rudd and Zeemach (2005) extended their test of SBC's universality by trying to induce the effect with regions of luminance (and edges) that lay beyond

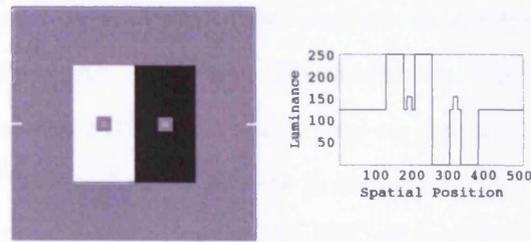


Figure 5.9: An example of the kind of stimulus used used to demonstrate the SBA phenomenon. The stimulus is identical to that used to demonstrate canonical SBC, with the exception that the test patches are not the larger squares closest to the large surround, but are in fact two smaller squares set into the original grey squares, but with a slightly higher luminance. The size of the outer grey squares is 0.75° , and the size of the inner grey test patches is 0.375° . The brightness of the test patches is observed to assimilate to that of the outer squares, whose brightness is under the influence of the canonical SBC effect.

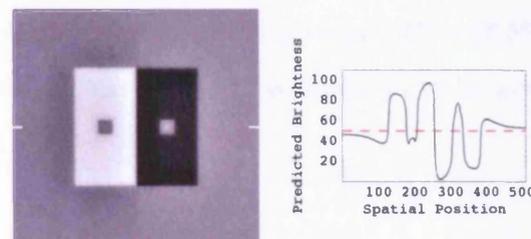


Figure 5.10: The model's output for the stimulus shown above (left), and its cross-sectional predicted brightness profile taken as indicated by the position of the markers (right).

the test patches immediate surround. Figure 5.7 demonstrates the type of stimulus used to probe for such effects: the immediate surrounds of both test patches are identical in terms of luminance, but the test patch on the right has an additional outer surround whose luminance differs from that of the surround. It was found that distant edges do indeed affect the brightness of the test patch, in the much the same way as the canonical SBC effect. Figure 5.8 illustrates the model's output for the same stimulus, which correctly predicts that the right hand test patch is darker than that on the left, by 2.9%. Rudd and Zeemach (2005) found the greatest magnitude of this effect to be $\approx 7\%$ for the optimal stimulus. This phenomenon is closely related to that of simultaneous brightness assimilation (SBA), an example of which is demonstrated in Figure 5.9. Here, the canonical SBC stimulus of

Figure 5.1 is modified with the addition of two smaller test patches within the original test patches, and of a slightly higher luminance. The brightness of the new test patches moves in the same direction as the SBC effect which acts on the original grey squares set into the large surrounds. That is, their brightness is assimilated into that of their immediate surround, which is itself under the influence of a SBC effect (Shapley & Reid, 1985). Figure 5.10 illustrates the model's output for this stimulus, which encapsulates this 'action-at-a-distance' effect: the left test patch has a brightness of 42.1%, while the right test patch has a brightness of 73.4%, thus predicting that the effect is also asymmetric with respect to the luminance of the surround. Using a similar stimulus, Shapley and Reid (1985) discovered that there was a linear relationship between the magnitude of the effect (C) and the difference in brightness between the slightly larger squares that the test patches lie within (dB), such that $C = 0.45dB$. For the model's output, $dB = 36.8\%$ and $C = 31.3\%$, which implies that $C = 0.85dB$. Thus the effect is predicted to be greater than what is found experimentally.

There are many other phenomena that may be considered as falling under the umbrella-term of simultaneous brightness contrast effects, especially when considered in the context of the model's account for their existence. Figures 5.11, 5.13, and 5.15 demonstrate some more well known phenomena of this type: Mach Bands, Vasarely's Pyramid, and Chevreul's Illusion. Traditionally, these phenomena have been attributed to the output of isotropic centre-ON / surround-OFF receptive fields of retinal ganglion cells (Adelson, 2000), in much the same way as the SBC has (as discussed in Chapter 2). However, here we offer an alternative explanation derived from the model (based on the the induced contrast that results from erroneous reconstruction of the *real* image, as discussed in Chapter 4). The aim here will *not* be to replicate parametric studies on these effects, but to demonstrate how the model can yield significant accounts for contrast effects other than SBC.

In Machs' Bands²³, the brightness of the regions at either side of the ramp are amplified with respect to that of the surfaces to which they are physically part of, thus producing light

²³which, it should be noted, is an essentially 1-D effect derived from 1-D stimuli

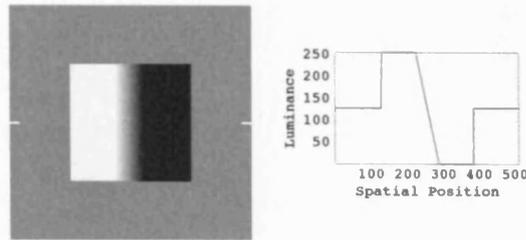


Figure 5.11: An example of the stimulus used to elicit the Machs' Bands phenomenon. The width of the inset stimulus is 5° , and the width of the ramp is 1.2° .

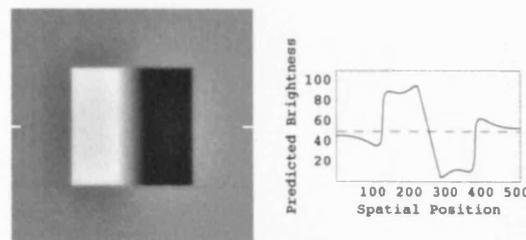


Figure 5.12: The model's output for the Mach Bands stimulus above (left), together with the mean predicted brightness cross-section, calculated over the range of the inset stimulus height (right). Machs' Bands are clearly visible at either side of the ramp.

and dark bands at the high and low end of the ramp (Ratliff, 1965; Ross, Morrone, & Burr, 1989). Figure 5.12 illustrates the model's output for the Mach Bands stimulus, with the mean predicted brightness cross-section clearly indicating the presence of Mach Bands²⁴, with a magnitude of $\approx 8\%$ (significant). The model generates the effect as a direct result of the same mechanism responsible for SBC. Consider the ramp as playing an analogous role to the test patch within a SBC stimulus. At the border between surface and ramp there is a relatively rapid change in luminance that facilitates an amplification in brightness, much the same as that predicted at the border between test patch and surface in Figure 5.1. However, the mean brightness of the ramp is not altered as there is an approximately equal and opposite shift coming from the opposite side. Thus, only the brightness at the borders is amplified.

²⁴it should also be noted that as the luminance of the light surface in the stimulus is that of white, and that the Mach Band has a higher brightness than this surface, the model once again demonstrates violation of the anchoring theory of lightness.

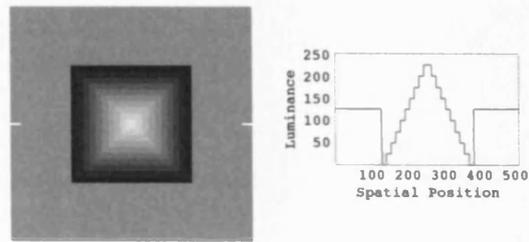


Figure 5.13: An example of the stimulus used to generate Vasarely's Illusion. The width of the inset stimulus is 5° , and the width of each step is 0.25° .

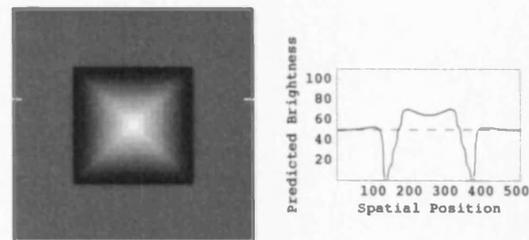


Figure 5.14: The model's output for the Vasarely's Pyramid stimulus above (left), together with the mean predicted brightness cross-section, calculated over the entire range of a step as indicated (right). The corner regions of the step are accurately predicted to be brighter.

In Vasarely's Pyramid, a stack of ever-decreasing squares, of ever-increasing luminance, are perceived as being brighter at their corners than at their mid-sections (Vasarely, 1970; Adelson, 2000). The stimulus and the corresponding model output are illustrated in Figures 5.13 and 5.14 respectively. The mean predicted brightness cross-section for one of the 'steps' clearly indicates that the brightness is higher towards the edges, or corners, by $\approx 8\%$. The model's account for this is much the same as that for Mach Bands: at every edge there is induced contrast as a result of the above surface having a higher luminance, and vice-versa. However, at corners this effect is compounded; it is significantly denser. Hence, corners are perceived as having a greater brightness than the mid-sections.

Chevreul's Illusion demonstrates how a homogeneous surface may appear inhomogenous. It is comprised of contiguous surfaces of increasing luminance; a luminance staircase. For

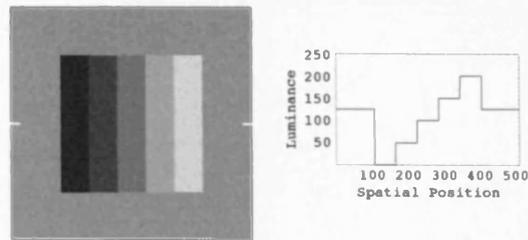


Figure 5.15: An example of the stimulus used to generate Chevreul's Illusion. The width of the inset stimulus is 5.25° , and the width of each step is 1.25° .

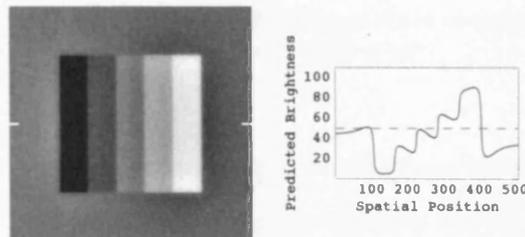


Figure 5.16: The model's output for the Chevreul's Illusion stimulus above (left), together with the mean predicted brightness cross-section, calculated over the range of the inset stimulus height (right). The brightness of the central surfaces are predicted to be inhomogeneous.

all but the first and last surfaces, the brightness of each surface is inhomogeneous, with the brightness towards the left-hand edge always being greater than that on the right (Chevreul, 1890; Morrone, Burr, & Ross, 1994). The model's output for such a stimulus is illustrated in Figure 5.16, and the mean predicted brightness cross-section clearly demonstrates this effect, with the brightness on the left-hand of each step predicted to be $\approx 9\%$ brighter than the right-hand side. Once again, the model generates the effect through induced contrast at boundaries, in exactly the same manner as Mach Bands are generated. However, this does not explain why the effect is not predicted for the first and last surfaces. To understand why this is, it is useful to consider the raw outputs for the model's largest and smallest scale channels, illustrated in Figure 5.17. To the large scale, the inset stimulus is treated as ramp with a rapid onset from the background, which produces very well defined peaks for the first and last surfaces. The small scale treats the first and last surfaces as it might test patches

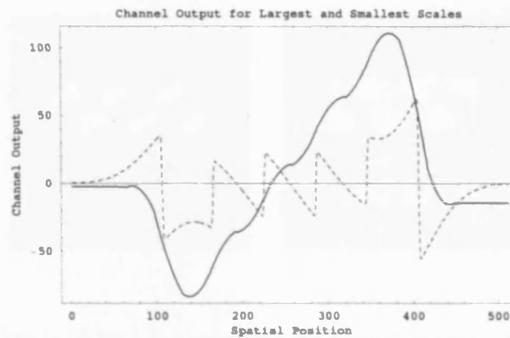


Figure 5.17: The response of the largest and smallest scale channels to the Chevreul's Illusion stimulus. The solid line represents the largest scale channel, $\sigma = 16$ (or 0.31°), and the dashed line the smallest scale channel, $\sigma = 1$ (or 0.02°).

in the canonical SBC stimulus, as the polarity of the edges on either side of each surface are identical. But, magnitude of the surrounding luminance is not identical on both sides, and it still predicts a ramp in brightness pointed towards the direction of greater brightness. However, this effect is dwarfed by the response from the largest scale and its presence in the resultant output is negligible. This also demonstrates the continual interplay between the assimilating effects of the larger scales and the contrast inducing effects of the smaller scales.

Lightness Illusions

When information deriving from a specific interpretation of a scene is used to infer the nature of a region's luminance, and thus compute the underlying surface reflectance of a region, it is often said that higher-level lightness computations are being made (see Chapters 1 and 2 for more). It is thought that such computations can have an affect on brightness, and some brightness illusions, as the visual system attempts to remain lightness constant. Such illusions are here referred to as 'lightness illusions'. It is not clear that what such illusions represent is indeed evidence of higher-level processing, and some illusions have been predicted by low-level models of brightness (for example, Blakeslee and McCourt (2001)). Here, the model (an essentially low-level model of brightness), is applied to four different types of such illusion with the aim of trying to differentiate where higher-level inferential

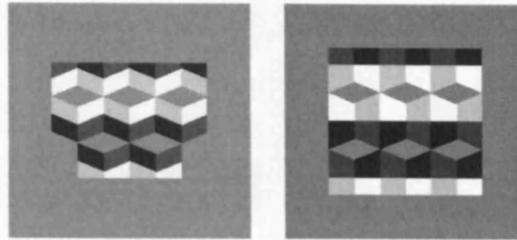


Figure 5.18: Adelson's 'Wall of Blocks' stimuli. The grey diamonds in both inset stimuli are of the same luminance as the background and are framed by surrounds of either high or low luminance, the areas of which are equal between stimuli. The width of the inset stimuli is 6.25° . The width of the diamonds is 2.1° , and the height is 0.9° . A SBC effect is perceived in the diamonds, but is greatly amplified in the right stimulus.

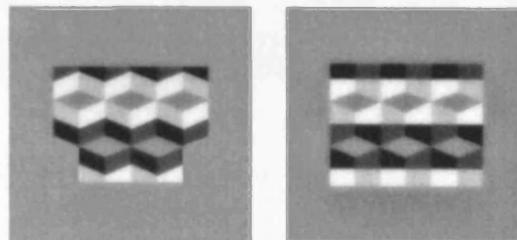


Figure 5.19: The model's output for Adelson's 'Wall of Blocks' stimuli.

processing may not be necessary to provide an accurate account for the phenomenon, and in what situations it might be required.

Figure 5.18 depicts an adaptation of the 'Wall of Blocks' stimuli as used by Adelson (1993) to demonstrate the affect of scene interpretation on brightness (and originally discussed in Chapter 1). In both stimuli, a SBC effect was found in the grey diamonds (the whole ones), which are the designated test patches. However, for the stimulus on the right, the strength of the effect was found to be over 300% greater than the one on the left (Adelson, 1993). While the low-level nature of the original SBC effect is not in doubt, the change in the strength of the effect is. The model's output for these stimuli is illustrated in Figure 5.19, and the effect's strength²⁵ was calculated at 6.3% for the left stimulus, and 5.3% for the

²⁵ defined here are the mean brightness of the test patches with the darker surrounds, less that of the mean brightness of the test patches with the lighter surrounds.

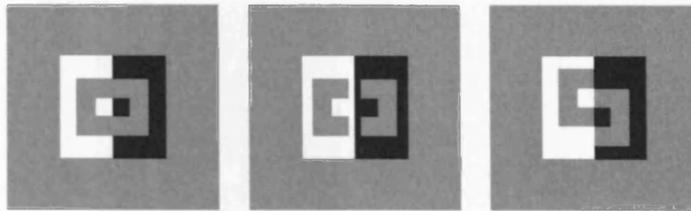


Figure 5.20: The Kofka's Rings triptych (A, B, and C, going from left to right). The width of each inset stimulus is 5° , the width of the 'ring' is 3.34° , and the height 3° .

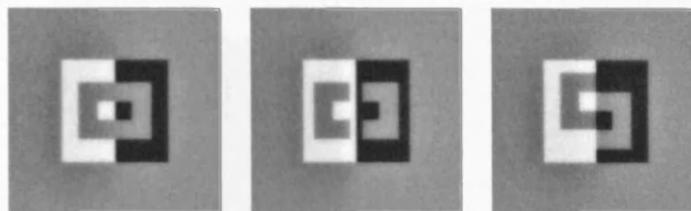


Figure 5.21: The model's output for the Kofka's Rings triptych.

right (both significant SBC effects). Thus, the model predicts that the effect should get weaker, and not stronger. Clearly, the low-level mechanisms described by the model cannot account for this phenomenon. The higher-level explanation offered by (Adelson, 1993) is indeed more favourable: for the stimulus on the right, the grey diamonds are interpreted as existing beneath light or dark transparencies. But, as they both have the 'same' luminance the diamonds under the darker transparency must have a higher surface reflectance, and hence appear brighter as the visual system attempts to remain lightness constant. On the other hand, for the stimulus on the left there is no such transparency interpretation and subsequent lightness computation, and hence only the residual SBC effect is observed.

Similarly, the model fails to capture the modulation of the SBC effect as seen in Kofka's Rings, demonstrated in Figure 5.20. In this phenomenon, the canonical SBC effect is observed in A, in the two different halves of the square-ring. In B, the two halves are separated and the strength of the effect increases. When the spatial configuration of the rings is altered, such that the half of the ring on the black surround interfaces with the white surround

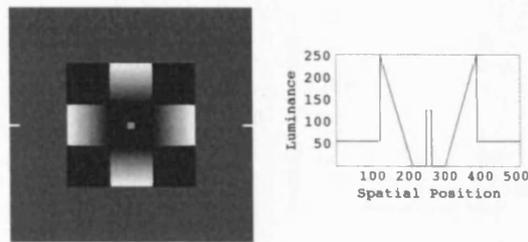


Figure 5.22: An example of the stimulus used to generate Agostini's amplification of SBC. The width of the inset stimulus is 5.3° and the width of the test patch is 0.28° .

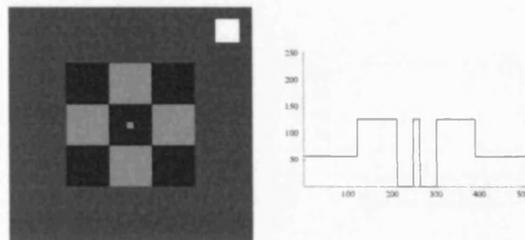


Figure 5.23: An example of the stimulus used as a control for Agostini's amplification of SBC. The width of the inset stimulus is 5.3° and the width of the test patch is 0.28° .

(and vice-versa), the effect strength increases yet further (Kofka, 1935; Adelson, 2000). The model's output is illustrated in Figure 5.21, and the effect strength was calculated as being 10.2% for all stimuli A-C. That is, the model does not predict perception. Once again, the higher-level interpretation is more appealing: In A, the ring is perceived as being a single entity existing above the surround. However, for B and C, the interpretation is of two separate entities beneath a light and dark transparency, but with the same 'luminance', hence the half of the ring on the dark background must have a lighter surface reflectance (and vice-versa). Hence, in the interests of remaining lightness constant, it's brightness is greater. The fact that the strength of the effect is greater for C than B, may be accounted for by appealing to Gestalt-like grouping cues, which would attach the half of the ring on the dark surround to the light surround (and hence it's lightness), and similarly for the other half of the ring (Adelson, 2000).

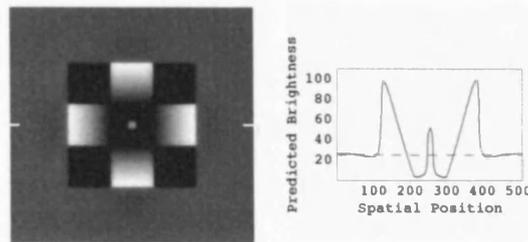


Figure 5.24: The model's output for the stimulus used to generate Agostini's amplification of SBC.

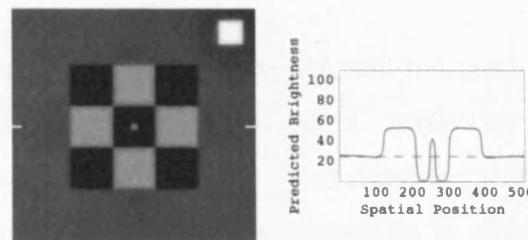


Figure 5.25: The model's output for the stimulus used as a control for Agostini's amplification of SBC.

Whereas the model failed to provide low-level accounts for the Wall of Blocks and Kofka's rings phenomena, it does provide an account for two other supposed lightness illusions. Figures 5.22 and 5.23 demonstrate a new variant of SBC, in which the addition of a luminance ramp in 5.22 yields a SBC effect that is approximately three times the strength of that found in 5.23, which has the same mean luminance surrounding the grey test patches (Agostini & Galmonte, 2002). One explanation that the authors give for this amplification of the effect, is that the luminance gradient is interpreted as a change in illumination toward the centre of the stimulus, which places the test patch in some degree of shadow with respect to the stimulus in Figure 5.23, and hence for the test patch to retain its brightness it must have a very light surface reflectance. In the interests of lightness constancy, it is then perceived as being brighter. As this interpretation cannot be ascribed to the stimulus in Figure 5.23, the obvious inference is made about the test patch's surface reflectance, and its brightness remains un-amplified. Figures 5.24 and 5.25 illustrate the model's output for both of these stimuli, with the test patch of the gradient variant being greater in brightness than that

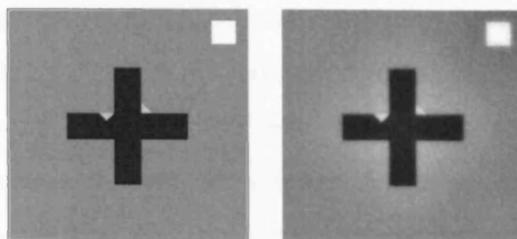


Figure 5.26: An example of a Wertheimer-Benary Cross stimulus (left) and the corresponding model output (right). The width of the cross's arms are 5° in length and 1° in width.

of the control. The mean predicted brightness for each test patch was calculated at 42.8% for the gradient variant, and 31.3% for the control. While this is far from the threefold increase that was initially reported, it is still a very significant shift in the magnitude of the effect²⁶. More importantly, this cannot be accounted for by appealing to the responses of isotropic centre-ON / surround-OFF receptive fields of retinal ganglion cells, whose output is proportional to the luminance of the test patch's immediate surround, the mean of which is identical in both versions of the stimulus. Through an understanding of how the model works by inducing contrast at borders, it is possible to account for this effect as follows: in the control version, the contrast between the central dark square (on which the test patch is mounted) and the surrounding squares is greater at the borders than that in the version with the ramps. This causes the dark square itself to experience a SBC effect making it darker. In accordance with the phenomenon of simultaneous brightness assimilation, this reduction in brightness will also be passed onto the test patch, regardless of the SBC effect in the opposite direction, that it experiences as a result of being mounted on a dark surround.

Finally, Figure 5.26 demonstrates the Wertheimer-Benary Cross and the model's output for this stimulus. In this stimulus, the isosceles triangle set into the cross appears brighter than that set outside the boundaries of the cross, regardless of the fact that the mean luminance of both surrounds is identical (Benary, 1924; Blakeslee & McCourt, 2001). Accounts for the effect once again appeal to notions of transparency and 'belonging'. The triangle on

²⁶it should be noted that the same effect, but in the opposite direction, was observed using stimuli that were inverted in luminance.

the cross may be interpreted as being underneath a dark transparency, but as it has the same luminance as the other triangle, its surface reflectance must be lighter, and hence it is perceived to be brighter in the interests of achieving lightness constancy. Alternatively, it has been said that the triangle on the cross is perceived as belonging to the cross, and hence it is processed within this context and contrasts only with the cross. Similarly, the triangle on the background is perceived as belonging to the background, is processed within this different context, and hence contrasts with the background. However, the model predicts the mean brightness of the triangle on the cross to be 5.8% brighter (significant), than the triangle on the background. Thus, it is possible to reject the higher-level accounts in favour of simpler low-level mechanisms. It is unclear exactly 'how' the model yields this effect. It may be that the contrast effect is greater for the triangle on the cross as directly adjacent there is a greater expanse of dark cross (especially to the right), than there is for the triangle on the background. Hence, as SBC is dependent on the area of the surround, the effect is larger for the triangle on the cross.

Assimilation Effects

One of the aims of this work to is provide biologically feasible accounts of both SBC and White's effect concurrently. It has already been shown (in Chapter 4), that the model encapsulates this behaviour. The purpose of this section is to test the model on the many parametric variations of White's effect, with an aim to demonstrating that the the phenomenon in its entirety is accounted for by the model.

Figure 5.27 demonstrates a typical White's effect stimulus (White, 1979; Blakeslee & McCourt, 1999). Recall that the effect's significance arises from the fact that an explanation based on our knowledge of SBC would predict that test patches surrounded by more dark than light (those on the white bars), should appear brighter than their counterparts surrounded by more light than dark. However, this is not the case and the effect proceeds in the opposite direction to that predicted by SBC²⁷. As a direct consequence, no straight-forward explanation based on the responses of isotropic centre-ON / surround-OFF receptive fields

²⁷or in the correct direction for SBC along the coaxial bars only.

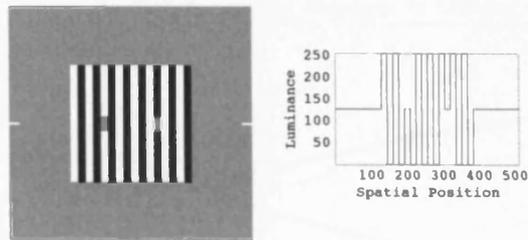


Figure 5.27: An example of a stimulus used to demonstrate White's Effect. The size of the inset stimulus is 5° , the frequency of the inducing grating is $1.6\text{ c}/^\circ$, the width of the test patches is 0.31° , and the height is 0.62° . The luminance of the test patches is equivalent to that of the background, which is set to mean luminance.

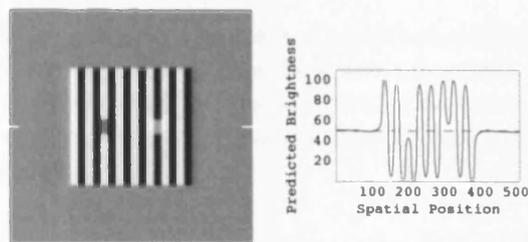


Figure 5.28: The model's output for the White's effect stimulus above (left), and its cross-sectional predicted brightness profile taken as indicated by the position of the markers (right). The test patch on the left can be seen to have a greater predicted brightness than the one on the right.

of retinal ganglion cells is possible. Figure 5.28 illustrates the model's output for such a stimulus, clearly predicting the effects existence in this situation (as discussed in Chapter 4). Quantitatively, the mean predicted brightness for the test patch on the light bar is 42.2% and 58.1% models units for the that on the dark bar. The effect is accounted for by the interaction between contrast and assimilation: each test patch receives the appropriate contrast signals at all surrounding boundaries, but in addition is assimilated to the mean luminance of it's surround by the intrinsic blurring in the reconstruction process.

One of the most well known parameters that determine the magnitude of White's effect is the SF of the inducing grating (White, 1981; Kingdom & Moulden, 1991; Blakeslee & McCourt, 1999, 2004; Anstis, 2004), with the caveat that the aspect ratio of the test patches

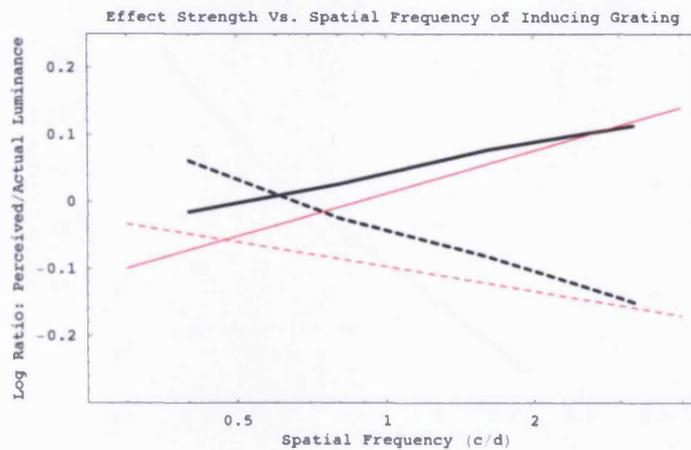


Figure 5.29: A plot of the logarithm of the ratio of predicted brightness to actual luminance, as a function of SF, for White's effect. The dark line represents the models' predictions, and the lighter line the data collected by Anstis (2004). The solid lines represent the values for test patches on the dark bar, and the dashed lines represent the value for test patches on the light bar of the grating.

must be held constant throughout the subsequent variations in spatial scale (Kingdom & Moulden, 1991). From $0.5c/^\circ$ onwards, the strength of the effect is seen to increase with the SF of the inducing grating. At around $0.5c/^\circ$ and below the data is less clear: Anstis (2004) reports data that suggests the effect reverses in polarity at around $0.5c/^\circ$, with the magnitude of this reversal increasing as SF decreases (a plot of this data is reproduced in Figure 5.29 along with the model's predictions). Kingdom and Moulden (1991) also report data which suggests that the effect can be reversed at low SFs, although he reports this as occurring at $0.63c/^\circ$ (for one observer). Blakeslee and McCourt (2004), however, report data which suggests that the strength of the effect does not reverse around the $0.5c/^\circ$ mark, but reaches a minimum that is also maintained for lower SFs. The model's output for White's stimuli consisting of inducing gratings with SFs between $0.5c/^\circ$ - $3.2c/^\circ$ is illustrated in Figure 5.29, plotted concurrently with the data collected by Anstis (2004). The magnitude of the predicted effect is shown to be similar to what is actually observed, unlike the model's predictions for SBC, which are less than what is found psychophysically. This is more so at higher SFs when the two plots intersect, and less so at low SFs (at $0.4c/^\circ$ the test patch on the black bar is predicted by the model to be 10% brighter than shown in the

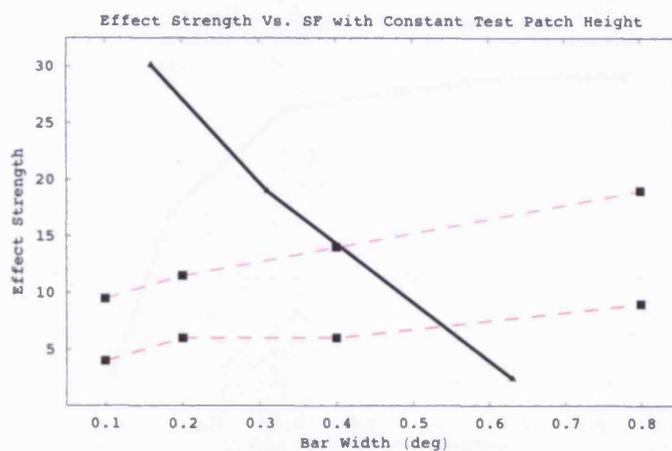


Figure 5.30: A plot of the predicted strength of White's effect, as a function of the spatial frequency of the inducing grating, when the test patch height is held constant at 0.4° . The solid line represents the predicted effect strength, and the dashed lines the data collected for the two observers in the experiment by Kingdom & Moulden (1991).

data). The point at which the effect reverses in polarity and becomes SBC is also predicted to occur at a higher SF than what is shown in the data ($0.6c/^\circ$ as opposed to $0.4c/^\circ$). Even with these differences, it is still fair to claim that the model's interplay between two separate mechanisms of contrast and assimilation appears to capture the behaviour of the data remarkably well.

It should be noted (as discussed in Chapter 4), that the point of reversal for the effect is dependent on the choice of spatial scales in the model. It is entirely possible that an alternative choice of spatial scales would fit this particular data better (although it is likely that this will be at the detriment to the SBC predictions, which will be reduced, and are already predicted to be weaker than the data suggests).

That White's effect is a result of an assimilation mechanism is a moot point. Kingdom and Moulden (1991) conducted experiments with the aim of showing that an assimilation mechanism could not fully account for White's effect. In their first experiment, the height of the test patch remained constant as SF (and thus the aspect ratio) was varied. The motivation

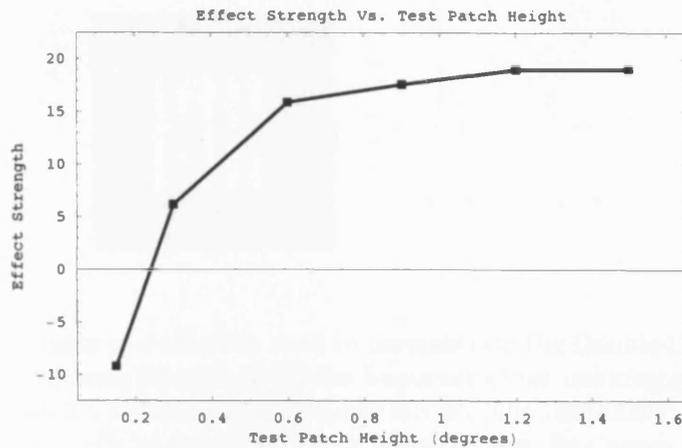


Figure 5.31: A plot of the predicted strength of White's effect, as a function of the test patch height on a White's Effect stimulus with an inducing grating of $1.6c/^\circ$ as in Figure 5.27.

behind this is as follows: If the effect is driven by assimilation across the vertical borders of test patch (as the model predicts), then while the length of the border is held constant and the length of the border with the coaxial bar is shortened (as SF increases), then the magnitude of the effect should increase as the assimilation signal penetrates further into the test patch. Figure 5.30 illustrates the model's predictions for this experiment, in alignment with the explanation just given and in direct conflict with the data of Kingdom and Moulden (1991), who found that the effect *decreased* as SF increased (the data for observers BM and FK are plotted with the model's prediction in Figure 5.30). Furthermore, it was shown that the magnitude of the effect *decreased* when the test patch height doubled. If assimilation across the vertical borders of the test patch is responsible for White's effect, then the magnitude of the effect should increase with the test patch height. Figure 5.31 illustrates the model's prediction for such an experiment, which is in good alignment with the explanation just given, once again contradicting the data collected by Kingdom and Moulden (1991). However, these studies are implicitly replicated in data collected by Blakeslee and McCourt (2004), and with twice as many observers²⁸. No evidence was found to support the findings of the first experiment, and in actual fact the magnitude of the effect

²⁸four as opposed to two.

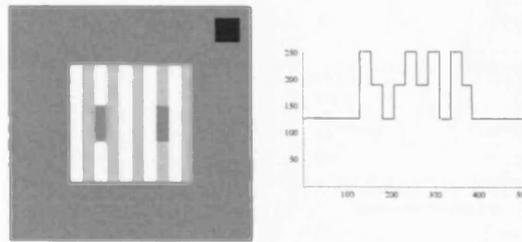


Figure 5.32: An example of a stimulus used to demonstrate the Double-Decrement White's Effect. The size of the inset stimulus is 5° , the frequency of the inducing grating is $1c/^\circ$, the width of the test patches is 0.5° , and the height is 1.5° . The luminance of the test patches is equivalent to that of the background, which is set to mean luminance.

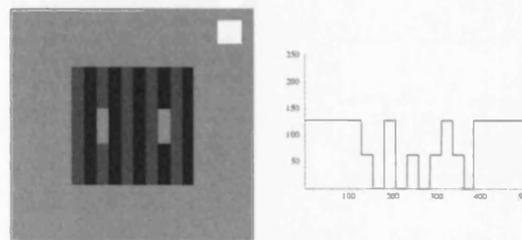


Figure 5.33: An example of a stimulus used to demonstrate the Double-Increment White's Effect. See Figure 5.32 for details about stimulus dimensions.

was found to increase with SF, as predicted by the model's assimilation driven mechanism. For the second experiment, only two of the four participants provided evidence in favour of the original findings, and only for high SF inducing gratings²⁹, rendering the situation unclear.

In addition to the parametric variation of spatial parameters, many other variants of White's Effect have been produced in an attempt to disentangle its source(s). Two of the more profound variants are provided by Spehar et al. (1995). In these versions, the luminance relationships between the test patches and the inducing grating have been altered, such that the test patches constitute either increments to the maximum-, or decrements to the minimum-luminance of the gratings. These two conditions are referred to as the Double-

²⁹the effect was not demonstrated for inducing gratings with SFs lower than $4c/^\circ$.

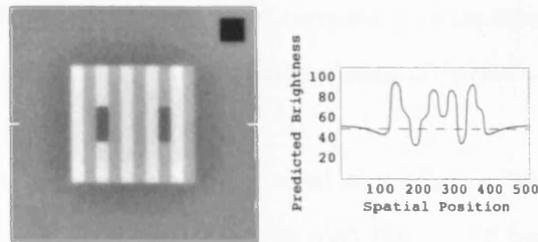


Figure 5.34: The model's output for the Double-Decrement White's effect stimulus (left), and its cross-sectional predicted brightness profile taken as indicated by the position of the markers (right). The magnitude of the effect has been suppressed.

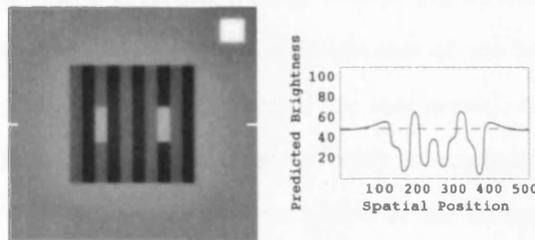


Figure 5.35: The model's output for the Double-Increment White's effect stimulus (left), and its cross-sectional predicted brightness profile taken as indicated by the position of the markers (right). The magnitude of the effect has been clearly suppressed.

Increment (DI) or Double-Decrement (DD) conditions, and are demonstrated in Figures 5.33 and 5.32 respectively. Spehar et al. (1995) found that White's effect was not demonstrated in either DI or DD conditions, and in some cases the effect reversed to give SBC. Table 5.1 lists the model's output for these stimuli, for inducing gratings of $0.8c/^\circ$ and $1.0c/^\circ$. Although a large and significant suppression of the effect has taken place, some degree of White's effect remains as a direct result of assimilation, although the overall strength of the effect is $< 5\%$, and thus probably not discriminable. This minimum residual effect may be accounted for by that fact that in the original experiment an inducing grating of $0.5c/^\circ$ was used when collecting data. The magnitude of White's effect is already known to be very weak for inducing gratings at these SFs (as discussed above), and so the effect of changing the luminance relationships in this way need only be small to nullify, or reverse the effect. Stimuli with inducing gratings of $0.5c/^\circ$ could not be used with the model, as

the model already predicts that the effect is reversed for this stimulus. It is claimed that the model does indeed explain the DI and DD variants of White's Effect.

Another stimulus which demonstrates the reversal or nulling of White's Effect is illustrated in Figure 5.36, and was originally put forwards with the aim of testing the junctions-based explanation of the effect³⁰. The stimulus keeps the area local to the test patch equivalent to that in the canonical White's Effect stimulus, such that any explanation dependent on assimilation ought to still predict that White's effect is present, when in actual fact it is perceived as being either nulled or reversed (Howe, 2001). Figure 5.37 illustrates the model's output for this stimulus. The mean predicted brightness of the test patch is calculated at 48.9% (1.1% darker than the actual luminance of the test patch), clearly predicting that the effect has been nulled. How then does the model, with an explicit assimilation mechanism, predict this effect? Once again, this is a direct result of the interplay between assimilation and contrast. When whole bars are present in the inducing grating, the magnitude of the SBC effect between the light bars and the test patch is relatively small and assimilation dominates. However, when parallel sections of the dark bars are removed, the area of the test patch's surround is increased, thus increasing the size of the SBC effect until it comes to dominate over assimilation. To add weight to this explanation, the number of incomplete bars in Figure 5.36 was varied and the predicted size of the effect calculated. Pairs of bars (one either side of the bar on which the test patch was mounted), were systematically

³⁰the junction structure in this version of the stimulus is identical to that in the canonical stimulus, but the effect is no longer present, an effect which cannot then be attributed to junctions.

Table 5.1: The mean predicted brightness for the test patches in the DI and DD versions of White's Effect, together with the % change in the strength of the effect, with respect to a canonical White's stimulus with an inducing grating of the same SF.

	TP on Light Bar	TP on Dark bar	% Change in Effect Strength
0.8 c/° DD	38.4%	39.8%	75% decrease from control
0.8 c/° DI	61.1%	62.4%	76% decrease from control
1.0c/° DD	22.7%	25.4%	77% decrease from control
1.0c/° DI	74.8%	77.4%	78% decrease from control

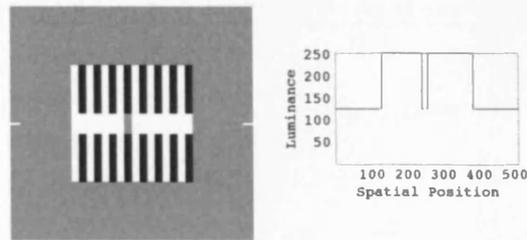


Figure 5.36: An example of the type of stimulus used by Howe to demonstrate how White's Effect may be nulled or reversed. The size of the inset stimulus is 5° , the SF of the inducing grating is $1.6\text{ c}/^\circ$, the width of the test patch is 0.31° , and the height is 0.93° .

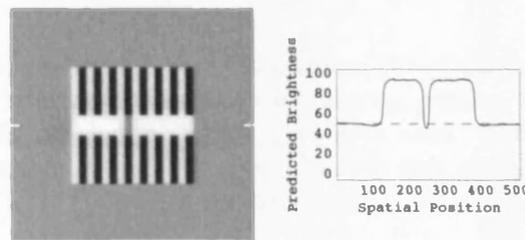


Figure 5.37: The model's output for the Howe's variant on the White's Effect stimulus (left), and its cross-sectional predicted brightness profile taken as indicated by the position of the markers (right). The effect has been nulled.

completed, starting from the perimeter. If the model is indeed displaying an interaction between assimilation and contrast, then as the size of the bar-less area surrounding the test patches decreases, then so should the magnitude of the White's Effect suppression. Figure 5.38 illustrates the results, which are in clear agreement with the known functionality of the ILC model. These results have not been replicated psychophysically and represent a novel prediction of the model.

Finally, in another attempt to investigate the importance of junctions in White's effect, a circular version of the effect without any junctions was constructed, an example of which is demonstrated in Figure 5.39 (Monnier & Shevell, 2003; Howe, 2005). The model's output for this stimulus is shown in Figure 5.40, from which it can clearly be seen that grey test contours surrounded by black contours appear darker than test contours surrounded

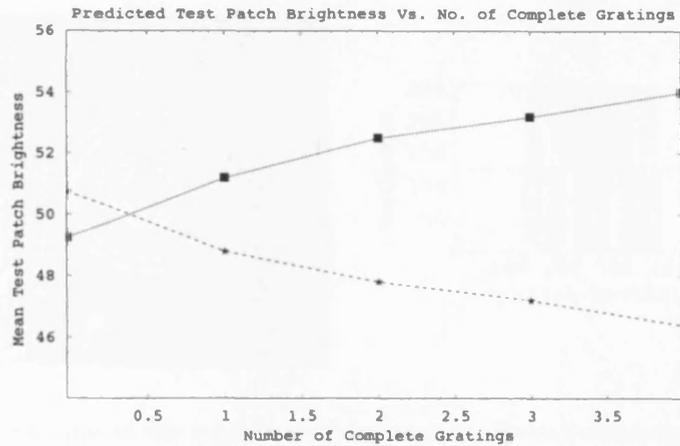


Figure 5.38: A plot of the mean predicted brightness of test patches in White's effect, as a function of the number of complete bars in the inducing grating. Stimuli with an inducing grating of $1.6c/\circ$, similar to that in Figure 5.36 were used. The solid line represent the mean predicted brightness of a test patch on the dark bar, and the dashed line the mean predicted brightness of a test patch on the light bar.

by white contours, in keeping with the traditional version of the effect. Interestingly, an additional prediction of the model is evident in Figure 5.4 - the brightness of the contours flanking the test contours are amplified with respect to the other contours, depending on their original brightness. The dark bars are predicted to be darker, and the light bars lighter, once again violating the Anchoring Theory of Lightness.

Another set of phenomena which have are classified as similar to that of White's effect are those of Todorovic (1997), demonstrated in Figure 5.41. Stimulus A induces the canonical SBC effect and the grey square (of same luminance as the background) appears brighter than the background and its inverse counterpart on a light surround; in B, the addition of four white squares results in the grey 'cross' having equal amounts of grey-white and grey-black border, but the SBC still persists, albeit suppressed; and finally in C, the 'arms' of the cross are extended further than in B, such that the grey cross now has more grey-white border than grey-black, but yet a SBC effect still persists in much the same manner as White's Effect. The model's output for these stimuli are illustrated in Figure 5.42, and the predicted effect strength from A-C was calculated as: 19.8%, 10.7%, and 8.2 % respectively.

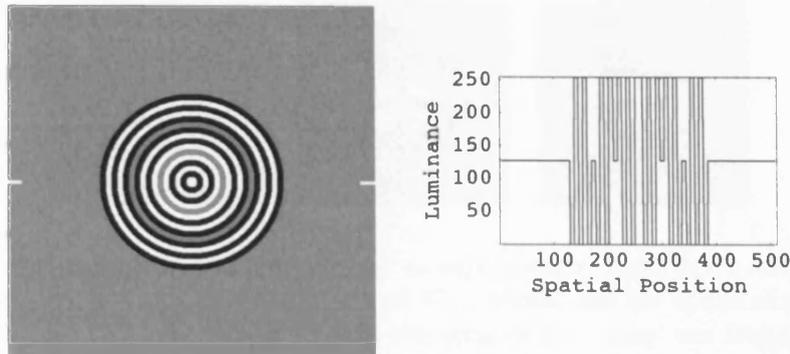


Figure 5.39: An example of the type of stimulus used by Howe to demonstrate how White's Effect is not dependent on junctions within the stimulus. The size of the inset stimulus is 5° , the SF of the inducing grating is $1.6\text{ c}/^\circ$.

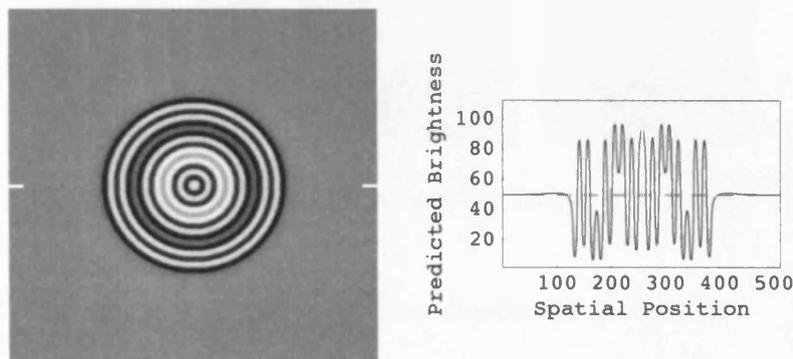


Figure 5.40: The model's output for the circular variant on the White's Effect stimulus (left), and its cross-sectional predicted brightness profile taken as indicated by the position of the markers (right). The effect is clearly visible.

These follow the same trend, but have a far larger magnitude than the psychophysical results reported by (Blakeslee & McCourt, 1999), which when converted to the units used in this report are 6.5%, 4.2% and 2.8%. Despite this, it is still possible to claim with a significant degree of confidence that the model and its contrast and assimilation mechanisms generalise and correctly predict the existence of this assimilation phenomenon in addition to White's effect.

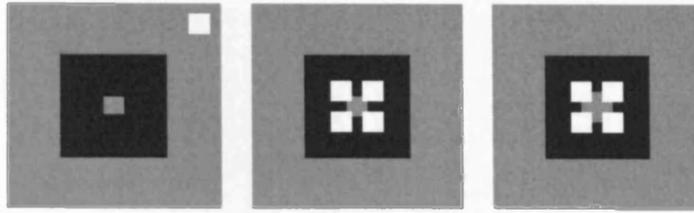


Figure 5.41: An example of the stimuli used to demonstrate Todorovic's effect (A, B, and C, from left to right). The inset stimuli are all 5° in width, and the width of all the squares in A and B is 1° . C differs from B in that the arms of the 'cross' are longer. Versions of these stimuli with the luminance inverted (black to white and white to black), were also processed so that the predicted effect strength could be calculated.

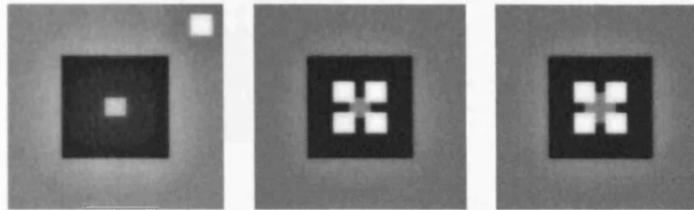


Figure 5.42: The model's output for Todorovic's stimuli..

Perceptual Filling-In and Edge-Driven Surface Perception

The term filling-in is an umbrella term describing a group of effects where a region of the visual field is 'filled-in' with brightness values that are not present in the original image, either by design of a stimulus, or due to lack of input at the retina either in the blind-spot at the optic nerve, or in a scotoma (real or artificial). It is likely that the mechanisms that mediate these differing effects are themselves different, although all three are pertinent to surface perception as a whole (Komatsu, 2006). Currently, only the effects related to illusory brightness values between edges are of interest, as the model has not explicitly been designed to operate in any other situation. The difference between these and the lightness and brightness illusions discussed above is subtle: in the latter the surfaces are usually well defined and explicitly represented, whereas in the former they are not, and the phenomenon often results from the explicit inference of the surface.

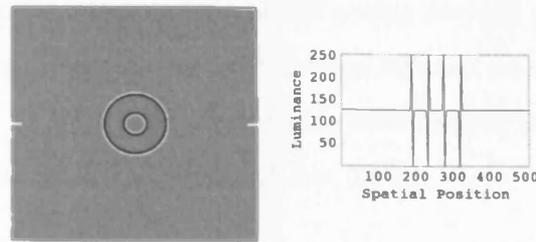


Figure 5.43: An example of the type of CCOB stimulus as used by Dakin and Bex (2003). The diameter of the inner disc is 0.64° and the diameter of the outer disc is 1.7° .

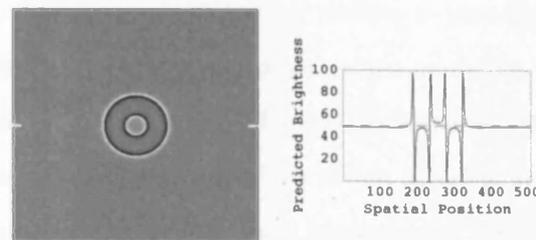


Figure 5.44: The model's output for the CCOB stimulus above (left), and its cross-sectional predicted brightness profile taken as indicated by the position of the markers (right). The effect is clearly visible and the centre disc is predicted to be lighter than the surrounding annulus.

The Craik-Cornsweet-O'Brien Effect

One of the most famous examples of such a phenomenon is the Craik-Cornsweet-O'Brien (CCOB) effect (O'Brien, 1959; Craik, 1966; Cornsweet, 1970), an example of which is demonstrated in Figure 5.43. With the exception of the circular edge profiles, the luminance of the entire stimulus is uniform. Yet, the central disc, which is surrounded by an edge which goes from dark to light, appears to be brighter than the surrounding annulus, which is defined by two concentric circular edge profiles that go from light to dark into the annulus³¹(Moulden & Kingdom, 1990; Dakin & Bex, 2003). Hence, the brightness of the surfaces enclosed by the circular edges has been explicitly filled-in based on the information provided by the edge. Figure 5.44 illustrates the the model's output for the the stimulus and it can be seen that the annulus is predicted to be darker than the central disc - when measured, this difference is predicted to be 7.7% (significant). This is a direct result

³¹it should be noted that polarity of the effect is dependent on the the polarity of the edges, which if reversed would also reverse the direction of the effect.

of assimilation between the edges and the neighbouring surface, and as such the effect is predicted to persist even as the number of iterations, or time, increase indefinitely. This is in accordance with the findings from experiments investigating the temporal evolution of the CCOB effect, as conducted by Dakin and Bex (2003).

It should be noted that the circular edge profiles were created from filtering two concentric black and white discs³² with a Laplacian of Gaussian filter with scale $\sigma = 17$ pixels (or 0.33°). The motivation for using circular edge profiles is because they enclose the entire stimulus area and do not require any additional edges to define the surface(s) where the effect is present. Hence, purely the CCOB effect is being studied, and no edge-derived contrast effect is present. The Laplacian of Gaussian filtering produces edges with equal amounts of light and dark, and hence the mean luminance of the edge is neutral and cannot affect a significant contrast effect on the surrounding area. This latter point is a criticism of the types of CCOB stimuli used by Moulden and Kingdom (1990). In their studies into the effect, they made use of circular edge profiles with large light or dark gradients constituting the edge profiles. Given how the model is known to work, it was felt that any effect attributed to these stimuli would be as a direct result of contrast between the gradient on the edge and the subsequent surface, and not a genuine example of filling-in due to edge-based information. Hence, the stimuli were constructed using the method of Dakin and Bex (2003).

The model's predictions for magnitude of the effect as a function of the distance between edges, and hence the amount of filling-in required, was also studied. The scale of the Laplacian of Gaussian filter used to create the stimuli was kept at $\sigma = 0.85$ pixels (or 0.017°), and the width of the disc and the annular region were always equivalent (and increased in octave intervals), in keeping with the methodology of Dakin and Bex (2003). Care was taken when calculating the mean predicted brightness of the disc and annulus, and only brightness values where the original stimulus's edge gradient was not present, were used in the measurement of the model's output. Figure 5.45 illustrates the predictions of the model, which are that the effect should decrease as a function of the distance, and hence the size

³²in this case of Figure 5.43 the larger of the two discs was white.

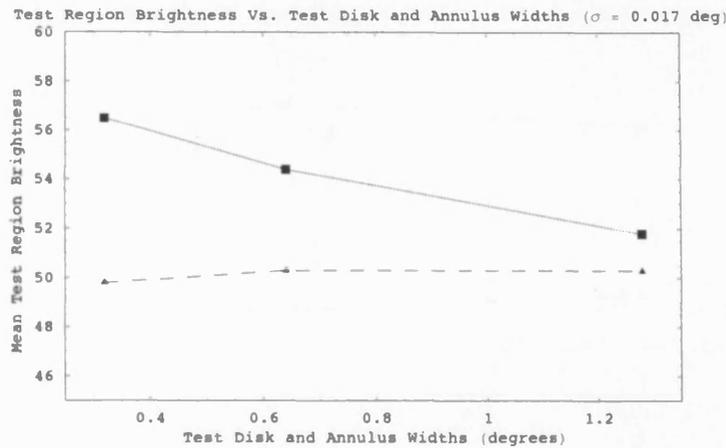


Figure 5.45: A plot of the model's output for the magnitude of the effect as a function of the distance between edges. The solid line represents the mean predicted brightness of the central disc and the dashed line represents the mean predicted brightness of the annulus. The scale of the Laplacian of Gaussian filter used to create the stimuli was kept at $\sigma = 0.85$ pixels (or 0.017°), in keeping with methodology of Dakin and Bex (2003).

of the area that requires filling-in. This behaviour is likely to result from the finite range of the intrinsic blurring in the model's reconstruction of the original image. Additionally, the strength of the effect in the annulus is predicted to be weaker than that in the disc. It should be noted that this was not explicitly observed in the experiments of Dakin and Bex (2003).

The parametric manipulation above was also repeated, but this time allowing the scale of the Laplacian of Gaussian filter used to construct the stimuli, and hence the size of the gradient, to increase proportionally with the distance between edges. Figure 5.46 illustrated the model's output, which predicts that the effect should remain approximately constant with respect to the previous study, as the greater the size of the gradient, then the greater the degree and distance over which the blurring can act. Although different stimuli were used, this effect of gradient size on effect strength was also implicitly observed in the experiments by Moulden and Kingdom (1990).

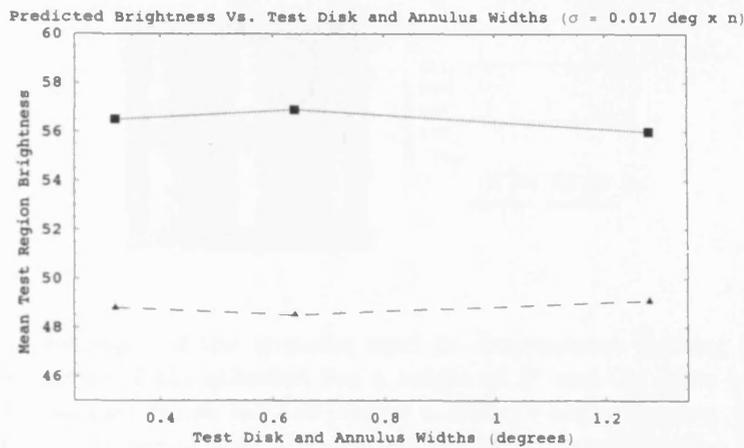


Figure 5.46: A plot of the model's output for the magnitude of the effect as a function of the distance between edges. The solid line represents the mean predicted brightness of the central disc and the dashed line represents the mean predicted brightness of the annulus. The scale of the Laplacian of Gaussian filter used to create the stimuli was doubled with each successive increase in stimulus size.

Grating Induction

Although sometimes regarded as a type of brightness illusion, the phenomena referred to as Grating Induction (GI) is also discussed as an example of filling-in, and sometimes referred to as a 'visual phantom' (Komatsu, 2006). Figure 5.47 demonstrates the type of stimulus which elicits the effect, as used by McCourt and Kingdom (1996). Within the test strip, which possesses the same luminance as the grey background, a cosine grating is in anti-phase with that above and below, is filled-in to the test strip. Figure 5.48 illustrates the model's output for this stimulus, and clearly predicts a significant effect. Unlike the CCOB, the model predicts that this effect is driven by induced contrast on the boundaries between the actual grating and the test strip, and the anti-phase nature of the induced grating is in synchrony with the polarity of the SBC effect projected into the test strip from the boundaries.

The magnitude of the GI effect has been shown to decrease as SF increases (McCourt & Kingdom, 1996; Blakeslee & McCourt, 1999). McCourt and Kingdom (1996) measured the strength of the effect using inducing gratings between $0.0625c/^\circ$ - $4.0c/^\circ$ (a plot of their

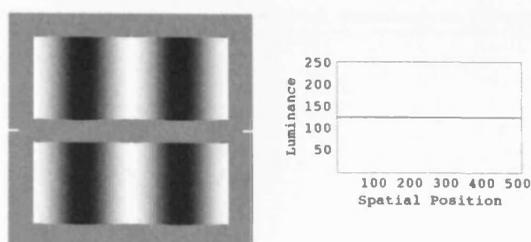


Figure 5.47: An example of the stimulus used to demonstrate Grating Induction. The test strip in the centre of the stimulus has a height of 1° and the same luminance as the background. Above and below are sections of a $0.25c/^\circ$ cosine grating with a minimum values of 0 and a maximum of 255. A cosine grating in anti-phase to that above and below is perceived within the test strip.

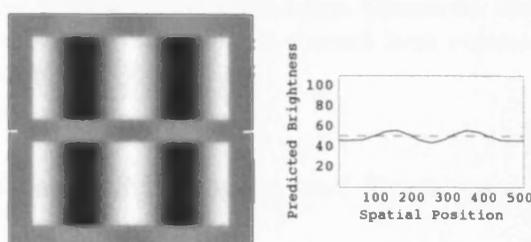


Figure 5.48: The model's output for the Grating Induction stimulus above (left), and its mean predicted brightness cross-section calculated over the range of the entire test patch (right). The presence of the anti-phase cosine grating within the test strip is clearly predicted.

data for observers BM and FK is reproduced in Figure 5.49 together with the model's predictions). For one participant, no effect could be measured above $0.5c/^\circ$, where as the cut-off was closer to $1.0c/^\circ$ for the other participants. This study on GI was modelled using inducing gratings between $0.125c/^\circ$ - $0.5c/^\circ$, and effect strength was defined as the mean predicted brightness of the light regions of the induced grating, less that of the dark regions³³. The models' predictions for this study are illustrated in Figure 5.49, and are in good accordance with the data collected by McCourt and Kingdom (1996), as the effect is predicted to decrease as SF increases, and the cut-off is predicted to be at $0.5c/^\circ$.

³³grating with SF lower than $0.125c/^\circ$ could not be measured as their spatial extent would be too large to be processed by the model.

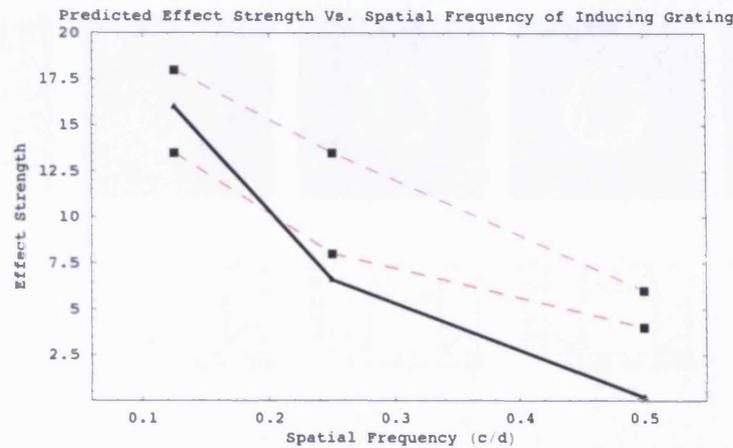


Figure 5.49: A plot of the model's output for the magnitude of the effect as a function of the SF of the inducing grating. The dashed line represents the predicted magnitude of the effect in the stimulus's test strip, and the dashed lines represent the data collected by Moulden & Kindgom (1996).

The Temporal Dynamics of Filling-In and Brightness

A novel aspect of this model with respect to the many models of lightness and brightness coding (as discussed in chapter 2), is that it can make predictions for the temporal dynamics of filling-in. Even though no explicit temporal dimension has been incorporated into the model, each iteration in the gradient descent of the biologically plausible tile alignment process, can be thought of representing the passing of a single 'unit of time'. The exact duration of each time unit depends on how long the underlying neurons need to align the local brightness tiles from the state they're in at iteration i , to the state at iteration $i+1$. Hence in this context, it may not be possible to make accurate quantitative predictions for the time necessary for a given surface to be perceived as being filled-in, but it is possible to make predictions on how the filling-in process will evolve through time.

Of particular relevance is the work on the temporal dynamics of the filling-in process by Paradiso and Nakayama (1991). In their study a white disk target on a black background was presented to participants for 16ms, after which a white circle on a black background was presented as a mask, several milliseconds afterwards. The duration between target and mask was varied, as was the diameter of the target disk, such that the mask may impede

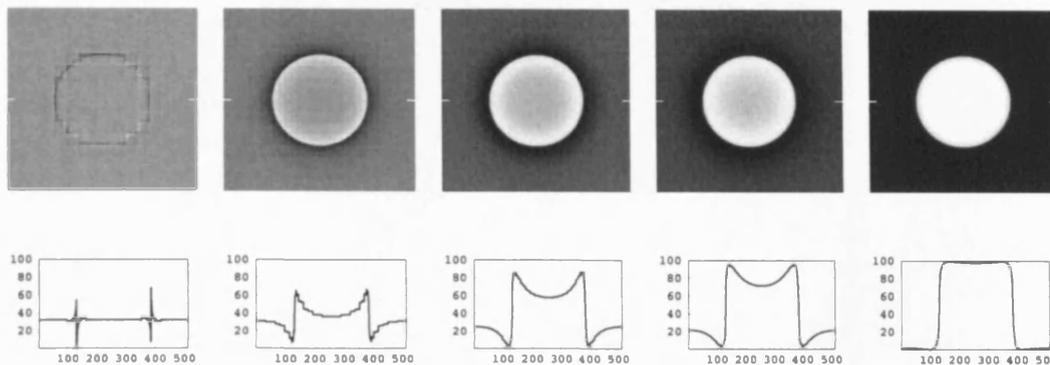


Figure 5.50: The model's output for the reconstruction of a white disk (5° in diameter) on a black background, at 0%, 5%, 40%, 100%, and 300% of the iterations used to reconstruct all other stimuli in this chapter (from left to right). The predicted brightness cross-sections are taken at the midpoint of the disk.

the filling-in process at earlier or later stages of its evolution. The observer's task was to perform an achromatic colour matching task on a concentric circular region at the centre of where the larger target disk was originally presented. The aim of this experiment was to test the concept that edges are responsible for the nature of the filling-in, and hence the filling-in 'signal' ought to travel outwards from them (from the edge of the white disk and into the centre). Thus, by presenting a circular contour (with a smaller diameter than the white disk), shortly after the initial white disk, and before it has been completely filled-in, the filling-in signal should be halted (or at least impeded) by the presence of this new edge, and the central test-region should appear darker, if this mode of filling-in is indeed correct.

It was found that for a fixed target disk diameter, the brightness of the matching region was dependent on the time before mask onset. Longer onsets produced greater brightness, converging on precise matches with the original white target disk. Shorter onsets produced the opposite effects. Furthermore, as the diameter of the target was decreased, the brightness of the matching was also decreased (as if the mask was catching the filling-in and impeding it sooner rather than later). Another variation of this task was performed using a discontinuous mask (a dashed circular contour), with varying gap sizes and a fixed target disk diameter. It was found that the larger the gaps the closer the brightness match was

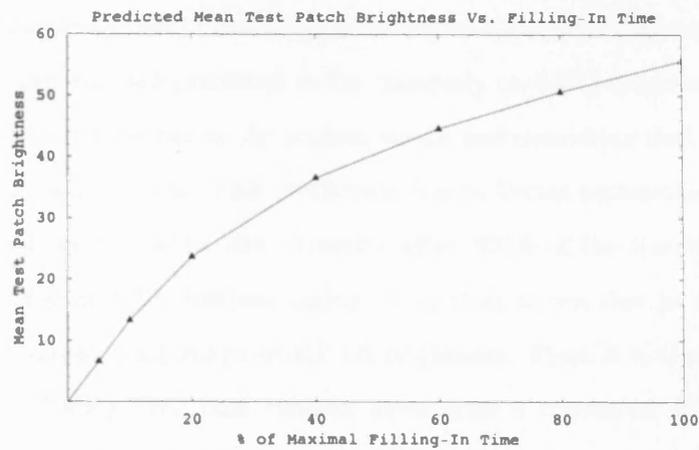


Figure 5.51: A plot of the model's output for the brightness of a 1° concentric disc within that depicted in Figure 5.50, as a function of temporal duration (% of maximum iterations used).

with the target disk luminance, as if the filling-in was seeping around the gaps. All of the above is taken as evidence for a diffusive filling-in process progressing from the edges in the disk.

Due to the lack of an explicit temporal dimension it is not possible to replicate these studies with the model. However, it is still possible to make a qualitative comparison with the model's predictions for the evolution of the filling-in process in the same stimulus, with the conclusions of Paradiso and Nakayama (1991). Figure 5.50 depicts the model's predicted brightness for a 5° white target disk at 0%, 5%, 40%, and 100% of the iterations used to reconstruct all other stimuli in this chapter. Before any iterations, all that is present is the multi-scale reconstruction of the disk's edge. As 'time' passes, the disk's brightness first becomes apparent at the edges, before becoming manifest at the centre. At all stages the edges are always predicted to be brighter than the centre, a direct consequence of induced contrast at edges, and as time progresses this effect becomes visible weaker. Figure 5.51 plots the mean predicted brightness of a concentric 'matching disk' 1° in diameter. The model predicts that the brightness of the matching disk is a non-linear monotonic function of time. A similar non-linear relationship between brightness and time was also found by

Paradiso and Nakayama (1991). However, what Figures 5.50 and 5.51 also demonstrate is how the edges of the disc are predicted to be massively ($\approx 45\%$) brighter than the centre, something that was not reported in the original study, and something that is not observed in our everyday visual experiences. This prediction can be better understood if one considers the model's output for the white disc stimulus after 300% of the iterations used process all other stimuli (Figure 5.51, furthest right). It is clear to see that in this instance that the disc is approximately uniform in predicted brightness. Thus, it is the same mechanism responsible for producing SBC (the residual error after a truncated minimisation of the global LBT error function), that yields this anomalous prediction. The number of iterations chosen to use in the model was primarily based on the number required to accurately reconstruct natural scene images, which do not have such large surfaces and as much error in the resultant reconstruction. What the reconstructions in Figure 5.50 may imply is that for stimuli like the white disc to be attributed the correct brightness, a separate mechanism may be required for setting the number of iterations to use, based on the information content of the scene being viewed.

The Temporal Evolution of SBC and White's Effect

The model's implicit temporal dimension can also be used to make predictions for the temporal evolution of visual phenomena. At this stage we return to SBC and White's effect with the aim of using the model to predict which of the two effects is the first to manifest. The variant of SBC chosen for this study is that illustrated in Figure 5.18, and has two 1° test patches. The variant of White's effect used is illustrated in Figure 5.27, and has an inducing grating of $1.6c/^\circ$. These particular variants of SBC and White's effect were chosen because the magnitude of the effect between their two test patches are almost identical at 19.5% and 16.9%, respectively (within one JND). Thus, when comparing the temporal evolution of the two phenomena, any effect strength derived biases will not be present. The magnitude of the effects, as a percentage of the magnitude at the maximal filling-in time, were computed for 0%, 0.5%, 1%, 2%, 5%, 10%, 20%, 50%, 100% of the number of iterations that constitute the maximal filling-in time. Figure 5.52 illustrates the results of this study,

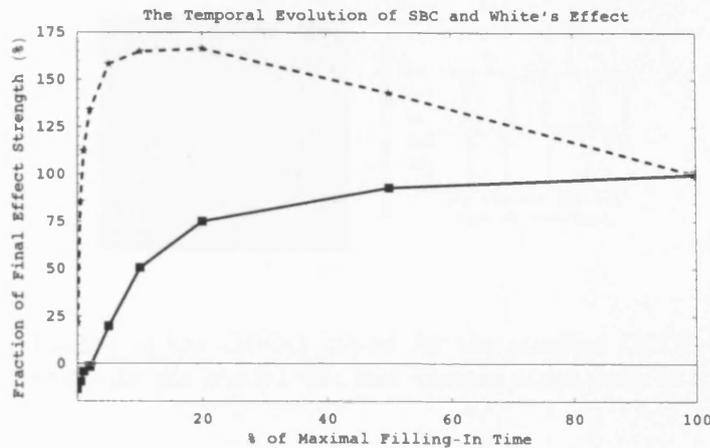


Figure 5.52: A plot of the relative predicted magnitudes of SBC and White's effect as a function of the number of iterations in the model (measured as a percentage of the number of iterations of the maximal filling-in time). The dashed line represents the magnitude of SBC as a percentage of the magnitude at the maximal filling-in time, and the solid line represents the magnitude of White's Effect as a percentage of the magnitude at the maximal filling-in time.

from which two important predictions can be drawn: Firstly, at onset White's effect is not predicted to be present (there is a -3% effect strength that indicates an insignificant SBC effect), while SBC is predicted to be perceived at onset (there is a 5.2% effect that can be considered significant); secondly, the magnitude of the SBC effect rises rapidly, and becomes greater than the maximal effect strength before converging on the maximal value, while White's effect, in comparison, rises steadily to the maximal value. To the knowledge of the author no experimental data exists with which to verify these predictions, and they represent novel predictions of the ILC model.

A Comparison with Other Models

With the exception of the parametric studies, all of the effects modelled above were tested with the ODOG model of Blakeslee and McCourt (1999) and the NSS model of Dakin and Bex (2003) (unless this output has already been published). The aim of this process was to isolate differences in the outputs of the three different models, and hence isolate the computational mechanisms required to generate the phenomena.

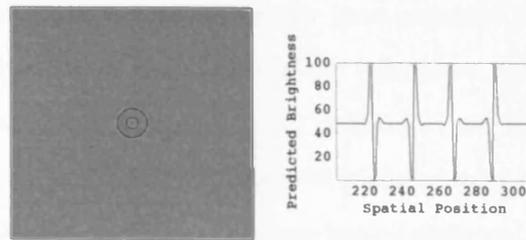


Figure 5.53: The output of the ODOG model for the smallest CCOB stimulus, with a diameter of 0.85° , and with the central disc and annulus width both at 0.32° .

It is possible to report that the NSS model is capable of accounting for all of the effects that the Implicit Luminance Coding (ILC) is capable of accounting for. However, this is not the case for the ODOG model. In addition to the circular variant of White's effect discussed in Chapter 2 (and seen in Figure 5.39), the ODOG model could not account for the circular variant of the CCOB illusion. The ODOG model's output for this stimulus is illustrated in Figure 5.53. The mean predicted brightness of the circular disk and surrounding annulus were calculated at 49% and 52.6% respectively, which results in the opposite polarity to that which has been determined experimentally (Dakin & Bex, 2003; Moulden & Kingdom, 1990).

Summary

The ILC model has been tested on a wide range of visual brightness phenomena that cover *both* contrast and assimilation effects, as well as some lightness phenomena that are thought to arise from scene interpretation. For the former, it has demonstrated that accounts can be provided by a biologically plausible low-level interaction between two assimilation and contrast mechanisms. Complete accounts of both the SBC and White's Effect, together with all pertinent variations are included. Furthermore, accounts of other examples of contrast and assimilation effects (Mach Bands, Vasarely's Pyramid, Chevreul's, and Todorovic's illusions) have been provided, demonstrating that the ILC model does not just predict the correct brightness for a few examples in each class of phenomenon investigated. In

general, SBC effects are predicted to be weaker than reported in the data (by $\approx 20\%$), while assimilation effects are in closer agreement (within $\approx 5\% - 10\%$). For lightness phenomena, it has shown that Adelson's 'Wall of Blocks' and Kofka's rings may require a better understanding of higher-level visual processing in order to be able to provide full accounts (in keeping with the views on visual processing discussed in Chapter 2). However, in the case of the Agostini and Benary Cross effects, it has shown that these phenomena do not *need* to be explained in terms of higher-level processing. The ILC model explicitly violates the Anchoring Theory of Lightness in some scenes, by predicting small variations in brightness ($\approx 3\% - 5\%$) for different regions of white in the scene. The ILC model is also capable of putting forward good accounts for two important filling-in phenomena, the CCOB and grating induction effects, as well as providing an understanding of the temporal dynamics of this process that is in qualitative agreement with experiment. Furthermore, its implicit inclusion of a temporal dimension also predicts that SBC will be present at stimulus onset, while White's effect will not, and that the temporal evolution of the two effects are very distinct, with SBC being greater in magnitude before it settles to the long-term predicted brightness value. These latter predictions are not possible with the other models that have been described in this work. However, in examining the temporal evolution of brightness it was shown that the model's fixed number of iterations (the fixed duration for reconstructing the image), may have to be relaxed in order to correctly predict the brightness of some simple non-natural scene configurations.

Conclusions

This thesis has pursued a single line of enquiry: lightness, brightness, and visual illusions. More specifically, this work has been centred on the view that lightness and brightness phenomena are the direct consequence of low-level visual mechanisms, but only when visual cues that support a specific interpretation of the scene are in absence. When visual cues pertaining to such things as depth perception, transparencies, and changes in illumination are present, it is conjectured that higher-level visual mechanisms modulate the output from low-level visual mechanisms, in an attempt to accurately infer the origins of the luminance and remain lightness constant. Within this context, lightness and brightness illusions could be thought of as either lightness *or* brightness phenomena. This is in contrast to some models of higher-level vision, which claim that *all* such phenomena are the result of higher-level mechanisms that attempt to accurately infer the origins of the luminance and remain lightness constant (Sinha & Adelson, 1993; Purves et al., 1999; Gilchrist et al., 1999).

In particular, this work has focused on White's effect, SBC, and theories that can account for both phenomenon. White's effect has been of particular interest, as it defies the canonical low-level neural explanations based on a simple filtering of the image with receptive fields of the type commonly found in the primary visual stream, whereas such theories are capable of accounting for SBC. However, ideas from higher-level vision have proved highly (although not completely) successful in predicting White's effect *and* SBC (Todorovic, 1997; Gilchrist et al., 1999; Anderson, 2001). This has generated a debate on the precise nature of the calculations that generate lightness and brightness. One of the aims of this work was to provide evidence that phenomena classified as higher-level, such as White's effect, can be accounted for by neural models of low-level vision, which also account for brightness induction effects.

In Chapter 2, current low- and high-level theories of lightness and brightness were reviewed in the context defined above. It was concluded that higher-level theories could not provide accounts for all variants of White's effect, such as the circular variant (Howe, 2005) or the

change in effect strength with the SF of the inducing grating (White, 1981; Blakeslee & McCourt, 1999; Anstis, 2004), and that the NSS model of Dakin and Bex (2003), and the ODOG model of (Blakeslee & McCourt, 1999), were the only low-level theories of brightness capable of accounting for both White's effect and SBC, based on neural processes. As such, in Chapter 3 both models were reconstructed, analysed, and evaluated through a systematic analysis of their constituent parts. It was shown that both models are reliant on the amplification of low SF information, to accurately reconstruct images and account for the illusory brightness apparent in White's effect. Furthermore, it was shown that the ODOG model could not account for the circular variant of White's effect, as the account it provides for the conventional stimulus configuration (White, 1979; Blakeslee & McCourt, 1999), is based on kind of 'neural T-junction analysis', and in the circular variant of the effect there are no junctions of any kind.

It was argued that the large spatial RFs used by the ODOG and NSS models to represent low SF information, are not likely to exist in V1. Based on recent physiological studies which have aimed to measure the spatial extent over which visual information can be integrated in the cortex (Sceniak et al., 2001; Angelucci et al., 2002; Cavanaugh et al., 2002; Stettler et al., 2002), it has been shown that the existence of $0.2c/^\circ$ and $0.1c/^\circ$ tuned Log-Gabor RFs does not fit with even the most accommodating data. Similarly, the global re-weighting and re-normalisation schemes employed by the NSS and ODOG model are not constrained by the local nature of intra-cortical connections. Hence, it was concluded that these models are not biologically plausible.

However, physiological recordings from V1 and upstream have shown that low SF information ($0.2c/^\circ$ - $0.2c/^\circ$) does exist (Rossi & Paradiso, 1999; Ringach, 2001; Xing et al., 2004), and cells have been found with responses to either the actual or perceived luminance of surfaces (Rossi & Paradiso, 1999; Kinoshita & Komatsu, 2001; Roe et al., 2005). A framework for interpreting these results was proposed: cells which appear to consistently respond to absolute surface luminance, are those which leak local D.C (or local mean luminance) information in their response - they do not play a functional role in surface representation

as their response to a surface would be hard to disambiguate from that to a sub-optimal grating. ‘Blob’-like cells (see Figure 3.23) (Ringach, 2001) *are* involved in surface representation, but are not part of an explicit feed-forward luminance code. If arrays of such cells did exist, and the retinal image is explicitly represented in the cortex, then there would be little reason for the existence of contrast coding cells upstream from V1 - a view supported by information theoretic analyses of the primary visual stream (Barlow, 1961; Kelly, 1962; Barlow, 1981; Nirenberg et al., 2001; Z. Li, 2006).

Following on from the conclusions of Chapter 3, a new model of luminance coding was proposed in Chapter 4. Firstly, the issue of recovering low SF and local mean luminance information without explicitly sampling it was considered. The problem was defined in the framework of Scale-Space, and based on the output from a set of GD spatial filters. Recovering information about local mean-luminance (and hence low SFs), is equivalent to predicting the output of zero-order GD filters (equivalent to the ‘blob’-like RFs found by Ringach (2001) and illustrated in Figure 3.23), from the information implicit in the contrast coding higher-order GD filters. That is, that cells with ‘blob’-like RFs code *implied luminance*. The functionality of Scale-Space was leveraged to solve the problem: it was proposing that V1 simple cells provide a measure of the first- and second-order derivatives of a blurred version of the retinal image (that is, that Scale-Space is a model of V1 simple cells), and that this information is sampled at discrete intervals by overlapping receptive fields. An approximation of the Taylor series (or local jet) representation of the image, operating with the available differential information, is used to ‘fill-in’ local brightness values in a square region about each sampling point. Reconstruction of the zero order term (the global luminance structure), is performed by enforcing continuity between neighbouring regions of local brightness. This is achieved by shifting the absolute brightness level of each local region, and optimising the shifts such that the variance in the differences in brightness within overlapping regions is minimized. This technique can effectively reconstruct natural scene images.

However, the optimisation process is not biologically plausible; neurons are not capable

of solving systems of linear equations analytically, especially when it involves integrating information from the entire extent of visual space, as discussed at the end of Chapter 3. An alternative method by which the optimisation can proceed was then proposed. It is based on the gradient-descent numerical approximation to the problem, and is shown to utilise local interactions between neighbouring local brightness regions to recover the global brightness structure in a biologically-plausible manner, consistent with the known functionality of neurons. This process introduces error into the reconstruction of the local brightness structure, which has the overall effect of inducing contrast at borders in an image, as observed in the simultaneous brightness contrast effect, introducing a new way of looking at a well known phenomenon.

The final model consists of five such reconstructions, at five different scales, combined in an un-weighted linear sum to produce the final predicted brightness image. The use of multiple scales was motivated by for several reasons: information at multiple scales is available to the cortex (Georgeson & Sullivan, 1975; Hubel, 1988); information at differing scales is of equal importance (Koenderink, 1984); and, because it was shown that the reconstruction of the global brightness structure was relatively invariant to noise for larger scales. The implications of the latter are that a robust reconstruction of the image with good acuity, for both small and large scales can be achieved by integrating information from multiple scales. The scales were chosen such that the RFs used in each reconstruction do not exceed the sizes commonly found within V1 (De Valois et al., 1982; Sceniak et al., 2001). The use of larger scales also introduces blurring over greater distances into the reconstruction, which has the affect of providing an explanation for Whites effect.

In Chapter 5, a wide range of brightness, lightness, and filling-in stimuli were tested on the biological ILC model, whose output was compared to experimental data. The model proves to be very successful at providing accounts for these phenomenon, based on the interplay between the contrast and assimilation mechanisms described above. In particular the model was shown to be capable of providing comprehensive accounts of all known variants of White's effect and SBC (although in general the magnitude of the former were found to

be greater than that found psychophysically, and the magnitude of the latter were found to be smaller). As a consequence of this, White's effect is predicted to be the result of an assimilation mechanism acting independently of the contrast mechanism responsible for the accounts of SBC. Furthermore, the model provides support for the view that *both* SBC and White's effect are derived purely from low-level visual mechanisms, and that higher-level visual process do not have a significant part to play in accounting for such phenomena, as has been suggested (Sinha & Adelson, 1993; Purves et al., 1999; Gilchrist et al., 1999). Similarly, the model's accounts for Agostini's SBC illusion (Agostini & Galmonte, 2002) and the Benary Cross (Benary, 1924) imply that these effects are also not the result of higher-level visual processing.

However, the model was not capable of providing accounts for the 'Wall of Blocks' (Adelson, 1993) and Kofkas' rings (Kofka, 1935) demonstrations. Both of these demonstrations could involve changes in cues for spatial configuration, which could alter the interpretation of the scene, the inferred origin of the luminance, and hence the attributed lightness. Thus, in its failure the model has supported the view that higher-level visual mechanisms modulate the output from low-level visual mechanisms only when genuine visual cues are present in the scene, where a genuine visual cue could now be defined in relation to what the model predicts.

One novel aspect of the model with respect to other low-level models of brightness, is that it is possible to make qualitative predictions for the temporal evolution of brightness. In particular, it was shown that the temporal evolution of brightness filling-in is a non-linear function of time, as determined empirically by (Paradiso & Nakayama, 1991). Furthermore, it was possible to make unique predictions for the temporal evolution of White's effect relative to SBC. It is predicted that SBC will be present at stimulus onset, while White's effect will not, and that the temporal evolution of the two effects are very distinct, with SBC being greater in magnitude before it settles to the long-term predicted brightness value.

That the model is in general agreement with empirical data, suggests that the explicit sam-

pling of local mean luminance information is not essential for determining the brightness in images. Given that information theoretic arguments exist to support the view that this information is not explicitly transmitted to the cortex, then it has been shown how the visual system can circumvent this limitation. Furthermore, the model's competent predictions also serve as a justification for the choice of spatial scales used to determine brightness.

The model has, however, proved to be insufficient in some circumstances. It was shown that the model failed to accurately reconstruct a solitary white disc, 5° in extent, on a black background. However, when the model was allowed to iterate 200% more times than the that deemed necessary to reconstruct natural scenes, an accurate reconstruction could be obtained. This implies that a separate mechanism to control the number of iterations in the model, which converges to those chosen in the case of natural scenes, may be required, and points to a stream of investigation for future work.

Additionally, future research efforts could be made to investigate whether the model could be used to understand how other filling-in mechanisms operate. For example, when filling-in regions that are not present in the original image, either by design of a stimulus, or due to lack of input at the retina either in the blind-spot at the optic nerve, or in a scotoma (real or artificial), as discussed by Komatsu (2006). Information in the derivative structure surrounding a region of the image for which there is no response, could be used to infer the derivative structure within this region, possibly by extending the size of the LBTs of the surrounding regions, before proceeding with the iterative tile aligning algorithm.

References

- Adelson. E. H. (1993). Perceptual organization and the judgement of brightness. *Science*, 262, 2042-2044.
- Adelson. E. H. (2000). Lightness perception and lightness illusions. In M. Gazzaniga (Ed.), *The new cognitive neurosciences* (2 ed., p. 339-351). Cambridge, MA: MIT Press.
- Adelson. E. H., & Pentland. A. P. (1996). The perception of shading and reflectance. In D. Knill & W. Richards (Eds.), *Perception as bayesian inference* (p. 209-223). New York: Cambridge University Press.
- Agostini. T., & Galmonte, A. (2002). A new effect of luminance gradient on achromatic simultaneous contrast. *Psychonomic Bulletin and Review*, 9(264-269).
- Allman, J., Miezin. F., & McGuinness, E. (1985). Stimulus specific responses from beyond the classical receptive field: Neurophysiological mechanisms for local-global comparisons in visual neurons. *Annual Review of Neuroscience*, 8, 407-430.
- Anderson, B. L. (2001). Contrasting theories of white's illusion. *Perception*, 30, 1499-1507.
- Angelucci. A., Levitt, J. B., Walton, E. J. S., Hupe, J., Bullier, J., & Lund, J. S. (2002). Circuits for local and global signal interaction in primary visual cortex. *Journal of Neuroscience*, 22, 8633-8646.
- Anstis. S. M. (2004). White's effect in color, luminance, and motion. In L. Harris & M. Jenkin (Eds.), *Proceedings of the york university conference in honour of d m regan*.
- Barlow. H. B. (1961). Possible principles underlying the transformations of sensory messages. In W. Rosenblith (Ed.), *Sensory communication* (p. 217-234). Cambridge, MA: MIT Press.
- Barlow. H. B. (1981). The ferrier lecture. 1980: Critical limiting factors in the design of the eye and visual cortex. *Proceedings of the Royal Society of London*, 212, 1-34.
- Barrow, H. G., & Tenenbaum, J. M. (1978). Recovering intrinsic scene characteristics from images. In A. Hanson & E. Riseman (Eds.), *Computer vision systems* (p. 3-26). New York: Academic Press.
- Benary. W. (1924). Beobachtungen zu einem experiment uber helligkeitskontrast. *Psychol-*

ogische Faschung, 5, 131-142.

- Bisti, A., Clement, R., Maffei, L., & Mecacci, L. (1977). Spatial frequency and orientation tuning curves of visual neurones in the cat: Effects of mean luminance. *Experimental Brain Research*, 27, 335-345.
- Blakeslee, B., & 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., & McCourt, M. E. (2001). A multiscale spatial filtering account of the wertheimer-benary effect and the corrugated mondrian. *Vision Research*, 41, 2487-2502.
- Blakeslee, B., & McCourt, M. E. (2004). A unified theory of brightness contrast and assimilation incorporating oriented multiscale spatial filtering and contrast normalization. *Vision Research*, 44, 2483-2503.
- Brainard, D. H., & Wandell, B. A. (1986). Analysis of the retinex theory of colour vision. *Journal of the Optical Society of America A*, 3, 1651-1661.
- Brewer, A. A., Press, A. W., Logothetis, N. K., & Wandell, B. A. (2002). Visual areas in macaque cortex measured using functional magnetic resonance imaging. *Journal of Neuroscience*, 22, 10416-10426.
- Brown, S. P., & Masland, R. H. (2001). Spatial scale and cellular substrate of contrast adaptation by retinal ganglion cells. *Nature Neuroscience*, 4, 44-51.
- Buchsbaum, G., & Gottschalk, S. (1983). Trichromacy, opponent colours coding, and optimum colour information transmission in the retina. *Proceedings of the Royal Society of London*, 220, 89-113.
- Burr, D. C., & Morrone, M. C. (1994). The role of features in structuring visual images. In G. R. Bock & J. A. Goode (Eds.), *Higher order processing in the visual system* (p. 129-146). Cambridge, UK: Cambridge University Press.
- Carandini, M., Heeger, D. J., & Movshon, J. A. (1997). Linearity and normalisation in simple cells of the macaque visual cortex. *Journal of Neuroscience*, 25, 8521-8544.
- Cavanaugh, J. R., Wyeth, C., & Movshon, A. M. (2002). Nature and interaction of signals from the receptive field centre and surround in macaque v1 neurons. *Journal of*

Neurophysiology, 88, 2530-2546.

- Chevreul, M. E. (1890). *The principle of harmony and contrast of colours* (C. t. Martel, Ed.). London: Bell.
- Cornsweet, T. (1970). *Visual perception*. New York: Academic Press.
- Cornsweet, T., & Pinsky, H. (1965). Luminance discrimination of brief flashes under various conditions of adaptation. *Journal of Physiology*, 176, 294-310.
- Craik, K. (1966). *The nature of psychology* (S. Sherwood, Ed.). Cambridge, UK: Cambridge University Press.
- Dakin, S. C., & Bex, P. J. (2003). Natural image statistics mediate brightness 'filling in'. *Proceedings of the Royal Society of London*, 270, 2341-2348.
- Dayan, P., & Abbott, L. F. (2001). *Theoretical neuroscience*. Cambridge, MA: MIT Press.
- DeAngelis, G. C., Freeman, R. D., & Ohzawa, I. (1994). Length and width tuning of neurons in the cat's primary visual cortex. *Journal of Neurophysiology*, 71(347-374).
- De Valois, R. I., Albrecht, D. L., & Thorell, L. G. (1982). Spatial frequency selectivity of cells in macaque visual cortex. *Vision Research*, 22, 545-559.
- 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*, 4, 2379-2394.
- Florack, L., Haar Romeny, B. M. ter, Viergever, M., & Koenderink, J. (1996). The gaussian scale-space paradigm and the multiscale local jet. *International Journal of Computer Vision*, 18, 61-75.
- Foster, K. H., Gaska, J. P., Nagler, M., & Pollen, D. A. (1985). Spatial and temporal frequency selectivity of neurones in visual cortical areas v1 and v2 of the macaque monkey. *Journal of Physiology*, 365, 331-363.
- Freeman, W. T., & Adelson, E. H. (1991). The design and use of steerable filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13, 891-906.
- Gallant, J. L., Connor, C. R., Rakshit, S., Lewis, J. W., & Essen, D. C. van. (1996). Neural responses to polar, hyperbolic, and cartesian gratings in area v4 of macaque monkey. *Journal of Neurophysiology*, 26, 2718-2793.
- Georgeson, M. A., & Sullivan, G. D. (1975). Contrast constancy: Deblurring in human vision by spatial frequency channels. *Journal of Physiology (London)*, 252, 627-656.

-
- Gilchrist, A., Kossyfidis, C., Agostini, T., Li, X., Bonato, F., Cataliotti, J., et al. (1999). An anchoring theory of lightness perception. *Psychological Review*, *106*, 795-834.
- Goldstein, E. B. (2003). *Sensation and perception* (6 ed.). Independence, Kentucky, USA: Thomson Wadsworth.
- Grossberg, S. (2003). Filling in the forms: Surface and boundary interactions in visual cortex. In L. Pessoa & P. DeWeerd (Eds.), *Filling-in: From perceptual completion to skill learning* (p. 13-37). New York: Oxford University Press.
- Grossberg, S., & Hong, S. (2006). A neural model of surface perception: Lightness, anchoring, and filling-in. *Spatial Vision*, *9*(263-321).
- Grossberg, S., & Todorovic, D. (1988). Neural dynamics of 1d and 2d brightness perception: A unified model of classical and recent phenomenon. *Perception and Psychophysics*, *41*, 241-277.
- Haar Romeny, B. M. ter. (2003). *Front-end vision and multi-scale image analysis*. Dordrecht, the Netherlands: Kluwer Academic Publishers.
- Hecht, S. (1920). Human retinal adaptation. *Proceedings of the National Academy of Sciences USA*, *6*, 112-115.
- Helmholtz, H. V. (1887). *Handbuch der physiologischen optik*. Leipzig: Voss.
- Hering, E. (1964). *Outline of a theory of the light sense (translated from the 1920's original in german)* (L. Hurvich & D. Jameson, Eds.). Cambridge, MA: Harvard University Press.
- Hong, S., & Grossberg, S. (2004). A neuromorphic model for achromatic and chromatic surface representation of natural images. *Neural Networks*, *17*, 787-808.
- Horn, B. K. P. (1974). Determining lightness from an image. *Computer Graphics and Image Processing*, *3*, 277-299.
- Howe, P. D. L. (2001). A comment on the anderson (1997), the tolorovic (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 illusion. *Perception*, *34*, 557-564.
- Huang, X., MacEnvoy, S., & Paradiso, M. A. (2002). Perception of brightness and brightness illusions in the macaque monkey. *Journal of Neuroscience*, *22*, 9618-9625.
-

-
- Hubel, D. H. (1988). *Eye, brain, and vision*. New York: Scientific American Library.
- Hubel, D. H., & Wiesel, T. N. (1977). Ferrier lecture: Functional architecture of macaque monkey visual cortex. *Proceedings of the Royal Society of London*, 198(1-59).
- Kelly, D. H. (1962). Information capacity of a single retinal channel. *IEEE Transactions on Information Theory*, 8, 221-226.
- Kingdom, F., McCourt, M. E., & Blakeslee, B. (1996). In defence of "lateral inhibition" as the underlying cause of induced brightness phenomenon: A reply to spehar, gilchrist, and arend. *Vision Research*, 37, 1039-1044.
- Kingdom, F., & Moulden, B. (1991). White's effect and assimilation. *Vision Research*, 31, 151-159.
- Kingdom, F., & Moulden, B. (1992). A multi-channel approach to brightness coding. *Vision Research*, 32, 1565-1582.
- Kinoshita, M., & Komatsu, H. (2001). Neural representation of the luminance and brightness of a uniform surface in the macaque primary visual cortex. *Journal of Neuroscience*, 42, 1495-1517.
- Knill, D., & Kersten, D. (1991). Apparent surface curvature affects lightness perception. *Nature*, 351, 228-230.
- Koenderink, J. (1984). The structure of images. *Biological Cybernetics*, 50, 363-370.
- Kofka, K. (1935). *Principles of gestalt psychology*. New York: Harcourt.
- Komatsu, H. (2006). The neural mechanisms of perceptual filling-in. *Nature Reviews Neuroscience*, 7, 220-231.
- Kries, J. von. (1902). Chromatic adaptation. In D. L. MacAdam (Ed.), *Sources of colour science (1970)* (p. 109-119). Cambridge, MA: MIT Press.
- Kries, J. von. (1905). Influence of adaptation on the effects produced by luminous stimuli. In D. L. MacAdam (Ed.), *Sources of colour science (1970)* (p. 120-126). Cambridge, MA: MIT Press.
- Land, E. H. (1983). Recent advances in retinex theory and some implications for cortical computations: Colour vision and the natural image. *Proceedings of the National Academy of Sciences USA*, 80, 5163-5169.
- Land, E. H., & McCann, J. J. (1971). Lightness and retinex theory. *Journal of the Optical*
-

Society of America, 61, 1-11.

- Levitt, J. B., & Lund, J. S. (1997). Contrast dependence of contextual effects in primate visual cortex. *Nature*, 387, 73-76.
- Li, X., & Gilchrist, A. (1999). Relative area and relative luminance combine to anchor surface lightness values. *Perception and Psychophysics*, 61, 771-785.
- Li, Z. (2006). Theoretical understanding of early visual processes by data compression and data selection. *Network: Computation and Neural Systems*, 4, 301-334.
- Lindeberg, T. (1996). Scale-space: A framework for handling image structures at multiple scales. In E. van Zee (Ed.), *Proceedings of the cern school of computing* (p. 1-12). The Netherlands.
- MacKay, D. J. C. (2003). *Information theory, inference, and learning algorithms*. Cambridge, UK: Cambridge University Press.
- Maloney, L. T. (1999). Physics based approaches to modelling surface colour perception. In K. R. Gegenfurtner & L. T. Sharpe (Eds.), *Colour vision: From genes to perception* (p. 387-422). Cambridge, UK: Cambridge University Press.
- Marr, D. (1982). *Vision*. San Francisco: W. H. Freeman.
- Masland, R. H. (2001). The fundamental plan of the retina. *Nature Neuroscience*, 4, 877-886.
- McCourt, M. E., & Kingdom, F. (1996). Facilitation of luminance grating detection by induced gratings. *Vision Research*, 36, 2563-2573.
- Monnier, P., & Shevell, S. K. (2003). Large shifts in color appearance from patterned chromatic backgrounds. *Nature Neuroscience*, 6, 801-802.
- Morrone, M. C., & Burr, D. C. (1988). Feature detection in human vision. *Proceedings of the Royal Society of London*, 235, 221-245.
- Morrone, M. C., Burr, D. C., & Ross, J. (1994). Illusory brightness step in the chevrel illusion. *Vision Research*, 34(1567-1574).
- Moulden, B., & Kingdom, F. (1990). Light-dark asymmetries in the Craik-Cornsweet-O'Brien illusion and a new model of brightness coding. *Spatial Vision*, 5, 101-121.
- Nirenberg, S., Carcieri, S. M., Jacobs, A. L., & Latham, P. E. (2001). Retinal ganglion cells act largely as independent encoders. *Nature*, 411, 698-701.

-
- O'Brien, V. (1959). Contrast by contour enhancement. *American Journal of Psychology*, *72*(299-300).
- Olzak, L. A., & Wickens, T. D. (1997). Discrimination of complex patterns: Orientation information is integrated across spatial scale; spatial frequency and contrast information are not. *Perception*, *26*, 1101-1120.
- Paradiso, M. A., & Nakayama, K. (1991). Brightness perception and filling-in. *Vision Research*, *31*, 1221-1236.
- Pollen, D. A., & Ronner, S. F. (1981). Phase relationships between adjacent simple cells in the visual cortex. *Science*, *212*(1409-1411).
- Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. (2002). *Numerical recipes in c++* (2 ed.). Cambridge, UK: Cambridge University Press.
- Purves, S., Shimpf, A., & Beau-Lotto, R. (1999). An empirical explanation of the cornsweet effect. *Journal of Neuroscience*, *19*, 8542-8551.
- Ratliff, F. (1965). *Mach bands*. San Francisco: Holden-Day.
- Ringach, D. L. (2001). Spatial structure and symmetry of simple-cell receptive fields in macaque primary visual cortex. *Journal of Neurophysiology*, *88*, 455-463.
- Rodieck, R. W. (1965). Quantitative analysis of cat retinal ganglion cell response to visual stimuli. *Vision Research*, *5*(583-601).
- Rodieck, R. W. (1998). *The first steps in seeing*. Sunderland, MA: Sinauer.
- Roe, A. W., Lu, H. D., & Hung, C. P. (2005). Cortical processing of a brightness illusion. *Proceedings of the National Academy of Sciences USA*, *102*, 3869-3874.
- Ross, J., Morrone, M. C., & Burr, D. C. (1989). The conditions under which mach bands are visible. *Vision Research*, *29*, 699-715.
- Rossi, A. F., & Paradiso, M. A. (1999). Neural correlates of brightness in the retina, lgn, and striate cortex. *Journal of Neuroscience*, *19*, 6145-6156.
- Rudd, M. E., & Arrington, K. F. (2001). Darkness filling-in: A neural model of darkness induction. *Vision Research*, *41*, 3649-3642.
- Rudd, M. E., & Zeemach, I. K. (2005). The highest luminance anchoring rule in achromatic color perception: Some counterexamples and an alternative theory. *Journal of Vision*, *5*, 983-1003.
-

-
- Sceniak, M. P., Hawken, M. J., & Shapley, R. (2001). Visual spatial characterization of macaque v1 neurons. *Journal of Neurophysiology*, *85*, 1873-1887.
- Shapley, R., & Enroth-Cugell, C. (1984). Visual adaptation and retinal gain controls. *Progress in Retinal Research*, *3*, 263-364.
- Shapley, R., & Reid, R. C. (1985). Contrast and assimilation in the perception of brightness. *Proceedings of the National Academy of Sciences USA*, *82*, 5983-5986.
- Silverman, S. M., Grosf, D. H., De Valois, R. I., & Elfar, S. D. (1989). Spatial frequency organization in primate striate cortex. *Proceedings of the National Academy of Sciences USA*, *86*, 711-715.
- Sinha, P., & Adelson, E. H. (1993). Recovering reflectance and illumination in a world of painted polyhedra. In *Proceedings of the fourth international conference on computer vision* (p. 156-163). California, USA: IEEE Computer Society Press.
- Spehar, B., Clifford, W. G., & Agostini, T. (2002). Induction in variants of white's effect: Common or separate mechanisms. *Perception*, *31*, 189-196.
- Spehar, B., Gilchrist, A., & Arend, L. (1995). The critical role of relative luminance relations in white's effect and grating induction. *Vision Research*, *35*, 2603-2614.
- Stettler, D. D., Das, A., Bennett, J., & Gilbert, C. D. (2002). Lateral connectivity and contextual interactions in macaque primary visual cortex. *Neuron*, *36*, 739-750.
- Stockman, A., & Sharpe, L. T. (2000). Spectral sensitivities of the middle- and long-wavelength sensitive cones derived from measurements in observers of known phenotype. *Vision Research*, *40*, 1711-1737.
- Tadmor, Y., & Tolhurst, D. J. (1994). Discrimination of changes in the second order statistics of natural and synthetic images. *Vision Research*, *17*, 977-981.
- Todorovic, D. (1997). Lightness and junctions. *Perception*, *26*, 379-395.
- Vasarely, V. (1970). *Vasarely ii*. Neuchatel: Editions du Griffon.
- Weber, E. H. (1978). *The sense of touch (translated from the german original of 1834)* (H. E. Ross & M. D. J, Eds.). New York: Academic Press.
- White, M. (1979). A new effect of pattern on perceived lightness. *Perception*, *8*, 413-416.
- White, M. (1981). The effect of the nature of the surround on the perceived lightness of grey bars within square-wave test gratings. *Perception*, *10*, 215-230.
-

- Williams, S. M., McCoy, A. N., & Purves, D. (1998). An empirical explanation of brightness. *Proceedings of the National Academy of Sciences USA*, *95*, 13301-13306.
- Xing, D., Ringach, D. L., Shapley, R., & Hawken, M. J. (2004). Correlation of local and global orientation and spatial frequency tuning in macaque v1. *Journal of Physiology*, *557*, 923-933.
- Young, R. A., & Lesperance, R. M. (2001). The gaussian derivative model for spatial-temporal vision: 2. cortical data. *Spatial Vision*, *14*, 321-381.
- Young, R. A., Lesperance, R. M., & Weson Meyer, W. (2001). The gaussian derivative model for spatial-temporal vision: 1. cortical model. *Spatial Vision*, *14*, 261-319.
- Yund, E. W., & Armington, J. C. (1975). Color and brightness contrast effects as a function of spatial variables. *Vision Research*, *15*, 917-929.
- Zeki, S. (1983). The distribution of wavelength and orientation selective cells in different areas of monkey visual cortex. *Proceedings of the Royal Society of London*, *217*, 449-470.
- Zhaoping, L. (2002). Optimal sensory encoding. In A. A. Arbib (Ed.), *The handbook of brain theory and neural networks* (2 ed., p. 815-819). Cambridge, MA: MIT Press.