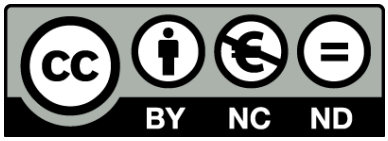
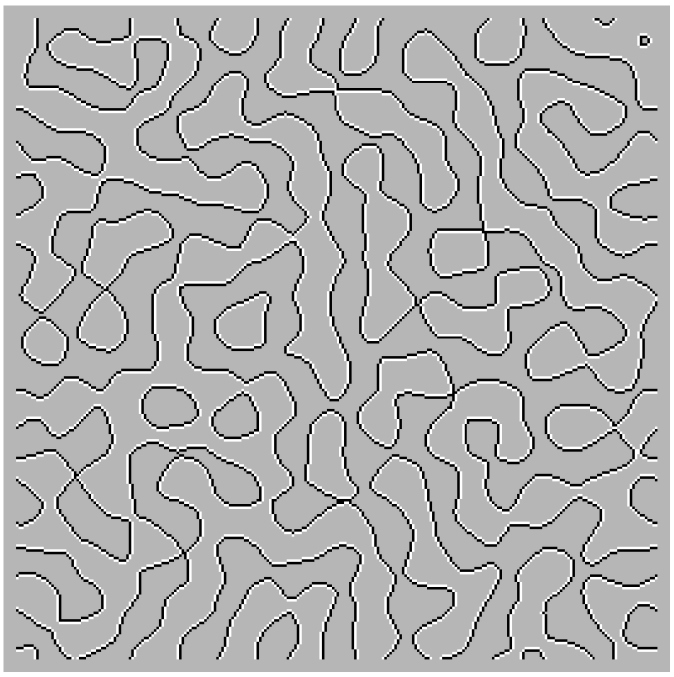
UNIVERSITAT DE BARCELONA

Computational principles of early vision for explaining a broad range of brightness and lightness phenomena

Alejandro Lerer Gornatti



Computational principles of early vision for explaining a broad range of brightness and lightness phenomena



Alejandro Lerer Gornatti

Doctoral Dissertation

Programa de Doctorat en Biomedicina, Departament de Cognició,
Desenvolupament i Psicologia de l'Educació, Facultat de Psicologia,
Universitat de Barcelona

UNIVERSITAT DE BARCELONA

Computational principles of early vision for explaining a broad range of brightness and lightness phenomena

---

Autor

Alejandro Lerer Gornatti

Directors

Dr. Hendrik Anne Super

Dr. Matthias Sven Keil

Tesi Doctoral del Programa de Doctorat en Biomedicina

Departament de Cognició, Desenvolupament i Psicologia de l'Educació, Facultat de Psicologia,

Universitat de Barcelona

Setembre 2021

Alejandro Lerer | Hans Supèr | Matthias S. Keil

Esta tesis está dedicada a mis padres y hermanos, a los amigos más cercanos, a mis compañeros de piso que durante largo tiempo me han soportado hablando de este proyecto, a todos aquellos que me han animado a seguir peleando, a Charo por sostenerme y darme luz en los peores momentos, y en especial a mis compañeros "Cangrejos al Sol" que su compañía siempre ha servido de inspiración y fuerza para estar lleno de vida y alegría.

Resumen

Aunque la percepción visual se ha estudiado extensamente durante siglos, los mecanismos neuronales subyacentes de la percepción visual siguen siendo confusos. El presente estudio tiene como objetivo desarrollar un modelo computacional basado en el procesamiento de bajo nivel del sistema visual (es decir, consistente con propiedades conocidas de áreas visuales tempranas), para desvelar mecanismos y estrategias de codificación computacional subyacentes a las percepciones básicas de brillo y luminosidad para estímulos visuales acromáticos. El modelo implementa dos mecanismos unificadores hipotéticos: uno basado en la variabilidad en la respuesta que segrega el gradiente de luminancia frente a las características sin gradiente, y el otro en la redundancia-reducción de las representaciones de los bordes. Mediante estos mecanismos, el modelo predice con éxito los fenómenos de brillo y ligereza de gran cantidad de ilusiones visuales. Muchas de estas predicciones coinciden con algunos resultados de experimentos psicofísicos, lo que indica que los mecanismos hipotéticos presentados podrían ser principios computacionales subyacentes prominentes para las percepciones de brillo y luminosidad en el sistema visual.

Abstract

Although visual perception has been extensively studied for centuries, the underlying neural mechanisms remain puzzling. The present study aims to develop a computational model based on the low-level processing of the visual system (i.e., consistent with known properties of early visual areas), to unveil mechanisms and computational encoding strategies underlying basic perceptions of brightness and lightness for achromatic visual inputs. The model implements two hypothetical unifying mechanisms: one based on variability in response that segregates gradient vs. non-gradient features, and the other based on dynamic filtering that reduces redundancy of edge representations. By these mechanisms, the model successfully predicts the brightness and lightness phenomena of many visual illusions. Many of these predictions concur with some results of psychophysical experiments, indicating that the variability-based segregation mechanism and dynamic filtering could be prominent underlying computational principles for brightness and lightness perceptions in the visual system.

Keywords: *brigthness, lightness, computational model, visual illusions*

Contents

Chapter 1

Global Introduction

The brain analyses the visual world through the luminance patterns that reach the retina. When luminance reaches the retina, it stimulates the retinal photoreceptors, which transmit information about the visual world to our brain (in the visual system) as visual input. Formally, luminance (as measured by the retina) is the product of illumination and reflectance. The illumination is defined as the amount of source light incident to the scene, while the reflectance is defined as the amount (in percentage) of light object surface reflected in the visual scene. Because reflectance is multiplicatively coupled to illumination when illumination is homogeneous in a scene (very rare), the reflectance directly correlates with luminance. However, when the illumination is inhomogeneous (very common), infinite combinations of reflectance and illumination can produce ambiguous visual input that reaches the retina. Because of this ambiguity, reflectance and illumination are intrinsically mixed in visual input (luminance), and then, the brain is faced with solving an inverse problem (Pizlo, 2001) to distinguish them. This problem has many solutions, but the brain successfully achieves only one or a few of them, depending on certain constraints, tasks, and context. As a result of this process, the physical properties of visual input are phenomenologically distinguished (i.e., perceived) by the visual system.

For instance, look at Adelson's checker-shadow illusion (adapted version) in Figure 1.1 (left). At first sight, it is a checkerboard under a shading. Easily, by distinguishing a bright vs. shaded region in the visual scene, we can perceive the appearance of illumination. In addition, by distinguishing different gray levels of luminance (from "white to black") that compose the whole visual scene (i.e., visual input), we can directly

perceive the appearance of luminance. Beyond the appearance of luminance, we also can distinguish the appearance of surface reflectance across checker by "white vs. black" squares. Notice that the difference between the appearance of luminance is distinct from the appearance of reflectance: In Adelson's checker-shadow illusion, the sensation of "white vs. black" remains despite the squares lying in different regions but, indeed, the "white" squares under the shaded region have the same luminance as the "black" squares under the bright region (compare with Figure 1.1, right).

Formally, the process of distinguishing the appearance of luminance is usually defined as brightness perception (Arend et al. 1993), while the process of distinguishing the appearance of surface reflectance is defined as lightness perception (Arend et al. 1993). Thus, formally, brightness perception comprises all scene perceptual aspects (appearance of luminance), while lightness perception is implicitly included in brightness. How the brain achieves and distinguishes brightness and lightness is still an ongoing scientific debate.

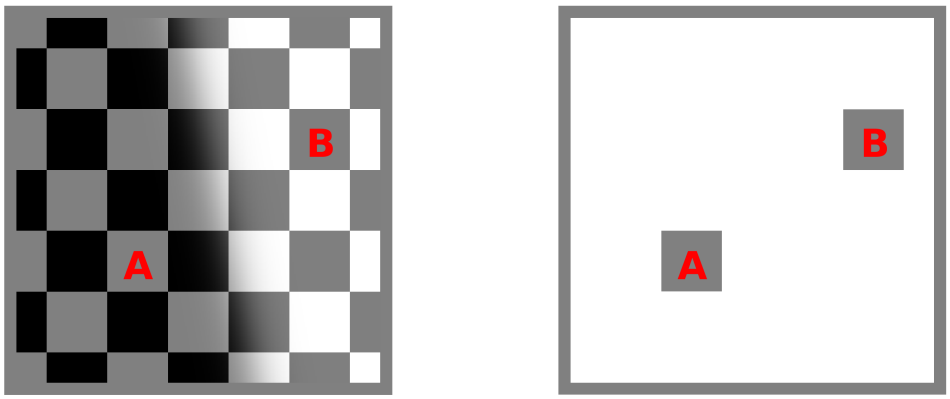


Figure 1.1: Left. Adapted version of Adelson's checker-shadow, squares with red A and red B have the same luminance but are perceived as lighter and darker, respectively (Adelson 1995). **Right.** The A and B squares without the context of the visual scene.

Historically, visual illusions, brightness, and lightness phenomena are natural ways to understand brightness and lightness perceptions. Understanding how the visual system makes accurate perceptions are often best clarified by understanding why it sometimes does not. A typical phenomenon concerning lightness is the "lightness constancy": it

refers to the appearance of an object's lightness that generally remains constant despite changes of external conditions such as illumination and distance (e.g., Adelson's checker-shadow illusion). Unlike lightness constancy, slight modifications of adjacent luminance patterns or spatial configurations in certain conditions could lead to different brightness/lightness phenomena such as contrast, assimilation, or induction effects. In the contrast phenomenon, the brightness/lightness of the test fields are contrasted with the background; the Simultaneous Contrast illustrates this phenomenon (see Figure 1.2A). On the contrary, the assimilation phenomenon means that the brightness/lightness[1] of the test fields averages with the background; Dungeon illustrates this phenomenon (see Figure 1.2B). Further, other brightness (or lightness) phenomena also emerge depending on certain configurations of visual features like gradients, edges, shapes (see Figure 1.2C, 1.2D, 1.2E).

These kinds of brightness/lightness phenomena and visual illusions are not only of great utility for building hypotheses about computational mechanisms or perceptual rules that underlie visual perception but also highlight the incompleteness or inconsistency of the proposed mechanisms. For instance, to explain Hering Grid (see Figure 1.2E), the classical explanation deems the lateral inhibition produced center-surround receptive fields of retinal ganglion cells as the principally acting mechanism (Baumgartner, 1960). However, the mechanism is insufficient to explain why the effect is considerably reduced (or even removed) if the bars are slightly corrugated (see Figure 1.2F). Similarly, Simultaneous Contrast can be explained "also" by lateral inhibition between a target region (center) and its context (surround). However, the mechanism is insufficient to explain why the appearance of self-luminosity (e.g., halo) enhances "contrast," although the local contrast luminance of visual input remains unchanged (e.g., Figure 1.2D). Unlike contrast phenomenon, assimilation (e.g., Figure 1.2B) pulls a target's brightness towards that of its immediate context, and therefore cannot be explained by mechanisms based on direct lateral inhibition. Overall, the neural mechanisms involved in generating brightness and lightness, respectively, appear to be more intricate (e.g., Grossberg, 2017; Schmid & Anderson 2017; Anderson et al. 2014; Kingdom 2011; Anderson & Winawer, 2005).

[1] As a clarification, here, for homogeneous illumination, we refer to brightness/lightness (together) due to the judgments of apparent reflectance (lightness) for homogeneous illumination are either identical to judgments of apparent intensity (brightness) (Blakelsee et al., 2008; Keil 2008).

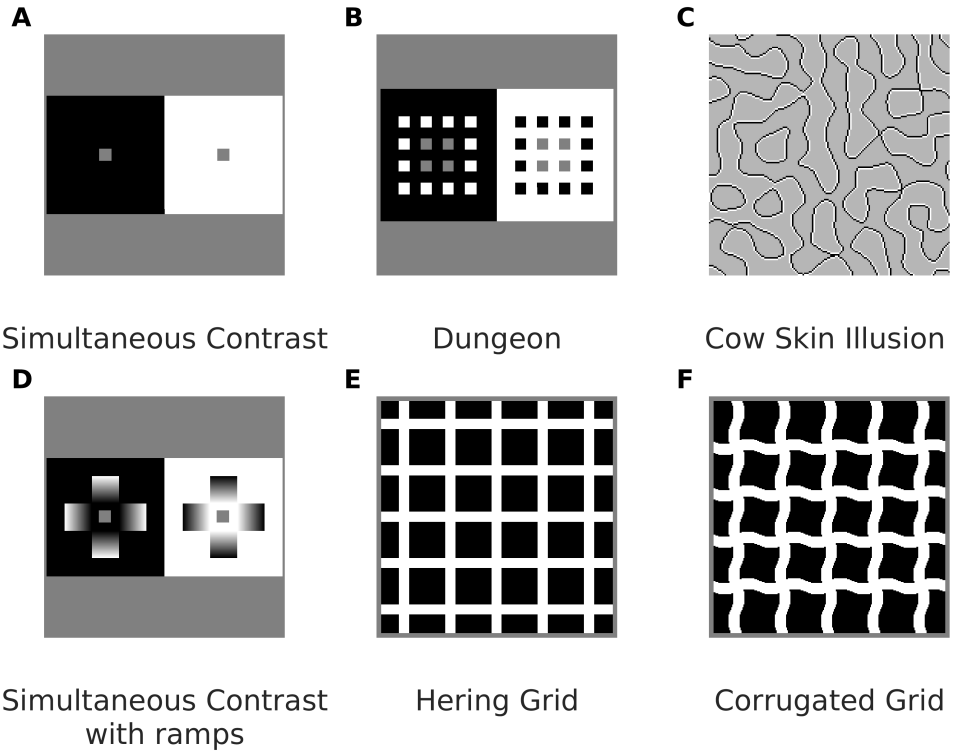


Figure 1.2: A. Simultaneous brightness contrast: The two gray patches with identical luminance increase their brightness/lightness difference with their respective backgrounds. B. Dungeon: The two gray areas with identical luminance reduce their brightness/lightness difference with their respective backgrounds. C. Cow Skin Illusion: a variation of Craik-O'Brien-Cornsweet illusion. All regions have identical luminance; however, such regions are differently perceived because the opposite luminance gradients ("cusps "almost unperceived) at the edges. D. Simultaneous brightness contrast with ramps: although the white and black regions have the same level of luminance, respectively, the ramps contribute to a sensation of self-luminosity (halo vs. glow) between regions. Notice a stronger sensation of contrast effect than the classic simultaneous contrast in A. E, F. Hermann/Hering (HG) illusion and a corrugated version. At the intersections of the white grid lines, illusory gray spots are perceived in the HG but less intense (or even removed) in the corrugated grid.

It is particularly challenging for computational proposals to simultaneously explain contrast assimilation effects or distinguish brightness or lightness with identical model parameters or what "neural coding principles" are involved. For instance, filling-in (FI) models (Cohen MA & Grossberg S, 1987; Gerrits HJM & Vendrik AJH. 1970) can be considered a neural implementation of edge integration, where boundary activity is propagated laterally to generate perceptible surface properties. Whereas the prediction of brightness contrast or lightness surface usually is straightforward with (most) filling-in architectures, assimilation effects remain challenging. Typically, a solution using FI models consists of distinguishing two types of contours between local and global (or "remote") boundary maps (Grossberg S & Todorovic D. 1988; Rudd Michael E. 2017). If local vs. global boundaries are chosen "carefully," FI models can achieve compatibility between contrast and assimilation. However, their implementations are based on artificial descriptors (but see Grossberg S. & Todorovic D. 1988; Domijan 2015) or limited to synthetic images (i.e., incompatible with real-world images).

Other computational models (the majority) for computing brightness estimates decompose an input image typically with filters of different orientations and multiple spatial frequencies, which mimic some neural response properties in the early visual system. Blakeslee and McCourt (Blakeslee B, McCourt ME. 1999; 2004) proposed a highly successful image decomposition with oriented difference of Gaussians (ODOG) filters. In the ODOG-model, the main mechanism is a “normalization response” consisting of two steps. First, a weighted sum of filter responses across spatial frequencies is computed for each orientation (orientation channels). Second, each orientation channel is divided by its root mean square level before they are summed to yield the final brightness map. The ODOG model successfully predicts contrast brightness and many assimilation displays. However, the correctness of these predictions is limited to visual input with homogeneous illumination (i.e., when brightness = lightness). The predictions fail to deal with the effects of non-homogeneous illumination on lightness perception (e.g., see Figure 1.2D) or perceived illumination as the glare effect (e.g., see Figure 1.3). Further, these models also fail to predict visual illusions where the figure-ground relations modify brightness or lightness perception; as an example, see Benary-Cross and Reverse Contrast in Figure 1.3.

Unlike spatial filter models, layers decomposition approaches are (computational) models which aim at deriving further images based on the characteristics of the depicted

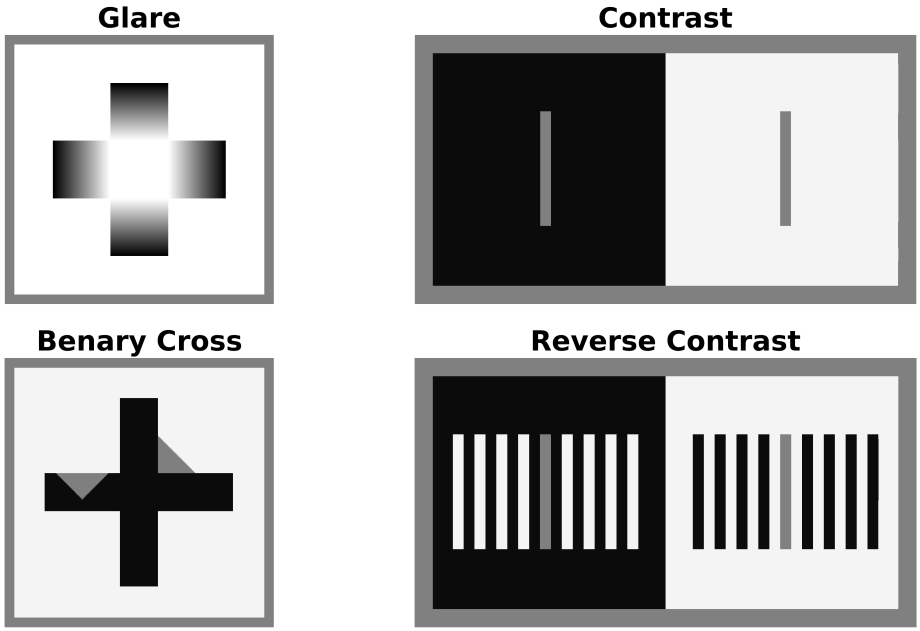


Figure 1.3: Glare. An example of the glare effect, the white region has the same luminance, but it appears self-luminous modifying the perceived brightness. **Benary Cross**, both grey triangles have the same luminance and are flanked by an equal amount of black and white; however, the triangle belonging to the cross is perceived as lighter than the other. **Contrast.** An example of Simultaneous Contrast (SC) similar to Figure 1.2A: The two gray patches with identical luminance *increase* their brightness/lightness difference with their respective backgrounds. **Reverse Contrast.** The simultaneous contrast is reversed (into "assimilation") by adding more structures to the original SC display: The two gray patches with identical luminance *reduce* their brightness/lightness difference with their respective backgrounds.

visual scene of an input image. Typical approaches with respect to layer decomposition are based on intrinsic images, which, in a wider sense, address surface qualities such as transparency, lightness, and occlusions (references). These models work very well to describe lightness for non-homogeneous illumination and successfully describe certain lightness phenomena as lightness constancy. However, typically they offer no robust mechanistic explanation of how layer decomposition (or intrinsic images) is combined with luminance values to compute lightness (or brightness) or how figure-ground relations (as reverse contrast) modify the perception of brightness or lightness.

Other proposal models (indeed theories) suggest higher-level processing to explain how

figure-ground relations modify lightness perception based on perceptual grouping. The best-known is the anchoring theory (Gilchrist et al., 1999), which suggests the stimulus is perceptually "anchored" into local and global perceptual frameworks in the visual scene: Each framework is assumed to be illuminated equally. The theory proposes that the highest luminance in each framework is "anchored" to correspond to a lightness of white, while the lightness of other regions (with lower luminance) are rescaled (or codetermined) relative to the other "anchors" within its local framework in relation to the entire global framework. Although this simple approach (or rules) describes some lightness phenomena even for non-homogeneous illumination, it fails to describe brightness phenomena (e.g., glare effect). Further, the "rule" is limited only to visual illusions (or images) with low complexity, being the "rule" un-computable to specifying frameworks in real-world images.

A completely different approach for explaining visual illusions is based on a statistical analysis of real-world images (Purves D et al., 2004; Yang & Purves, 2004). This approach suggests that the perception of brightness or lightness is related to knowledge about the statistical relationships between visual patterns across space. The perception of the target then depends on its expected luminance given its current context: It is perceived as brighter if the expected luminance is lower input luminance, and it is perceived as darker otherwise. This approach is successful in predicting contrast and assimilation for several visual illusions and suggests a statistical relationship between luminance patterns and brightness perception. Unfortunately, no attempts have been made to unveil any underlying mechanisms involved in the computation of brightness and lightness from the statistical analysis (but see Morgenstern Y et al. 2014).

In summary, the neuronal underpinnings of brightness and lightness perception remain a mystery. Many existing models for proposing neural mechanisms only work with synthetic input images but often struggle with real-world images. Yet different models are only phenomenological descriptors based on psychophysical studies and, as such, cannot make predictions about neural mechanisms. Other models present some inconsistencies in robustness and fail in their predictions if illusions are slightly modified. On the other hand, neurophysiological studies attempting to link neuronal activity to lightness/brightness perception have proven difficult. On top of that, in psychophysical experiments, lightness is often confused with brightness, which complicates the comparison of corresponding results. It is due in large to the most extended consensus that

defines brightness as the "appearance" of luminance and lightness as the "appearance" of surface reflectance (Arend et al. 1993). However, since "appearance" is subjective, the interpretation or judgments of brightness or lightness may account for numerous "errors," especially in achromatic images (see Blakeslee & McCourt 2015; with replies).

In this study, we propose hypothetical unifying mechanisms for estimating brightness perception and lightness perception for achromatic images. On the one hand, for visual scenes containing inhomogeneous illumination (i.e., ambiguity in visual patterns), we hypothesize that the visual system could distinguish the apparent illumination and apparent surface reflectance (i.e., lightness) by a segregation mechanism. On the other hand, independently of illumination configuration (with homogenous or inhomogeneous illumination), we hypothesize that the different kinds of brightness/lightness phenomena emerge due to a mechanism that reduces context redundancy in edge representations.

Goals

The present study aims to develop a computational model based on the low-level processing of the visual system, i.e., consistent with known properties of early visual areas. We propose hypothetical unifying mechanisms for estimating brightness perception and lightness perception for achromatic images. In particular, we hypothesize that the visual system could distinguish the apparent illumination and apparent surface reflectance (i.e., lightness) by a segregation mechanism. In contrast, that different kinds of brightness/lightness phenomena emerge due to a mechanism that reduces context redundancy in edge representations. The main goals of the present study and proposed model are:

1. Propose computational encoding strategies according to the hypotheses to compute the perception of brightness, lightness, and perceived illumination features.

2. The model must be well-defined: Robust in successfully processing real-world images and consistent in predicting a challenging set of brightness and lightness phenomena.

3. The model must explain unifying principles of the different kinds of brightness and lightness phenomena as possible and reproduce some of the trends of corresponding psychophysical data.

4. The proposed mechanisms must be published to be reviewed and contrasted by the scientific community.

Chapter 2

First Publication: "Luminance Gradients and non-Gradients as a cue for distinguishing reflectance and illumination in achromatic images: A computational approach".

Chapter abstract: This chapter presents a published study describing a computational model compatible with the early visual system consisting of a segregation mechanism to distinguish the apparent illumination and surface reflectance of an achromatic visual input. Since the publication corresponding to this chapter is not open access, there is a post-print copy instead. Here, the corresponding "published version" DOI https://doi.org/10.1016/j.neunet.2018.11.001

Luminance Gradients and non-Gradients as a cue for distinguishing reflectance and illumination in achromatic images: A computational approach

Alejandro Lerer[a], Hans Supèr[a,b,c,d], Matthias S. Keil[a,b]

[a]*Departament de Cognició, Desenvolupament i Psicologia de l´Educació, Faculty of Psychology, University of Barcelona, Barcelona, Spain*
[b]*Institut de Neurociéncies, Universitat de Barcelona, Barcelona, Spain*
[c]*Institut de Recerca Pediàtrica Hospital Sant Joan de Déu, Barcelona, Spain*
[d]*Catalan Institute for Advanced Studies (ICREA), Barcelona, Spain*

Abstract

The brain analyses the visual world through the luminance patterns that reach the retina. Formally, luminance (as measured by the retina) is the product of illumination and reflectance. Whereas illumination is highly variable, reflectance is a physical property that characterizes each object surface. Due to memory constraints, it seems plausible that the visual system suppresses illumination patterns before object recognition takes place. Since many combinations of reflectance and illumination can give rise to identical luminance values, finding the correct reflectance value of a surface is an ill-posed problem, and it is still an open question how it is solved by the brain. Here we propose a computational approach that first learns filter kernels ("receptive fields") for slow and fast variations in luminance, respectively, from achromatic real-world images. Distinguishing between luminance gradients (slow variations) and non-gradients (fast variations) could serve to constrain the mentioned ill-posed problem. The second stage of our approach successfully segregates luminance gradients and non-gradients from real-world images. Our approach furthermore predicts that visual illusions that contain luminance gradients (such as Adelson's checker-shadow display or grating induction) may occur as a consequence of this segregation process.

Keywords: reflectance, illumination, computational model, lightness, image processing, brightness illusions

October 26, 2018

1. Introduction

Based on two slightly different retinal images, the brain assigns a rich set of perceptual attributes ("features", such as color, depth, etc.) to object surfaces. Each visual object is composed of surfaces, and each surface can be characterized by its reflectance, which is a physical property. Here we define reflectance as specifying the ratio between impinging and reflected light intensity, which theoretically can range from zero (black) to one (white) in case of achromatic images. Reflected intensity (luminance) is eventually measured by the retinal photoreceptors, and represents the signal that the brain uses to interpret the visual world.

Specifically, lightness cannot be extracted directly from luminance, because reflectance is multiplied with illumination: if an illumination source gets more intense, so does the luminance that reaches the eye from a surface. This implies that for determining (absolute) reflectance, one ideally needs knowledge about the illumination. Since many combinations of reflectance and illumination can produce the same value of luminance that reaches the retina, the brain has to approximate the solution to an inverse problem. A good estimation of reflectance and/or illumination, respectively, can be expected to make object recognition performance largely independent of illumination conditions. Notice that one and the same object can be found in a myriad of illumination conditions, and it is unlikely that the brain memorizes each corresponding condition, since the number of illumination conditions is infinite. In what follows, we consider the spatial aspects of vision, neglecting temporal variations. Furthermore, we will consider only achromatic reflectance (the brain interprets chromatic reflectance as color, and achromatic reflectance as lightness).

Rather than relying directly on absolute luminance, the visual system starts with comparing (spatially) adjacent luminance values (contrasts), because if luminance is sampled from any adjacent surfaces with different reflectance values, then these samples normally co-vary with illumination. But at which distance should we compare luminance values? A possible assumption is that reflectance changes are associated with abrupt jumps in luminance within relatively short distances, while illumination varies more smoothly, across larger distances [1, 2, 3, 4, 5]. This in turn relates to spatial frequencies and scale, respectively: High spatial frequencies are generated by luminance variations across relatively short distances, whereas lower spatial frequencies involve larger distances.

The idea of decomposing a (luminance) image into oriented contrasts at different spatial scales (i.e. layers with spatial frequencies) is exploited by most computational approaches to "brightness perception" (brightness is perceived luminance). Albeit these models are often successful in explaining many visual illusions (e.g., [6, 7]), they implicitly hold that simple and complex cells in the primary visual cortex create a kind of "all-in-one" representation of the retinal image(s). Otherwise expressed, they do not propose any concrete mechanism for determining reflectance, or suppressing illumination, respectively. This is a major difference to so-called "filling-in" models [8, 9, 10]. In filling-in models, contrasts are extracted at first, possibly across different scales. These contrast maps can be processed further such that they delineate the contours of surfaces (surface boundaries). Subsequently, contrast measurements are propagated from the boundaries to the interior of surfaces, such that ideally each representation of a surface is "tagged" with a single activity value.

Similar to the Retinex mechanism [2], activity propagation (filling-in) acts to suppress any smooth variations in luminance across a surface (discounting gradient luminance, [11]. In our paper, we build on the idea that reflectance changes are often associated with fast luminance variations (edges), and that shading is often associated with slow variations [12, 13, 14]. Especially in achromatic vision, this cue (fast versus slow luminance variations) is exploited by the visual system in order to estimate reflectance given luminance. Of course, it may fail in certain situations (e.g., shadows with sharp boundaries), and further achromatic cues (e.g. textural continuity, [15]) as well as chromatic cues (e.g. chromatic versus achromatic variations, [16, 17]) are evaluated concurrently in everyday vision [18, 19]. Thus, the final estimation of reflectance and shading relies on the interaction between several cues, and this fact should be taken into account when interpreting our results. In what follows, we denote the shading-cue with gradients, and the reflectance-cue with non-gradients [12, 13, 14], and here we propose a computational model that generates corresponding representations. Again, we like to emphasize that the non-gradient representations as computed by our model may be different from the true estimation of reflectance (i.e. lightness) because our model relies only on one cue for identifying putative reflectance changes in luminance patterns. Furthermore, our model does not classify the detected gradients according to their source (e.g. shading, curvature, shadows).

At first, our model learns two classes of filter kernels from natural (or real-world) gray-level images. One class of filters is constrained to high spatial

frequencies. This class of filters is therefore sensitive to non-gradient features such as surface boundaries and edges. The other class is sensitive to gradient features and thus captures smooth luminance variations such as shading. The second part of our model segregates gradients from non-gradients in gray-level images, and builds corresponding representations. As a consequence of this segregation process, we can predict visual illusions that contain luminance gradients. Since our approach successfully processes real-world images, and at the same time accounts for gradient-based visual illusions, we believe that it represents a plausible model for information processing in the first stages of the visual system, apart from demonstrating its suitability for estimating achromatic reflectance of object surfaces in gray-level images.

2. Methods

Our computational model consists of two stages. In the first stage, two sets of filter kernels were learned from patches that were randomly chosen from real-world images. The correlation matrix of all patches was computed, and their principal components were determined. With the latter, we extracted the low and high frequency content from the patches. Using sparse coding techniques, two sets of filter kernels were learned from the (frequency-specific) patches, one from the low spatial frequency content, and another one from the high spatial frequency content. The low-spatial frequency filters respond better to smooth intensity variations and thus prefer luminance gradients. The high spatial frequency filters are more sensitive to non-gradient features (lines and edges). In the second stage, each set of learned filter kernels was used to compute corresponding representations (or intrinsic image): one for gradients and another one for non-gradients. The sum of both intrinsic images yields what we call a "full reconstruction". The full reconstruction serves to verify that no information has been lost in the segregation process. Notice, however, that some information is lost before the segregation takes place as a result of image encoding by the learned filter kernels.

2.1. Learning of Filter Sets

2.1.1. Training patches

Figure 1.A illustrates the following description. We extracted square patches (samples) with size 13×13 pixels from gray-level images in the same way as described in [20]. Subsequently, we analyzed the sample set with singular value decomposition (SVD), and produced 169 (13×13) principal

components. With natural images, the statistics of luminance values depend only on their relative distance, but not on the absolute position from where they were extracted [21]. In that case, the principal components are the sinusoidal basis functions of the Fourier transform, with decreasing amplitudes from low to high spatial frequencies [22]. In this way, we were able to modify the spatial frequency content of our sample set by selecting a subset of the principal components (spectral filtering). We used a log-normal kernel for spectral filtering:

$$f_\alpha(\epsilon) = \frac{\alpha}{\epsilon\sqrt{2\pi\sigma^2}} e^{\frac{(\log(\frac{\epsilon}{\alpha})-\mu)^2}{2\sigma^2}} \quad (1)$$

where α is a dilation parameter, $\mu = 0$, and $\sigma = 1.2$ and $\epsilon \in \{1, 2, ..., 169\}$, α indicates the principal component index. Figure 1.B shows an example of this function for different alpha values. When applied to the sample set, this procedure produced two training sets for low ($\alpha = 8$) and high ($\alpha = 64$) spatial frequencies, respectively. Figure 1.C shows examples from the training set.

Although different techniques for filtering exist to generate the training patches (e.g., whitening [20]), we used our procedure for two reasons: (i) The filtering process is directly done on the sample set, which allows a more precise control of the frequency content of the two generated training sets. (ii) We observed that it seems to guarantee a better convergence of our algorithm for filter learning.

2.1.2. Filtering learning

We used SPARSENET, an unsupervised learning algorithm [20], in order to learn the two sets of filter kernels from the corresponding training data. As a result, two sets of filter kernels were learned: one selective for low spatial frequencies (LSF), and another one selective for high spatial frequencies (HSF). Figure 1.D shows some of the learned filters. Each filter set was further pruned by computing the spatial correlation between the learned kernels with suitably parameterized 2D Gabor functions, where those with a correlation smaller than 0.8 were discarded. There are two main reasons for this procedure: (i) We wanted our learned filters to be as similar as possible to simple cells in the primary visual cortex, whose receptive fields can be modeled by Gabor functions [23]. (ii) SPARSENET returns a set of filter kernels that encodes the training samples [20]. However, some of the filter kernels do not account for much encoding variability (low variability or LV-kernels) with respect to other kernels in the set. In other words, when convolved with

natural images, LV-kernels have low amplitudes on the average, as compared to the other learned filters. The LV-kernels therefore can be omitted without significantly compromising the overall coding performance. Finally, we balanced the ON and OFF regions of our filters such that they did not respond to homogeneous regions of luminance.

2.1.3. Preferred luminance gradient direction for a filter

The preferred luminance gradient direction (GraDir) of a learned filter was evaluated by calculating the maximum correlation between the filter and an ideal luminance gradient. The latter was defined as a linearly increasing luminance ramp S_θ along some direction θ (the directional angle θ is measured in degrees):

$$G(\theta) = \sum_{p,q} S_\theta(p,q) w(p,q) \tag{2}$$

$$\hat{\theta} = \arg\max_{\theta} G(\theta) \tag{3}$$

where w represents the filter, the parameters $p,\ q$ denote pixel positions, and $\hat{\theta}$ is the preferred GraDir for the filter w.

2.2. Image Decomposition

Figure 2 illustrates how our model segregates and generates representations of gradients and non-gradients, respectively. In the first stage (encoding stage), an input image is encoded by the learned filters, resulting in two sets of response maps (one response map for each learned filter kernel). In the second stage (contour detection), we used a variability measure in order to generate two respective sets of maps from the corresponding response maps; one for high variation (high variability map), and another one for low variation (low variability map). Finally, in the reconstruction stage, we computed three intrinsic images, the non-gradient layer, the gradient layer, and the full reconstruction, which is the sum of the previous two layers. As mentioned earlier, the full reconstruction layer serves to illustrate that no information is lost due to segregation (although some information is lost before segregation due to image encoding by learned filters).

2.2.1. Encoding an input image

Given one set of learned filter kernels $S \in \{LSF, HSF\}$, an input image was convolved (symbol "$*$") with the learned filter kernels w_i of S:

$$a_i = w_i * Im \tag{4}$$

resulting in a corresponding response map a_i. Note that the filters kernels within each S could not be expected to form an orthogonal set. As a consequence, the response maps $a \in \{a_1, ..., a_n\}$ may contain redundant information. For redundancy reduction we introduced a local mechanism which aims at decorrelating the a_i:

$$b_i(x, y) = (K_S + Id)^{\dagger} a_i(x, y) \tag{5}$$

where K_S is the correlation matrix over the filter set S , Id is the identity matrix, $[.]^{\dagger}$ denotes the Moore-Penrose pseudo-inverse, and $b \in \{b_1, ..., b_n\}$ represents the decorrelated response set for each respective S. The identity matrix has a positive sign because we used an implicit formulation of the decorrelation problem.

2.2.2. Contour map, high and low variability maps

Figure 2B and Figure 3 illustrate the calculation of the contour map. First we computed a variability map m, where we assigned at each image pixel a value as follows:

$$L_i = \nabla^2 b_i$$

$$m(x, y) = \mathrm{std}(L_i(x + l, y + k)) \tag{6}$$

The activities L_i correspond to the highpass-filtered response maps b_i, and are computed by applying the Laplacian ∇^2 to a specific set S of b_i. The variability map m corresponds to the standard deviation (std) of the L_i, $i \in \{1, ..., n\}$, which are centered at pixel (x, y) and include the neighboring pixels ($l, k \in \{-1, 0, 1\}$) using mirror (or symmetric) boundary conditions (Figure 3.A). Next, we used a sigmoidal function $\sigma_{\alpha,\tau}$ for computing a contour map C from the variability map $m(x, y)$:

$$C(x, y) = \begin{cases} \sigma_{\alpha,\tau}(m(x, y)) & \text{if} \quad m(x, y) >= \tau \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

The slope of the sigmoid ("rate parameter") is set to $\alpha \equiv \frac{1}{mean(m(x,y))}$. The threshold τ (which is also the inflection point of the sigmoid) is set so that

the sum of the superthreshold $m(x, y)$ is 0.01 times the total sum of $m(x, y)$ across all spatial positions (x, y):

$$\tau = \arg'_{\tau}\{ \sum_{\substack{x,y \\ m(x,y)>\tau'}} m(x, y) = 0.01 \sum_{x,y} m(x, y)\}$$

meaning that 99 percent of the "total activity" is superthreshold. Figure 3.B shows the contour map of the Lena image. Finally, the high and low variability maps are computed by multiplicative gating of the response maps with the contour map $C(x, y)$ (Figure 3.C). Specifically, the high variability maps are defined by $b_i C$, and the low variability maps by $b_i(1 - C)$, where the matrices b_i and C are multiplied entrywise (Hadamard product).

2.2.3. Cost function for finding Gradient and Non-gradient representations

We formulated the reconstruction of an image $\hat{z}$ as an optimization problem. Because more than one filter is involved in the reconstruction of the input image, we modified Tikhonov regularization accordingly. Our cost function for gradients ($h(b_i, C) \equiv b_i(1 - C)$) and non-gradients ($h(b_i, C) \equiv b_i C$) is defined as:

$$\hat{z} = argmin_z \sum_i \parallel h(b_i, C) - v(z, w_i) \parallel_2^2 + \lambda \parallel \nabla^2 z \parallel_2^2 \tag{8}$$

where $\parallel . \parallel$ denotes the Frobenius norm, $h(b_i, C)$ are the contour-gated response maps, b_i are the decorrelated response maps (Equation 5), C is the contour map (Equation 7), λ is a regularization parameter, and ∇^2 is the Laplacian operator. The function $v(z, w_i)$ describes the encoding of the reconstructed image z by the learned filter kernels w_i (Equation 5):

$$v(z, w_i) \equiv (K_S + Id)^{\dagger} w_i * z \tag{9}$$

where $[.]^{\dagger}$ denotes the Moore-Penrose pseudo-inverse, and $*$ denotes convolution. The correlation matrix corresponding to filter set S is denoted by K_S.

In order to minimize Equation 8, we used the conjugate gradient method, which was terminated either when having reached a maximum number of iterations n, or if the following error criterion was satisfied:

$$\frac{\parallel z_{NG}^{n+1} + z_G^{n+1} - (z_{NG}^n + z_G^n) \parallel}{\parallel z_{NG}^{n+1} + z_G^{n+1} \parallel} < \delta \tag{10}$$

The last equation measures the difference in reconstruction between subsequent iterations ("relative error"), where z_{NG}^{n+1} and z_{G}^{n+1} denote the non-gradient layer and gradient layer, respectively, at iteration $n+1$. Notice that we have two free parameters in this process: One is the regularization parameter λ and the other one is the threshold δ. For all simulations, we used $\lambda = 0.4$ and $\delta = 10^{-5}$.

2.2.4. Generating representations for gradients and non-gradients

The non-gradient layer and the gradient layer were estimated by iterating Equation 8 for the low and high variability maps, respectively (i.e., with the correspondingly defined representation maps $h(b_i, C)$ (see response maps with high and low variability and Figure 3.C). The full reconstruction was computed by simply adding the non-gradient layer to the gradient layer. We included the full reconstruction for the purpose of verification. It illustrates that no features of the original images are lost as a consequence of segregating it into gradients and non-gradients. If not otherwise stated, we used the HSF filter set for computing the non-gradient layer, and the LSF set for computing the gradient layer. Figure 4 shows an illustration of segregating an input image into gradients and non-gradients, respectively: whereas the gradient layer captures only slow (or smooth) variations in luminance, these variations are suppressed in the non-gradient layer. As a consequence of the latter, the non-gradient contains only the abrupt luminance changes, that often correspond to changes in reflectance. Figure 4C furthermore demonstrates that the gradients and non-gradients cannot be obtained by simple lowpass and highpass filtering, respectively, of the input image.

3. Results

This section consists of four parts: (1) An analysis of the responses of the learned filter sets to luminance gradient direction (GraDir), and the role of the preferred GraDir in recovering the non-gradient and gradient layers for synthetic stimuli; (2) By means of an example we show that the creation of (non-)gradient representations is implemented by activity propagation (filling-in); (3) A demonstration of the robustness of our proposed mechanisms by processing real-world images; and (4) model predictions of visual illusions which contain luminance gradients (simultaneous brightness contrast, Adelson's checkershadow illusion, Logvinenko's illusion, and grating induction).

3.1. Filter responses to luminance gratings and luminance steps

We measured the preferred luminance gradient direction (GraDir) of the learned filters with Equation 2 (see methods section). Figure 5.A, shows the (computed) orientation preference versus the strength of the GraDir. Not surprisingly, we found that the filter's internal structure is essential for detecting the GraDir: The odd symmetrical filters with low spatial frequencies (LSF) respond better to luminance gradients than their HSF counterparts, which in turn respond better than the even symmetrical filters with LSF and HSF, respectively. In addition to being detectors for oriented bar and line features, the receptive fields of simple cells are suited to discern luminance gradients, [14]. Figure 5.B suggests that filters with odd symmetry outnumber the even filters in the two learned sets. In fact, the LSF set hardly contains any even filters. This observation is consistent with previous neurophysiological studies on spatial structure of simple cells in primary visual cortex [24].

In order to study the segregation performance of gradients versus no-gradients, we tested our model with synthetic input images. The latter images consisted of a sinusoidal grating (a gradient feature) with orientation ψ that was superimposed on a luminance edge (a non-gradient feature) with orientation ρ. Figure 6.A shows a pair of such images with different orientations ψ and ρ. Next, for each stimulus, and for each learned filter, we measured the reconstruction correctness (see legend of Figure 6) of the resulting gradient representation and the (input) grating, and that of the non-gradient representation the (input) edge. The results (shown at the bottom row of Figure 6.C) suggest that gradient segregation is determined by the correlation between a filter's orientation and grating orientation ψ. The lowest gradient activity is obtained when the filters orientation is perpendicular to the grating. With the latter configuration, the LSF result for the gradients in Figure 6.C suggests that the gradient layer is nevertheless successfully generated. This is because of the nearly isotropic filter kernels, which produce relatively high responses at all orientations. For generating a gradient representation, only relatively few iterations of Equation 8 were required for both filter sets (i.e., LSF and HSF).

The computation of non-gradient representations differs in three aspects from the gradient representations (second row of Figure 6.C): (i) non-gradient responses were less correlated with the filter's orientation relative to the orientation of the luminance edge ρ; (ii) in general more iterations of Equation 8 were required for building the non-gradient layer, but then (iii) the LSF

set required fewer iterations than the HSF set. For the full reconstruction (top row of Figure 6.C), the highest reconstruction correctness (for almost all filters) occurred if ψ, ρ and the filters orientation were the same. The latter case reveals a limitation of our model. As obvious from the lower left image of Figure 6.B, the segregation into gradients and non-gradients is not achieved cleanly, in the sense that the luminance gradient layer contains residual non-gradient components.

3.2. Representations are Created with a Filling-in-like Process

The previous results suggest that our model essentially implements a filling-in process. This can be understood by considering the filter kernels as diffusion operators. Specifically, odd-symmetrical filters implement a diffusion operator which propagates activity along a certain direction and orientation, while even symmetrical filters are bi-directional and oriented. As a consequence, image reconstruction by means of Equation 8 proceeds according to a filling-in process [25]. Figure 7 illustrates filling-in with a luminance staircase as input image. At 20 iterations a saw-tooth-like profile is represented in the non-gradient layer (Figure 7C). The reconstruction process continues until the stopping criterion is satisfied (Equation 10), and at around 70 iterations the staircase is fully recovered. The filling-in-like reconstruction process furthermore implies that non-gradient activity is propagated across edges. Observe that relatively few iterations are needed for building the gradient representation (Figure 7B). Note that the error function (Equation 10 as a function of iterations) can decrease in a non-monotonical way (Figure 7D). However, this did not cause any problems in our simulations, as our model converged (globally) to a stable solution in all considered cases.

3.3. Segregation Performance with Real-World Images

Although synthetic images are a valuable tool for the study of certain aspects of the visual brain, the visual system evolved for the efficient processing of real-world images. Real-world images provide therefore a test of robustness for any model of the visual system. Figure 8 and 9 illustrate that our model successfully segregates gradients and non-gradients from real-world images. The gradient layer converges faster than the non-gradient layer. In the latter, non-gradient activity is propagated rather slowly from the contours to the interior of surfaces. This is because of: (i) Solving Equation 8 proceeds according to a filling-in process when derivative filters (i.e. our even or odd symmetrical filter kernels, cf. Section 3.2) are used [25]; and (ii),

in natural images, gradients are more abundant and spatially more extended than non-gradients. This means that in the maps with low variability the activity is (already) distributed throughout the whole image, whereas in the high variability maps the activity is concentrated at surface contours. A perfect segregation of non-gradients would imply that any gradients are removed from the non-gradient representation. However, Figure 8 (both layers) and Figure 9 (second row at 80 iterations) suggest that the segregation is not (entirely) perfect, as residual gradients may still exist in the non-gradient layer (e.g., at the hat of Lena image). Specifically, for the Lena image (Figure 9), this can be explained with that the model erroneously detects an illumination edge as a non-gradient feature.

To study this issue in more depth, we tested the segregation performance of the model for natural images with a superimposed (simulated) sharp illumination edge or a soft shadow boundary (Figure 10). The gradient layer perfectly captures the soft shadow boundary (Figure 10B), and therefore the non-gradient layer contains by and large the input image (Figure 10C). For the sharp illumination edge (Figure 10A), on the other hand, the gradient representation is hardly distinguishable from the gradient representation of the original Peppers image (Figure 10C), because the sharp illumination edge is captured by the non-gradient layer. Notice the halo around the representation of the sharp illumination edge in the non-gradient layer. This indicates that the illumination edge has a spatially limited influence on the rest of the non-gradients, because the corresponding activity propagation is attenuated by the presence of adjacent boundary structures. As a consequence, the dynamic range of the input is compressed into the non-gradient layer; an effect that can also be appreciated in the full reconstruction. Dynamic range compression also occurs for the soft shadow boundary, but then in the non-gradient layer, and in the full reconstruction as well.

How does the removal of high spatial frequencies affect the distinction between gradients and non-gradients? The segregation criterion is not based on low versus high spatial frequencies in the first place, but rather on low versus high variability. Accordingly, the suppression of high spatial frequencies in an input image will cause a narrowing of the range of variability values (because variability is dispersed or diluted). The segregation still works for moderate amounts of lowpass filtering of the input image. In fact, a small amount of lowpass filtering increases the stability of the segregation process, because of noise suppression. For strong lowpass filtering, nevertheless, the segregation process breaks down. In the latter case, the input image is represented in

both layers, as illustrated in Figure 11.

3.4. Prediction of Visual Illusions that Contain Luminance Gradients

Our model can be used to predict a certain class of visual illusions that contain luminance gradients. Accordingly, we predict that simultaneous brightness contrast, Adelson's checkershadow illusion [3], Logvinenko's illusion [26] and grating inductions [27, 28] are reproduced in the non-gradient layer regardless of whether the HSF filter set or the LSF filter set it used.

3.4.1. Simultaneous Brightness Contrast

The simultaneous brightness contrast (SBC) display consists of two small and gray patches with identical luminance arranged over a background with different luminance. The patch on the dark background is perceived as brighter than the patch on the bright background. Figure 12 shows that our model correctly predicts this illusion with the non-gradient layer. SBC is usually attributed to low-level processing. For example, retinal ganglion cells may enhance the patch contrasts, but recent studies suggest that SBC may involve higher level processing as well [29]. In our model, SBC is reproduced by the difference between the low variability of the background and the high variability at the contours of the patches. As a consequence, our model assigns the patches to the non-gradient layer, where their (predicted) brightness depends on their respective luminance difference with the local background. Clearly, SBC is also perceived with a luminance step as background. It would be therefore interesting to study experimentally for which background (gradient or step) the effect is perceived stronger [26]. If the smooth luminance gradient was replaced by a sharp luminance step (Figure 13), then our model predicts only a minor difference between the patches in the non-gradient layer. Furthermore, if the stopping criterion (Equation 10) was removed from the reconstruction process, an assimilation effect (albeit small) would be predicted from a certain iteration on (see the profile plot of the non-gradient layer of Figure 13). This is because the response maps with high and low variability are generated via the contour map C, where $0 \leq C < 1$. Since non-gradient activity is determined by $1 - C$ (see methods), a small part of the activity of the non-gradient layer is absorbed by the gradient layer in the reconstruction process, even though the stimulus is not composed of gradients.

The stopping criterion that we adopted seems to be ad hoc at first sight. But a second look suggests that it could be a plausible metaphor for corre-

sponding mechanisms in the brain. Since perception is highly dynamic, it is conceivable that it occurs in time slices [30]. If this was true, then a neural mechanism could actively reset a representation at the beginning of a new time slice. In the context of filling-in, this implies that activity can only propagate within the duration of a time slice, what in essence corresponds to a stopping criterion. Further evidence for a mechanism that could possibly be related to a stopping criterion comes from [31, 32]. These results suggest that contrast integration is distance dependent, what essentially amounts to that distant boundaries will have a reduced effect on the activity in a layer that represents a target surface. [32] also suggested that contrast polarity (dark-light versus light-dark directional luminance changes) contributes differently to lightness. In summary, the stopping criterion could be related to more plausible mechanisms for determining non-gradient activity, and we plan to consider such mechanisms in future versions of our model.

3.4.2. Adelson's checker-shadow illusion

This display consists of a checkerboard with bright and dark squares, where a part of the checkerboard is shaded. The bright squares in the shaded region have the same luminance as the dark squares in the bright region, yet they are perceived as being different. Figure 14 shows two versions of the checker-shadow display: One with a smooth "shadow boundary", and another one with a sharp "illumination edge". As before, our model reproduces both versions of the illusion in the non-gradient layer, because it renders all white squares with nearly identical non-gradient activity (analogously all dark squares are rendered with approximately the same activity). Our model can reproduce both types of the illusion, albeit in different ways (see profile plots of Figure 14C). In the version with the smooth shadow boundary, the gradual changes are captured by the gradient layer, whereas the non-gradient layer contains only the checkerboard. With the sharp illumination edge, however, the gradient layer is "empty", and the illumination edge is represented in the non-gradient layer. Because of the boundaries of the small squares, the influence of the illumination edge stays local its contrast cannot propagate very far (cf. Section 3.3 and Figure 10). Therefore, irrespective of the (shadow or illumination) boundary being smooth or sharp, Adelson's checker-shadow illusion can principally be accounted for by low-level mechanisms, more specifically on grounds of variability. This result contrasts with the original study in which the checker-shadow display was explained in terms of (double-reversing) X-junctions that are formed by

any four adjacent squares. The presence of X-junctions means that a contrast change is interpreted (presumably at some higher level) as a change in non-gradients rather than a shadow [33].

3.4.3. Logvinenko's Illusion

In the Logvinenko's illusion (Figure 15A), all diamonds have the same luminance, but those in the shadow are perceived as being brighter than those in the lit regions (shadowed and lit regions are produced by a superimposed grating). The Logvinenko's illusion can be considered therefore as a variation of the checker-shadow display, and is correctly predicted by our model in the non-gradient layer. As with Adelson's checker-shadow illusion, our model suggests that Logvinenko's illusion can be produced by low level mechanisms: The gradient layer captures the sinusoidal grating, while the non-gradient layer captures the reflectance estimates. Notice the difference between the non-gradient activity and the activity of the full reconstruction in Figure 15B: Whereas the full-reconstruction activity of the brighter and darker perceived diamonds are identical (top plot), the corresponding activities in the non-gradient layer are different and consistent with perception.

3.4.4. Grating Induction

Figure 16 shows the grating induction (GI) display. The GI display consists of two sinusoidal gratings (inducers) separated by a gap (test field). When the two inducer gratings are in-phase, then an illusory brightness modulation is perceived across the test field. The brightness modulation has the opposite phase as the inducers. Brightness modulation decreases when shifting the phase of the inducer gratings, being minimum when the gratings are in anti-phase. The illusory modulation is furthermore attenuated with increasing separation and spatial frequency of the inducer gratings. Typically, the GI is explained in terms of multi-scale filtering [34, 35]. As before, our model predicts that the illusory brightness modulation is produced in the non-gradients layer. Figure 16E shows profile plots of the non-gradient activity for in-phase versus anti-phase inducer gratings, for increasing the separation between the inducers, and for a higher spatial frequency of the inducers. In all cases, our model predictions are consistent with psychophysical results. Observe that the gradient layer contains the opposite pattern than the non-gradients layer. This means that the modulation of the test field is predicted as a direct consequence of discounting the gradient pattern, which leaves a (illusory) modulation in the non-gradients layer.

4. Discussion and Conclusions

Here, we proposed a computational model for generating representations of gradients and non-gradients with achromatic real-world images. We propose that gradients versus non-gradients can be exploited by the visual system as a possible powerful cue for detecting changes in reflectance versus illumination. This is to say that non-gradient representations can be linked to lightness computations, whereas gradient representations can be associated with shading effects in natural images. Luminance is the product of reflectance and illumination effects (e.g. shading, shadows, specular highlights and inter-reflectance). Brightness is perceived luminance, and thus should be composed of perceived reflectance (lightness) and perceived shading [14]. In other words, we hypothesize that "normally" we perceive brightness directly and lightness implicitly, as normally smooth luminance gradients are superimposed on object surfaces. Lightness could be perceived directly for flat surfaces (i.e., in the absence of overlaid luminance gradients), because in this special situation brightness equals lightness. A further advantage of segregating gradients from non-gradients lies in dynamic range reduction. Intensity values of shaded and lit regions in an image often differ by several orders of magnitude (of dynamic range). This difference can be reduced by removing corresponding luminance gradients, thus revealing (reflectance) details in both lit and shaded regions. Moreover, our simulations suggest that dynamic range compression is still achieved with sharp illumination edges. Albeit a sharp illumination edge is represented in the non-gradient layer, its effect stays localized around the edge, meaning that it does not increase the contrast for the rest of the represented non-gradient features.

Assume a sensor array with finite spatial resolution, such as the retina or a CCD. Then, with increasing viewing distance, low spatial frequencies will eventually shift to high spatial frequencies on the sensor array (while high spatial frequencies cannot be resolved anymore). This raises the question about how the viewing distance interacts with the perception of luminance gradients, and thus with distinguishing between reflectance and illumination changes based on spatial frequency. Neglecting optical constraints, the physical limit for representing a luminance ramp across a sensor array is three pixels. Increasing the viewing distance would map this minimal ramp to two pixels. This means that at a short distance, the ramp would be represented in the gradient layer, whereas at a larger distance, it would be represented in the non-gradient layer. Our model therefore predicts that when viewing

smooth luminance gradients from a sufficiently large distance they could be interpreted as reflectance changes. However, as outlined in the introduction, the visual system uses further cues to distinguish between reflectance and illumination, such that the drop out of one cue could be compensated by the others. We are only aware of corresponding psychophysical data in the context of luminosity perception (i.e., the perception of light emitting surfaces), which were successfully predicted by one of us ([14], Figure 8). The computational approach of ([14] is based on a similar idea as the present one: The suppression of high spatial frequency structures with subsequent diffusive reconstruction of a representation for luminance gradients. According to the model of ([14], the perception of luminosity depends on the presence of linear luminance gradients ("ramps"). With increasing viewing distance or spatial frequency, respectively, these ramps become narrower. Accordingly, the dependence of perceived luminosity from spatial frequency (or viewing distance) shows an inverted-U behavior (perceived strength versus spatial frequency), exactly as it is the case with Mach bands [10, 12, 13]. Returning to our initial question, the inverted-U curve suggests that smooth luminance gradients caused by illumination may indeed be interpreted as reflectance when viewed from a large distance.

The relationship between brightness and lightness computations (and perceived gradients) is especially important for studying visual illusions. In line with previous approaches (e.g., [36]), our model successfully reproduces a couple of those that include luminance gradients: Simultaneous brightness contrast (with a smooth luminance gradient as background), Adelson's checker-shadow illusion [3], Logvinenko's illusion [26], and grating inductions [27, 28]. As our model predicts most of these illusions with non-gradient representations, we suggest that they could be explained as "errors" in lightness computations. Notice that some of the aforementioned illusions are frequently explained by higher-level mechanisms in the visual system which supposedly has "knowledge about the illumination". Our computational model suggests that high-level knowledge may not be strictly necessary for their explanation, but then our model does not distinguish between different types of luminance gradients (shadows, shading, etc.), neither can it classify a sharp illumination edge as illumination feature. Therefore, *higher level mechanisms and/or chromatic mechanisms* may still be critical for recognizing and interpreting the different types of luminance gradients, and for identifying sharp illumination edges, meaning that we do not claim that our low or mid-level computations (or intermediate representations) are sufficient for determining

lightness. Nevertheless, the advantages of having intermediate representations of (non-)gradients (on which higher level mechanisms can operate) were elucidated before [37].

Technically, our model could be subdivided into two independent computational steps: (i) two sets of filter kernels are first learned from natural (or real-world) gray-level images using sparse coding techniques, and (ii) once learning is completed, the filter kernels encode a luminance image in response maps. The response maps are analyzed for local variability and are transformed into two sets, where one corresponds to response maps with high variability (non-gradients), and another one with low variability (gradients). (In the absence of noise, computation of local variability essentially amounts to computing a contour map). Since the segregation of gradients from non-gradients is an inverse problem (more variables than equations), we used an optimization algorithm to find a solution.

On first sight, our approach underlies similar assumptions as Retinex theory [2]. In Retinex theory, high contrast changes across space are classified as reflectance edges, whereas low contrast changes (corresponding to illumination effects) are suppressed. It has been pointed out that Retinex would wrongly classify sharp illumination edges as reflectance changes (see [17], with references). This behavior is therefore identical with our model, where sharp illumination edges (that normally are associated with high activity in the variability map) enter the non-gradient representation. However, unlike Retinex, we do not classify gradients and non-gradients based on contrast strength: Gradients and non-gradients are distinguished according to blurred (i.e. low variability) versus sharp (i.e. high variability) contours, irrespective of contrast amplitude. In fact, this mechanism is essential for reducing dynamic range, as illumination edges can have far higher contrasts than reflectance edges. This is to say that we relax the constraint that an increase in the degree of blurring always leads to contrast reduction in natural images (assuming conservative diffusion mechanisms). This assumption seems to make sense because of the independence of reflectance and shading edges at the local level. At a more global level (e.g. whole scene), however, it is clear that, for example, any shadow boundary has to be caused by an object that occludes the illumination source. Furthermore, our contour map is adaptively computed based on the variability of the decorrelated response maps. Because of adaptation (Equation 7), when removing high spatial frequencies from an input image, it is still decomposed into gradients (low variability) and non-gradients (high variability).

Dynamic range reduction therefore cannot be achieved with the standard Retinex algorithm, as reflectance edges are assumed to have always higher contrast than illumination effects. Some published models classify reflectance boundaries based on chromatic variations in space (e.g. [17]), thereby assuming that achromatic variations in an image are due to illumination effects [16]. For instance, the model presented in [17] converted a color image from RGB into a red-green (RG) channel, a blue-yellow (BY) channel and a luminance-contrast (L) channel. The cue for distinguishing reflectance from illumination effects is that the latter cause variations predominantly in the L-channel, but also in the BY-channel, although to a less degree. By using an adaptive threshold, the two chromatic channels were combined into one (chromatic) map, and edge detection was applied. This binary edge map subsequently served to classify the boundaries of RG, BY, and L channels, where unclassified boundaries were suppressed. Finally, in order to recover estimates for reflectance and shading, the three pruned boundary maps were subjected to inverse filtering [38]. The similarity between this model and ours lies in the classification of edges, albeit according to different cues: [17] detect chromatic versus achromatic variations across space, whereas we detect sharp versus gradual achromatic variations.

Similarly, the model proposed by [39] assumes that illumination-related variations multiply the RGB vector corresponding to reflectance by a scalar value. With this setup, they framed the segregation of reflectance from illumination as an optimization of an energy function. The energy function consisted of a smoothness constraint on illumination effects, a Retinex term ("gradient consistency"), and a term that enforces a small number of different reflectance values based on clustering (global sparseness constraint on reflectance). The relative weighting of the three energy terms and the two threshold values amount to four free model parameters (optimized for an image or across a set of images). Their method outperformed other methods, for instance the one proposed by [40], or standard Retinex [2]. The latter method revealed a tendency to smooth across contours.

The approach from [41] used a multi-scale wavelet decomposition of the input image. In this way, the input image was transformed into a representation where actual reflectance values were stored at some location and scale, thereby implementing local sparseness (i.e., most of the wavelet coefficients are very small or zero). Similar to [39], a further (global) sparsity constraint was imposed on the number of different reflectance values. Again, the local and global sparsity constraints were used together with a smoothness con-

straint on illumination effects to minimize a cost function. In that sense, our approach also uses local sparsity, as the input image is processed by LSF and HSF filter kernels.

Our original idea was to use the initially learned HSF set for building the non-gradient layer, and the LSF set for composing the gradient layer. The underlying reason is that high variable patterns in the input are supposed to be better encoded in the filter responses of the HSF set, while low variability is associated with LSF filter responses. Nevertheless, because the two filter sets are not mixed during layer computation, and the segregation according to variability could be applied to any of the two learned sets (HSF and LSF), both sets would be equally suitable in order to compute any layer. In other words, the important step is the variability-based segregation, but not the choice of a specific filter set. Layer computation based on the LSF set nevertheless converges faster with Equation 8, because the filter kernels essentially implement a diffusion operator (cf. Section 3.2), and the (bigger) LSF kernels advance activity across larger distances per time step, as compared to the HSF filters. Therefore, this means that our model implements a filling-in process [42, 8, 9, 43, 44], where layer computations with the LSF set reaches a (semi-stable) solution within a few iterations.

We also identified a functional role of the filter kernels: Odd symmetrical LSF filters respond the best to smooth luminance gradients, followed by their HSF counterparts. Even symmetrical filters (both sets, i.e. LSF and HSF) respond only poorly to gradient features. This is in line with a previous proposal which holds that high (spatial) resolution odd symmetrical filters are best suited to detect reflectance changes, even symmetrical filters detect "texture" features (e.g. lines and points), whereas low-resolution filters of both symmetries are suitable for detecting smooth luminance gradients [10, 12, 13, 14].

The notion of identifying the variability of contrast patterns with certain semantic features of an image is quite different from multi-scale approaches to brightness perception (e.g. [7, 6]). Multi-scale approaches typically compute the responses (filter coefficients) of a set of oriented band-pass filters with different sizes to an input image. Subsequently, filter response amplitudes are modified according to pre-defined criteria, such as the contrast sensitivity function or normalization. Finally, a representation (normally identified with brightness) of the input image is computed by using the modified amplitudes as local weights for generative band-pass kernels. As these models are targeted to explain brightness phenomena, they usually do not address

the segregation of illumination and reflectance explicitly, and do not propose mechanisms for suppressing illumination patterns.

Although the filling-in process would be excellently suited to suppress low-contrast illumination patterns (similar to Retinex theory), low-contrast activity in filling-in models is often maintained (e.g. [9]), in order to enable "imprinting" of smooth gradient patterns on surfaces (boundary webs, [45]), and explaining in this way a larger set of brightness illusions. In other words, typical filling-in models (such as [9, 8] usually target brightness phenomena, and as such usually do not address the segregation of reflectance from illumination. In fact, in some of these models the output is identified with perceived reflectance (lightness, e.g. [36, 25]), yet they still contain illumination patterns. Moreover, different to our approach (and the filling-in mechanism proposed in [10]), some of the cited models add a lowpass-filtered [9, 7] or even a full version [25] of the input image to their output in order to recover absolute brightness levels.

The main limitation of our model with respect to visual illusions is that the present version cannot predict those which are devoid of luminance gradients. This includes "flat" displays that are well predicted by some of the cited models above, such as Whites effect or assimilation displays. A possible remedy for our model could be to introduce a gain control mechanism or response normalization into the high variability map. As a result, the classical simultaneous contrast could be perfectly reproduced (preliminary data not shown), but then again this requires to modify some of the models parameters. Another remedy could be to superimpose an "invisible" luminance gradient over the input, which, in terms of perception, would not reach threshold. For suitable amplitudes of these "invisible gradients", they will be discounted by our model, and lead to contrast enhancement or reduction in the non-gradient layer. Examples for visual illusions that could be explained by "invisible" gradients include the Craik-OBrien-Cornsweet effect or the "phantom illusion" [46], in which, narrow (or sharp, respectively) imperceptible luminance gradients may produce a contrast (or assimilation, respectively) effect. In this context, it is interesting to note that many filling-in models produce such shallow gradients from the boundaries towards the interior of filled-in surfaces (depending of course on the specific type of filling-in equation under consideration). Thus, an extension of our model could in principle operate on a filled-in brightness representation, assuming that brightness is composed of lightness and (perceived) luminance gradients [14].

A further limitation concerns the incomplete segregation of reflectance

from illumination if the orientation of a luminance gradient spatially coincides with, and has the same orientation as, a reflectance edge (Figure 6). This reflectance-gradient crosstalk may also influence the correct prediction of a bigger set of visual illusions. We therefore plan to study mechanisms for improving the segregation in future versions of our model.

5. Founding and Acknowledgments

We thank the Visca Group for laboratory assistance, Flavia Loren and August Romero for useful discussions. Funding for this project was provided by the Ministry of Economy and Competitiveness (Spain). Project numbers: PSI2010-18139-P, PSI2013-41568-P and PSI2014-57454-P.

6. Figures Caption

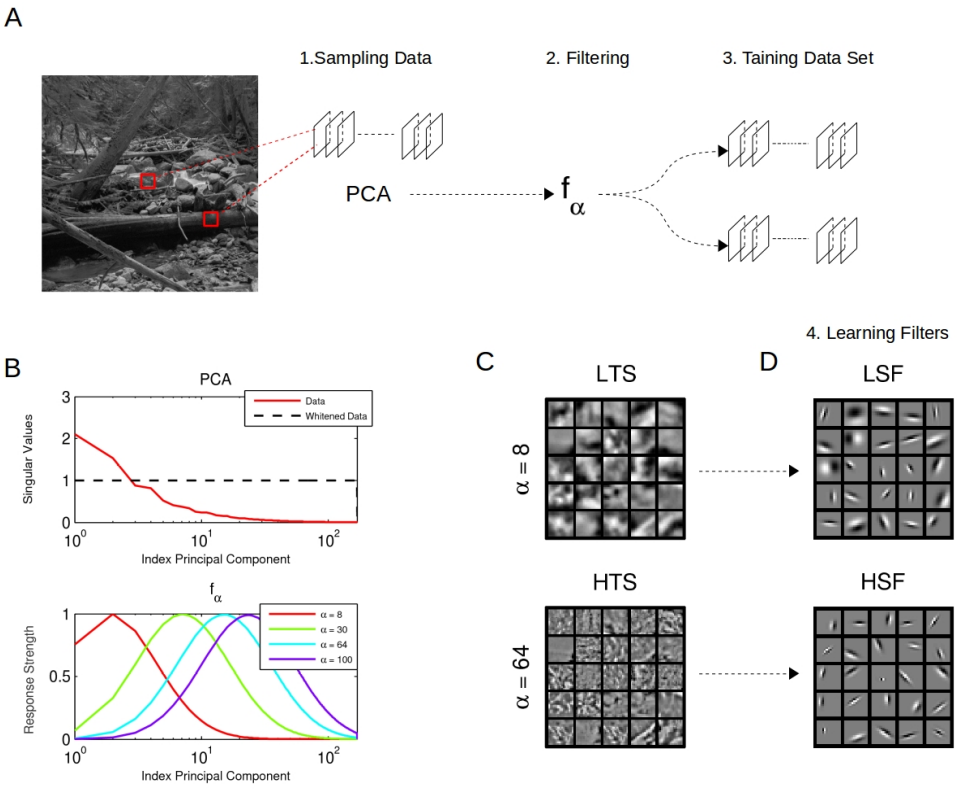


Figure 1: (A) Step-by-step illustration of the procedure for building our two training sets: (1) A set of samples is extracted from natural images and its principal components are computed, (2) the principal components are filtered according to low and high spatial frequencies, what yields (3) two training data sets. (B) Top: Singular values of our training data. Bottom: The response strength of the spectral filter f_α (Equation 1) for different values of alpha. (C) A subset of the training patches with high (HTS) and low (LTS) spatial frequency content. (D) Low (LSF filters, $\alpha = 8$) and high spatial frequency filter kernels (HSF filters, $\alpha = 64$) that were learned from the respective training data set. For illustration, only learned filters are shown.

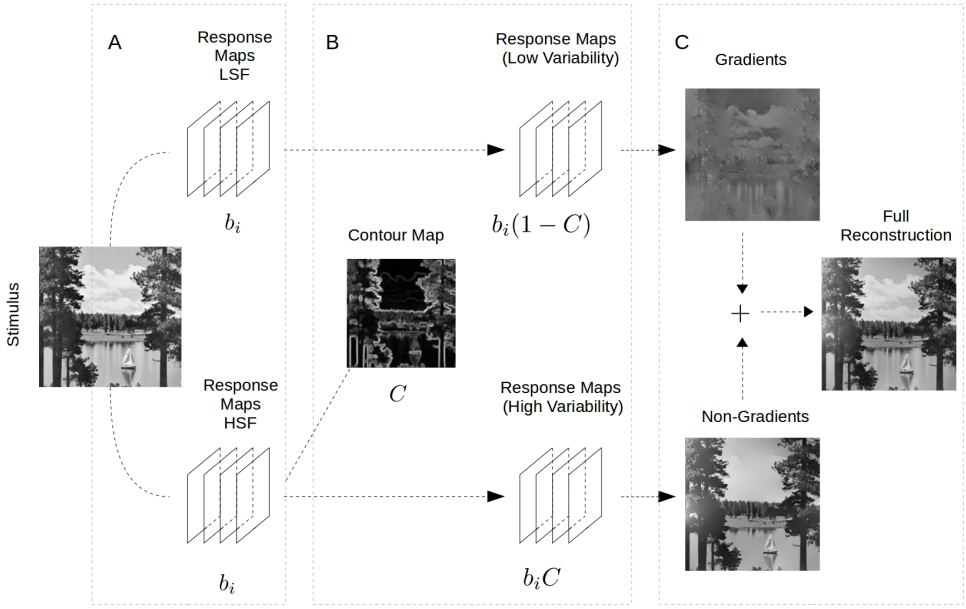


Figure 2: Illustration of the three principal stages of image decomposition. Formulas and mathematical symbols are defined in the text. (A) At the first stage we computed the responses of each filter to an input image, resulting in a set of response maps b_i for the learned HSF and LSF filters, respectively. Subsequently the response maps of a set were locally encoded by removing the correlations between themselves. (B) In the second stage, a variability map is computed across each set of response maps. This means extracting the pixels of a 3x3 neighborhood from each response map, where the center pixel has the same position across all response maps. We then computed the standard deviation with all the pixels corresponding to identical center positions (Equation 6). By thresholding of the variability map with its mean standard deviation, and applying a suitably parameterized sigmoidal function to the super-threshold values (Equation 7 and Figure 3, respectively), we obtained a contour map with values ranging from zero to one. With the contour map, we extracted a set of high variability maps and another one of low variability maps from the response maps. The former set results from multiplicative gating of the decorrelated response maps with the contour map, and the latter set is the result of multiplying the decorrelated response maps with one minus the contour map. (C) Luminance gradients (gradient layer) were estimated with the low variability maps, whereas the non-gradients were estimated with the high variability maps.

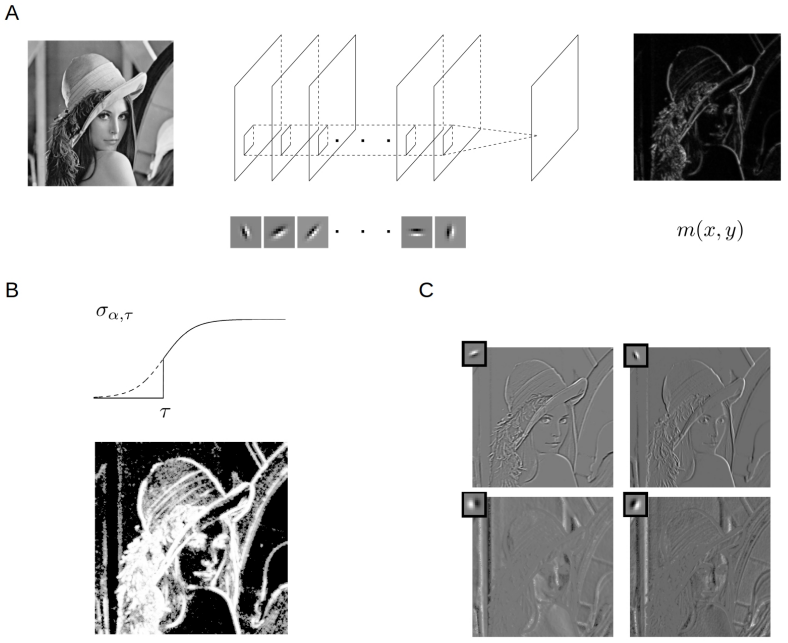


Figure 3: Illustration of the contour map, and the high and low variability maps. (A) The variability map $m(x, y)$ for the Lena image (Equation 6). (B) A sigmoidal function $\sigma_{\alpha,\tau}$ is applied to the variability map, with its threshold τ chosen such that the sum of those $m(x, y)$ with $m(x, y) > \tau$ (i.e. the superthreshold values) is just one percent of the total sum of all $m(x, y)$. The rate parameter α of the sigmoid is equal to $1/\text{mean}[m(x, y)$. In order to convert the sigmoidized variability map into a contour map, values with $\sigma_{\alpha,\tau}[m(x, y)] < \tau$ are set to zero. Note that the highest activity of the contour map is typically located at sharp contrast structures. (C) Two examples of high variability maps (top row) and two examples of low variability maps (bottom row), corresponding to the shown filter kernels (inset). These maps were computed via $b_i C$ and $b_i(1 - C)$, respectively, where b_i represents the decorrelated response maps (Equation 5).

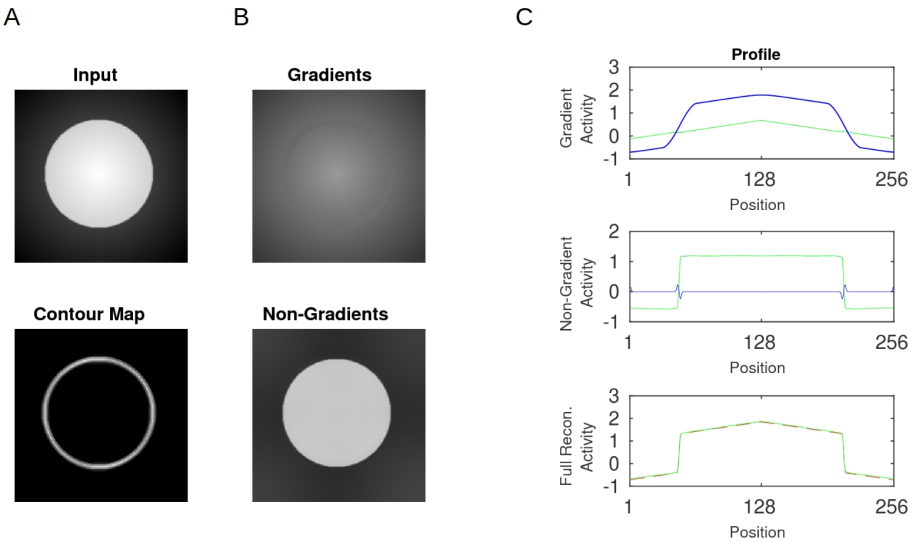


Figure 4: Illustration of segregating an input image into gradients and non-gradients. (A) A homogeneous disk with a overlaid gradient corresponds to the input image. The corresponding Contour Map as computed by Equations 6 and 7. (B) The result of segregating the input image into Gradients and Non-Gradients by means of Equation 8. (C) Profile plots of the reconstructed layers. Top: Gradient layer activity (green line) along with the profile plot of a lowpass filtered version of input image (blue line) illustrates demonstrates that gradient activity is not just equal to lowpass information. Middle: Analogous profile plots for the non-gradient layer compared with a highpass filtered version of input image show that non-gradient information is not the same as to highpass filtering of the input image. Bottom: The profile plot of the full reconstruction (green line) compared with the input (dashed red line) shows that no information of the input image is lost due to the segregation process.

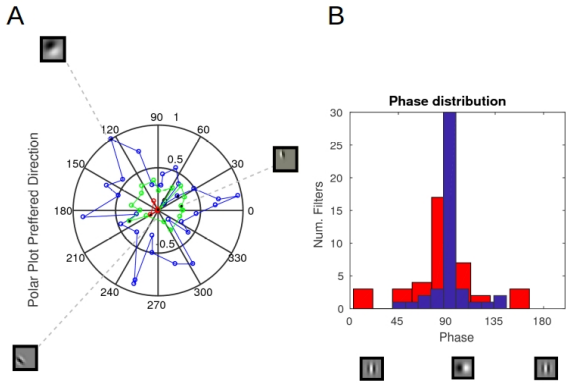


Figure 5: The spatial structure of the filters determines the response properties of luminance gradients. (A) Polar plot for the learned filters showing the normalized strength and the preferred luminance gradient direction (in degrees). The blue dashed line represents odd filters with LSFs (mean response 0.5314), the green dashed line the odd filters with HSFs (mean response 0.2704). The red dashed line represents even symmetric filters these are only found in HSF set (mean response 0.0409). (B) Phase histogram for the two sets of learned filters: red indicates the LSFs, and blue indicates HSFs, respectively, in units of degree. A phase of 90 degrees corresponds to an odd symmetric spatial structure of the filter, whereas 0 and 180 degrees to even symmetry.

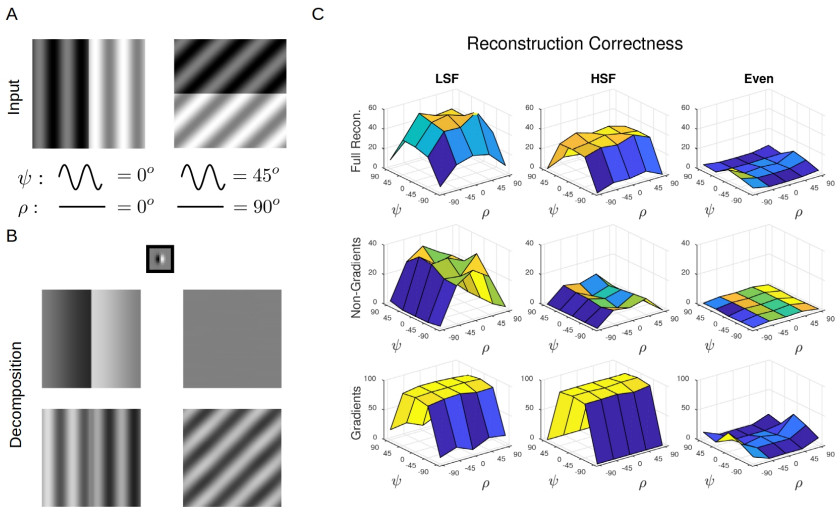


Figure 6: Segregation performance of non-gradients versus gradients. (A) Two examples of synthetic test images, where a luminance step represents a non-gradient feature and the grating a gradient feature. The test images are parameterized over grating orientation ψ and step orientation ρ. (B) Segregation of the test images of (A) into non-gradients and gradients, respectively, with the filter shown in the thumbnail image. (C) Reconstruction correctness for the Full reconstruction, Non-Gradients, and Gradients, using the LSF, HSF, and Even-symmetric filters. The "reconstruction correctness" is measured by $c = 100(1-r)$ where r is the relative error between the reconstructed feature and the corresponding feature of the test image. The surfaces represent the "reconstruction correctness" of all learned filters of LSF, HSF and Even. The domain of the surface plots is represented by the relative orientation and with respect to the gradient direction (GraDir) of the learned filters. All simulations involved 10 iterations of Equation 8.

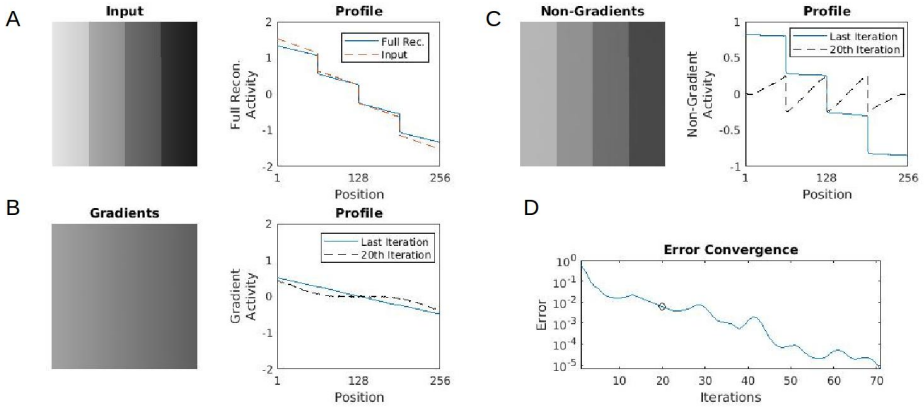


Figure 7: (A) Luminance staircase with an overlaid linear gradient along with corresponding profile plots: Original input (red broken line); full reconstruction (blue line). (B) Gradient layer and its corresponding profile plots: gradient activity (blue line); representation at 20 iterations (black broken line). (C) Non-gradient layer with corresponding profile plots: non-gradient activity (blue line); representation at 20 iterations (black broken line). Activity propagation is local at the beginning, but it eventually spreads out across the entire image. (D) Log-Error versus iterations. The black circle marks the error at 20 iterations. Despite local maxima, the overall error is decreasing, what indicates convergence.

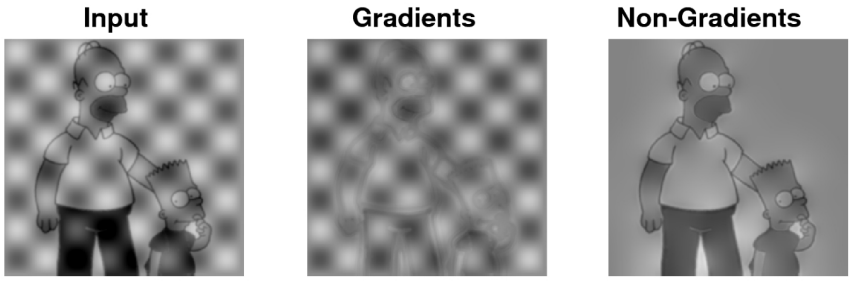


Figure 8: Illustration of gradient versus non-gradient segregation for a cartoon image with superimposed soft checkerboard grating. The first image is the Input, the second image the gradient representation and the last image the non-gradient representation. Notice the "ghost"-contours in the gradient layer, and the halo-like luminance variations around the contours in the non-gradient representation.

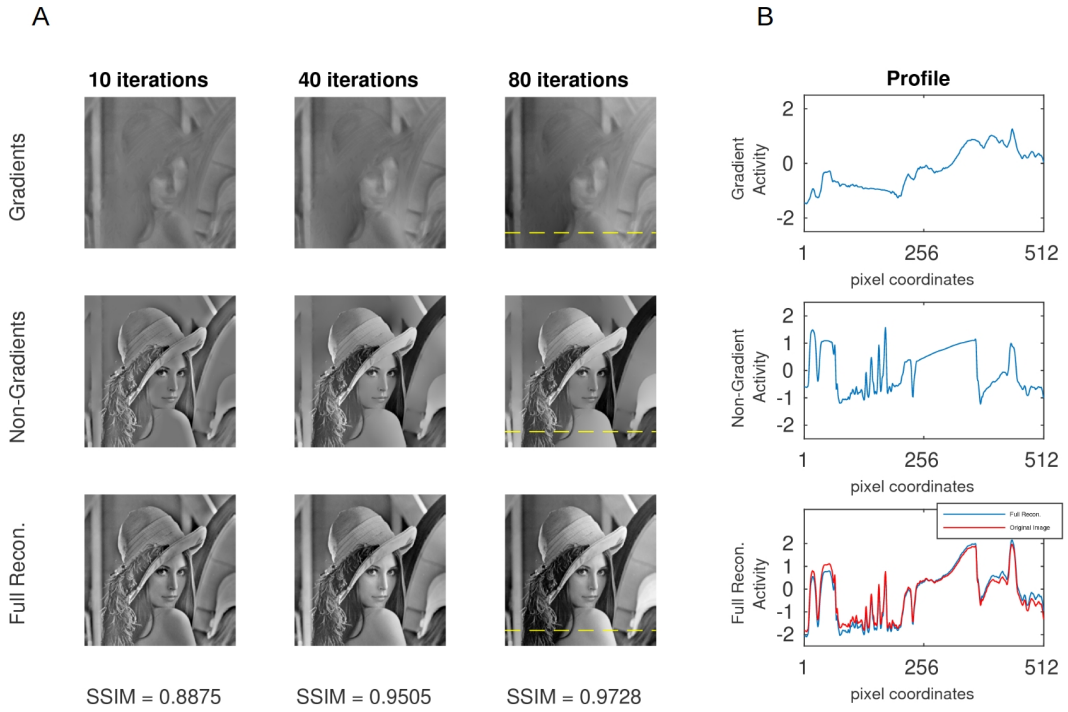


Figure 9: Gradient versus non-gradient segregation as a function of time ("Lena" image). (A) Each column is the representation at the indicated number of iterations (top) of Equating 8. At the bottom, the structural similarity index (SSIM: [47]) of the full reconstruction with respect to the original image is shown. Gradient and non-gradient representations were estimated using LSF and HSF filters, respectively. (B) At 80 iterations, the corresponding profile plots across the columns indicated by dashed yellow lines are shown.

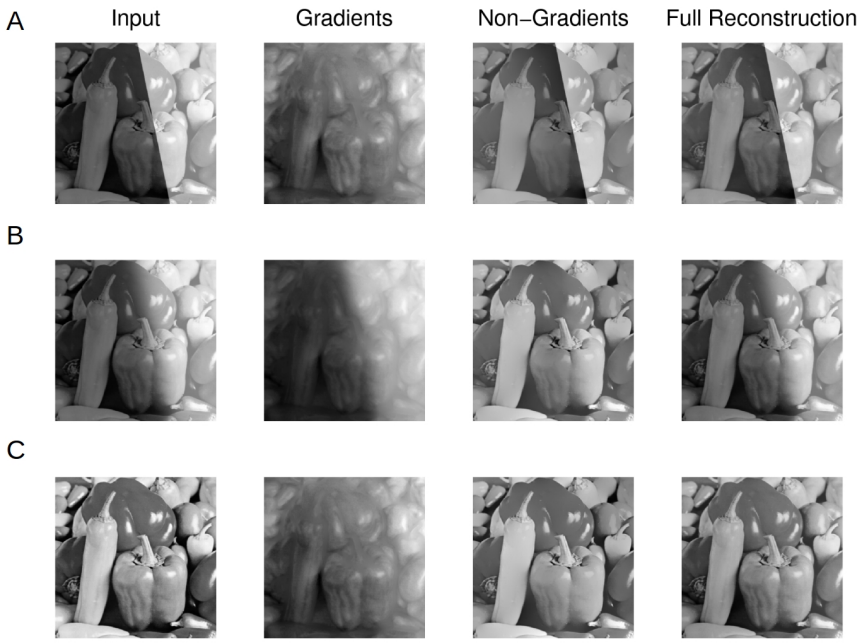


Figure 10: Gradient versus non-gradient segregation ("Peppers" image). Across the original pepper image (bottom row C) we superimposed a sharp illumination edge (top row A) and a soft shadow edge (middle row B). (A) The sharp illumination edge is captured by the non-gradient layer, but not by the gradient layer. The originally strong contrast of the edge is spatially confined in its non-gradient representation (halo around the edge), what leads to a compressed dynamic range both in the non-gradient layer and in the full reconstruction. The gradient representation is similar to the original Pepper image (row C). (B) The soft shadow boundary is captured by the gradient layer, what causes a corresponding contrast reduction in the non-gradient layer and thus dynamic range compression (also visible in the full reconstruction). (C) Results for the original Pepper image.

Figure 11: Segregation performance with suppressed high spatial frequencies. Strong lowpass filtering of the input image quenches the range of variability which is available to distinguish between gradients and non-gradients. As a consequence, the segregation gets ineffective, and the input image is principally represented in the non-gradient layer, and to a less extent in the gradient layer.

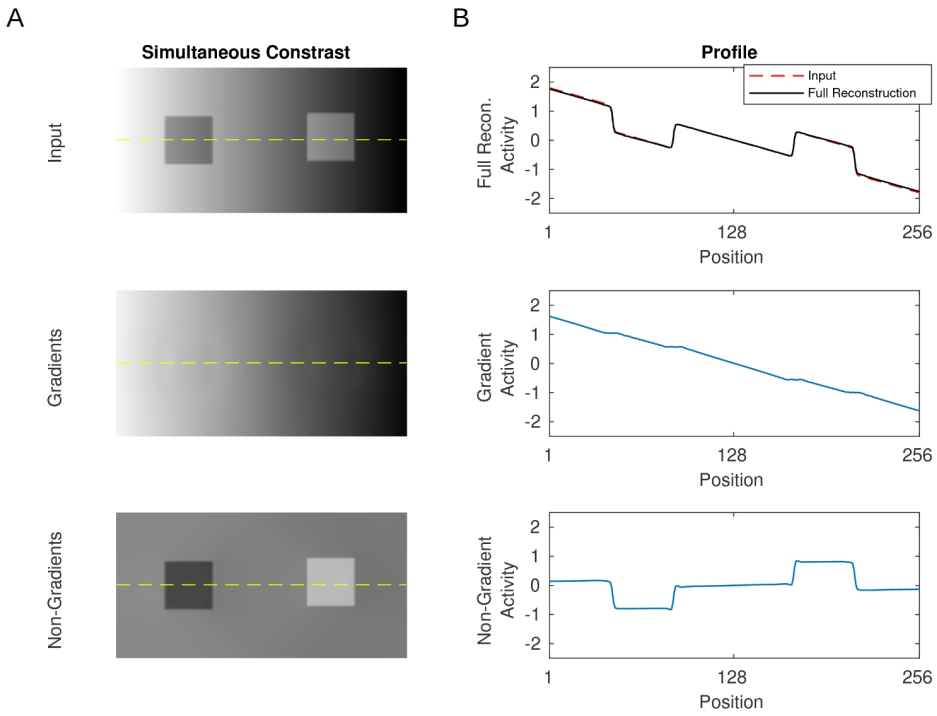


Figure 12: Simultaneous Brightness Contrast with a luminance gradient as background. (A) The estimated representations of non-gradient and gradient as predicted by our model. (B) Profile plots for the full reconstruction, gradients and non-gradients, corresponding to the yellow dashed lines in A.

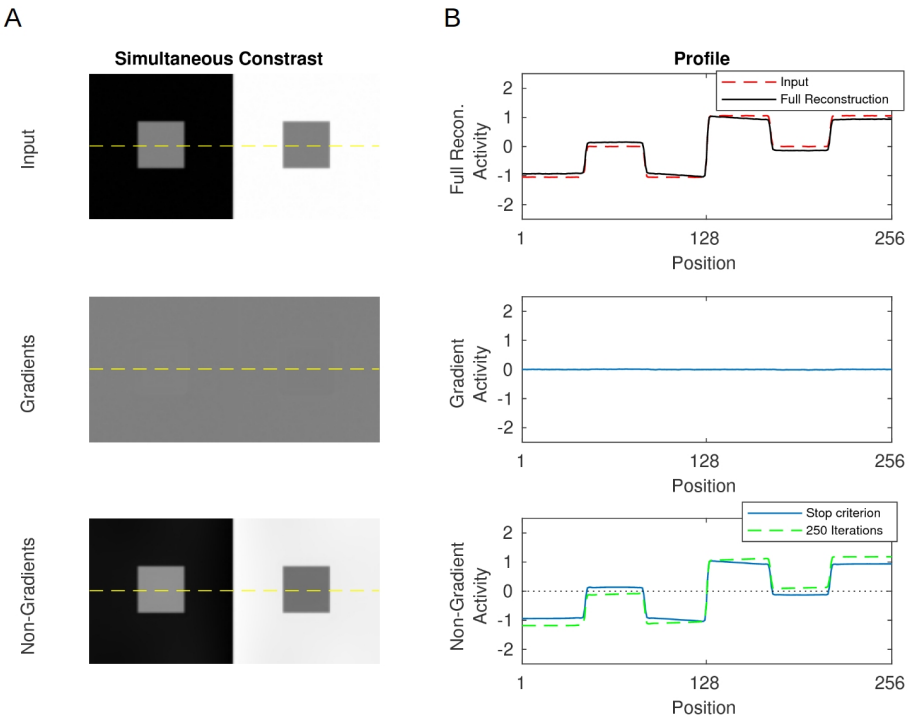


Figure 13: Simultaneous Brightness Contrast with a luminance step as background. (A) Gradient and non-gradient representations produced by our model. (B) Profile plots for the full reconstruction, gradient representation and non-gradients, corresponding to the yellow dashed lines in A. Bottom, the profile of the non-gradient layer with the default stopping criterion (blue line) versus without stopping criterion at 250 Iterations (green dashed line). Patch contrasts in the non-gradient layer are 0.2672 with the stopping criterion; -0.2059 at 250 iterations (opposite effect!); and 1.6048 in the SBC with gradient background (Figure 10).

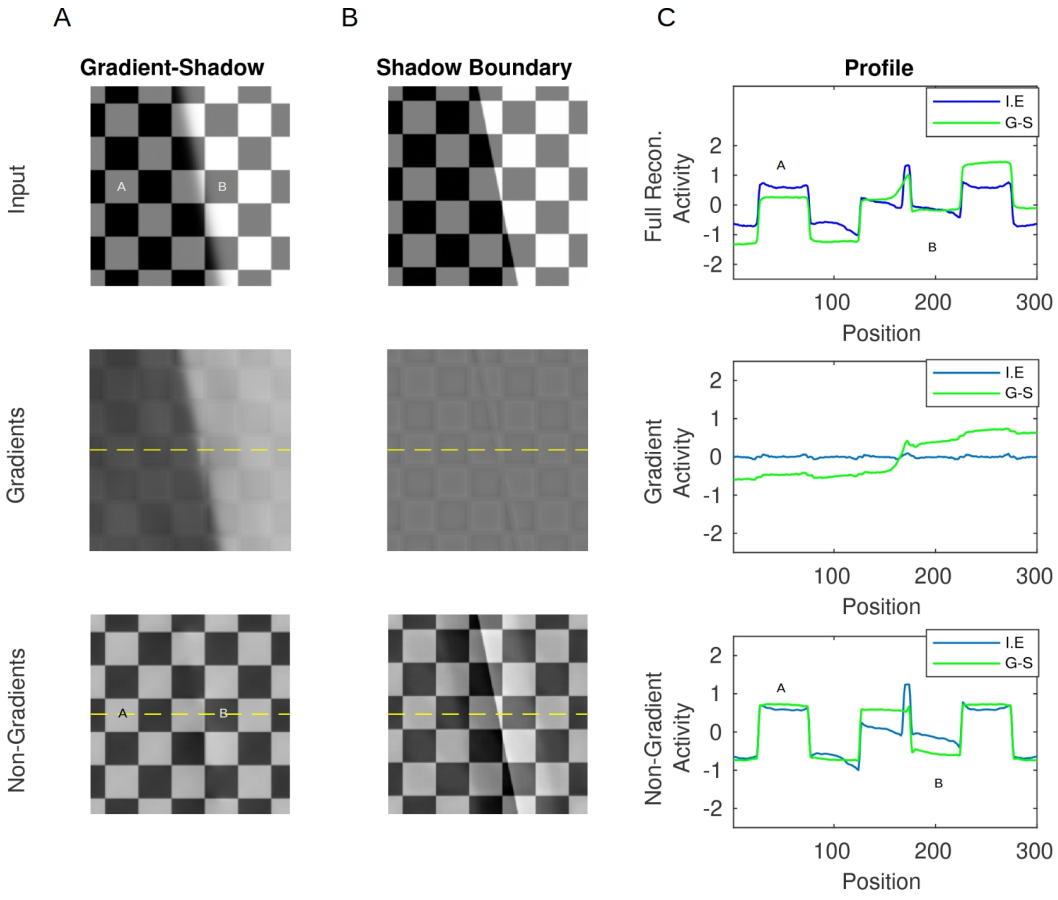


Figure 14: Model prediction for Adelson's checker-shadow illusion. (A) Adelson's checker-shadow illusion with a shadow with a smooth boundary with corresponding gradient and non-gradient representations. (B) Same as A, but now with a sharp illumination edge. (C) Profile plots corresponding to the dashed yellow lines in A and B. Top: Full reconstruction. Middle: Gradient representation. Bottom: Non-gradient representation. Legend labels refer to (I)llumination (E)dge and (G)rafient-(S)hadow, respectively.

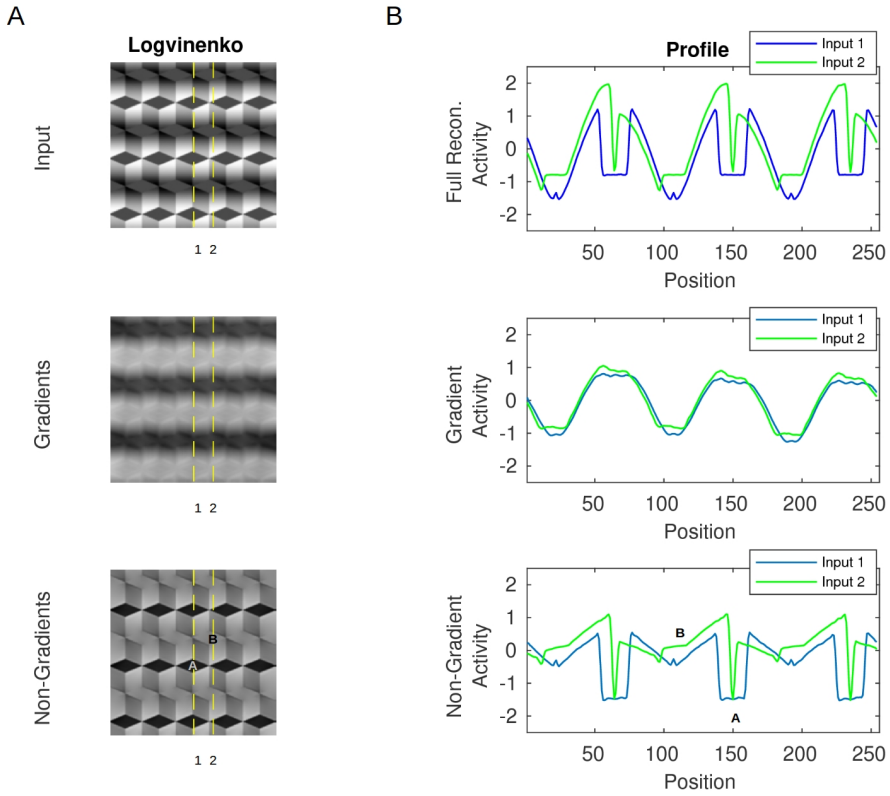


Figure 15: Model prediction for Logvinenko's Illusion. Although all diamonds have identical intensity values, we perceive them as light-gray and dark-gray, respectively, dependent on the amplitude of the superimposed sine wave grating. (A) Gradient and non-gradient representations produced by our model. (B) All profile plots correspond to the two vertical dashed yellow lines. These are labeled as "input 1" (left yellow line), and "input 2" (right yellow line) in the figure legends.

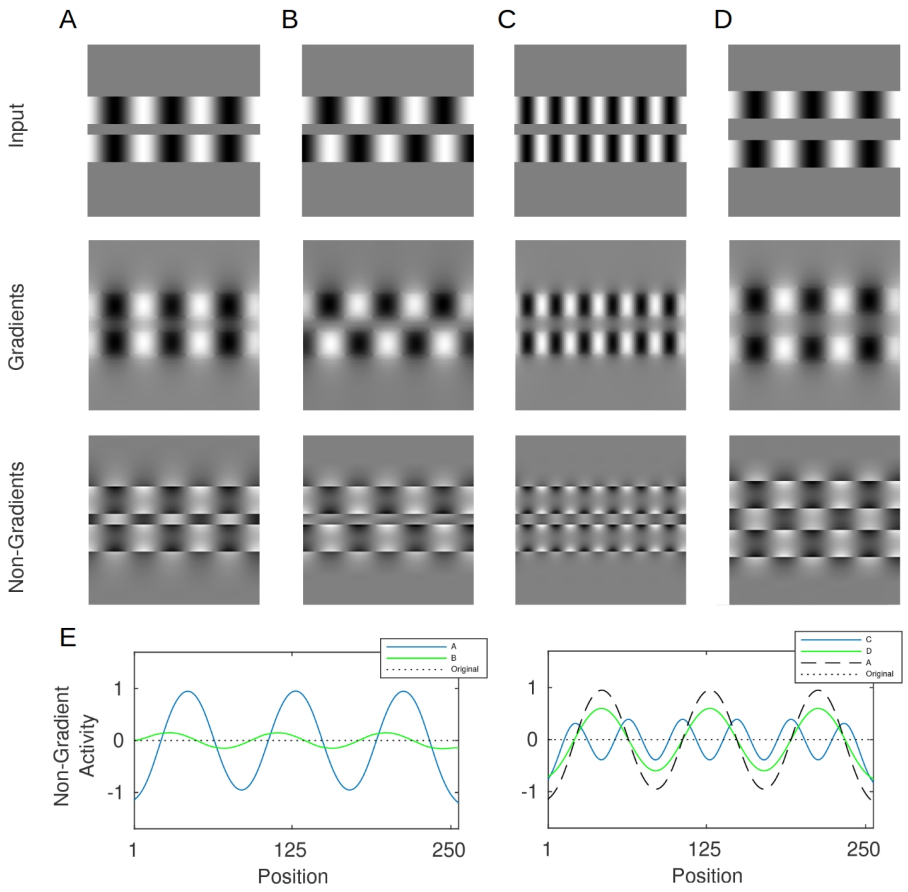


Figure 16: Grating induction (GI). The non-gradient representations for GI as they are produced by our model for varying the relative phase of the inducer grating, its spatial frequency, and test field width. (A) Standard GI display (3 cyc/image and a test field width of 16 pixels) (B) GI with anti-phase inducers (3 cyc/image and 16 pixels). (C) GI with the double spatial frequency (6 cyc/img and 16 pixels). (D) Grating induction with double test field width (3 cyc/img and 32 pixels). (E) The profile plots for non-gradient layers "in middle of the test field" compared to the original default physical grating (in dots). Notice that the non-gradient representations consistently predict that induced brightness falls off with distance from the inducing edge [48].

7. Bibliography

[1] E. Hering, Grundzge der Lehre vom Lichtsinn, Springer, Berlin, 1920, english version: Outlines of a Theory of the Light Sense. 1964. Translated by L.M. Hurvich and Dorothea Jameson. Cambridge, MA: Harvard University Press.

[2] E. H. Land, J. J. McCann, Lightness and Retinex Theory, Journal of the Optical Society of America 61 (1) (1971) 1. doi:10.1364/JOSA.61.000001.

[3] E. Adelson, Checkerboard optical illusion, Retrieved June 27, 2014 from MIT, Adelson website, http://web.mit.edu/persci/people/adelson/checkerboardshadowillusion.html. (1995).

[4] T. Agostini, A. Galmonte, A new effect of luminance gradient on achromatic simultaneous contrast, Psychonomic Bulletin & Review 9 (2) (2002) 264–269. doi:10.3758/BF03196281.

[5] J. M. Hillis, D. H. Brainard, Distinct mechanisms mediate visual detection and identification, Current biology : CB 17 (19) (2007) 17141719. doi:10.1016/j.cub.2007.09.012.

[6] B. Blakeslee, M. E. McCourt, When is spatial filtering enough? Investigation of brightness and lightness perception in stimuli containing a visible illumination component, Vision Research 60 (2012) 40–50. doi:10.1016/j.visres.2012.03.006.

[7] X. Otazu, M. Vanrell, C. Alejandro Párraga, Multiresolution wavelet framework models brightness induction effects, Vision Research 48 (5) (2008) 733–751. doi:10.1016/j.visres.2007.12.008.

[8] S. Grossberg, D. Todorovic, Neural dynamics of 1-D and 2-D brightness perception: A unified model of classical and recent phenomena, Perception & Psychophysics 43 (3) (1988) 241–277. doi:10.3758/BF03207869.

[9] L. Pessoa, E. Mingolla, H. Neumann, A Contrast- and Luminance-driven Multiscale Network Model of Brightness Perception, Vision Research 35 (15) (1995) 2201–2223. doi:10.1016/0042-6989(94)00313-0.

[10] M. S. Keil, G. Cristóbal, T. Hansen, H. Neumann, Recovering real-world images from single-scale boundaries with a novel filling-in architecture, Neural Networks 18 (10) (2005) 1319–1331. doi:10.1016/j.neunet.2005.04.003.

[11] S. Grossberg, 3-D vision and figure-ground separation by visual cortex, Perception & Psychophysics 55 (1) (1994) 48–121. doi:10.3758/BF03206880.

[12] M. S. Keil, Smooth Gradient Representations as a Unifying Account of Chevreul's Illusion, Mach Bands, and a Variant of the Ehrenstein Disk, Neural Computation 18 (4) (2006) 871–903. doi:10.1162/089976606775774705.

[13] M. S. Keil, G. Cristóbal, H. Neumann, Gradient representation and perception in the early visual system-A novel account of Mach band formation, Vision Research 46 (17) (2006) 2659–2674. doi:10.1016/j.visres.2006.01.038.

[14] M. S. Keil, Gradient representations and the perception of luminosity, Vision Research 47 (27) (2007) 3360–3372. arXiv:0709.3237, doi:10.1016/j.visres.2007.09.018.

[15] M. Sawayama, E. Kimura, Stain on texture: Perception of a dark spot having a blurred edge on textured backgrounds, Vision Research 109 (PB) (2015) 209–220. doi:10.1016/j.visres.2014.11.017.

[16] Kingdom Frederick A A, Color brings relief to human vision, Nature Neuroscience 6 (2003) 641. doi:http://dx.doi.org/10.1038/nn1060 10.1038/nn1060.

[17] A. Olmos, F. A. Kingdom, A biologically inspired algorithm for the recovery of shading and reflectance images, Perception 33 (12) (2004) 1463–1473. doi:10.1068/p5321.

[18] A. Gilchrist, Seeing Black and White, 2007. doi:10.1093/acprof:oso/9780195187168.001.0001.

[19] F. A. Kingdom, Perceiving light versus material, Vision Research 48 (20) (2008) 2090 – 2105, vision Research Reviews. doi:https://doi.org/10.1016/j.visres.2008.03.020.

[20] B. A. Olshausen, D. J. Field, Emergence of simple-cell receptive field properties by learning a sparse code for natural images, Nature 381 (6583) (1996) 607–609. doi:10.1038/381607a0.

[21] Ruderman D L., The statistics of natural images, Network: Computation in Neural systems 5 (4) (1994) 517–548.

[22] E. P. Simoncelli, B. A. Olshausen, Natural Image Statistics and Neural Representation, Annual Review of Neuroscience 24 (1) (2001) 1193–1216. doi:10.1146/annurev.neuro.24.1.1193.

[23] J. P. Jones, A. Stepnoski, L. A. Palmer, The two-dimensional spectral structure of simple receptive fields in cat striate cortex., Journal of Neurophysiology 58 (6) (1987) 1212–1232.

[24] D. L. Ringach, Spatial structure and symmetry of simple-cell receptive fields in macaque primary visual cortex., Journal of neurophysiology 88 (1) (2002) 455–63. doi:10.1152/jn.00881.2001.

[25] S. Grossberg, S. Hong, A neural model of surface perception: Lightness, anchoring, and filling-in, Spatial Vision 19 (2-4) (2006) 263–321. doi:10.1163/156856806776923399.

[26] A. D. Logvinenko, Lightness induction revisited, Perception 28 (7) (1999) 803–816. doi:10.1068/p2801.

[27] M. E. McCourt, A spatial frequency dependent grating-induction effect, Vision Research 22 (1) (1982) 119–134. doi:10.1016/0042-6989(82)90173-0.

[28] J. M. Foley, M. E. McCourt, Visual grating induction, Journal of the Optical Society of America A 2 (7) (1985) 1220. doi:10.1364/JOSAA.2.001220.

[29] A. D. Logvinenko, J. Kane, Hering's and Helmholtz's types of simultaneous lightness contrast, Journal of Vision 4 (12) (2004) 9. doi:10.1167/4.12.9.

[30] M. H. Herzog, T. Kammer, F. Scharnowski, Time slices: What is the duration of a percept?, PLOS Biology 14 (2016) e1002433.

[31] M. E. Rudd, How attention and contrast gain control interact to regulate lightness contrast and assimilation: A computational neural model, Journal of Vision 10 (14) (2010) 40. arXiv:/data/journals/jov/933538/jov-10-14-40.pdf, doi:10.1167/10.14.40.
URL + http://dx.doi.org/10.1167/10.14.40

[32] M. Rudd, Retinex-like computations in human lightness perception and their possible realization in visual cortex, Electronic Imaging 2016 (2016) 1–8.

[33] E. H. Adelson, Lightness Perception and Lightness Illusions, The New Cognitive Neurosciences 3 (2000) 339–351. doi:10.1068/p230869.

[34] B. Blakeslee, M. E. McCourt, Similar mechanisms underlie simultaneous brightness contrast and grating induction, Vision Research 37 (20) (1997) 2849–2869. doi:10.1016/S0042-6989(97)00086-2.

[35] B. Blakeslee, M. E. McCourt, A multiscale spatial filtering account of the White effect, simultaneous brightness contrast and grating induction, Vision Research 39 (26) (1999) 4361–4377. doi:10.1016/S0042-6989(99)00119-4.

[36] W. D. Ross, L. Pessoa, Lightness from contrast: A selective integration model, Perception & Psychophysics 62 (6) (2000) 1160–1181. doi:10.3758/BF03212120.

[37] H. G. Barrow, J. M. Tenenbaum, Recovering intrinsic scene characteristics from images (1978) 3–26.

[38] Y. Weiss, Deriving intrinsic images from image sequences, Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001 2 (2001) 68–75. doi:10.1109/ICCV.2001.937606.

[39] P. V. Gehler, C. Rother, M. Kiefel, L. Zhang, B. Schölkopf, Recovering intrinsic images with a global sparsity prior on reflectance, Advances in Neural Information Processing Systems 24 (NIPS 2011) (2011).

[40] M. F. Tappen, W. T. Freeman, E. H. Adelson, Recovering intrinsic images from a single image, IEEE Transactions on Pattern Analysis and Machine Intelligence 27 (9) (2005) 1459–1472. doi:10.1109/TPAMI.2005.185.

[41] L. Shen, C. Yeo, B.-S. Hua, Intrinsic Image Decomposition Using a Sparse Representation of Reflectance, IEEE Transactions on Pattern Analysis and Machine Intelligence 35 (12) (2013) 2904–2915. doi:10.1109/TPAMI.2013.136.

[42] H. J. Gerrits, A. J. Vendrik, Simultaneous contrast, filling-in process and information processing in man's visual system, Exp Brain Res. 1970 Nov 26;11(4):411-30. (1970). doi:10.1007/BF00237914.

[43] A. F. Rossi, C. D. Rittenhouse, M. A. Paradiso, The Representation of Brightness in Primary Visual Cortex, Science 273 (5278) (1996) 1104–1107. doi:10.1126/science.273.5278.1104.

[44] H. Komatsu, The neural mechanisms of perceptual filling-in, Nature Reviews Neuroscience 7 (3) (2006) 220–231. doi:10.1038/nrn1869.

[45] S. Grossberg, E. Mingolla, Neural dynamics of surface perception: Boundary webs, illuminants, and shape-from-shading, Computer Vision, Graphics and Image Processing 37 (1) (1987) 116–165. doi:10.1016/S0734-189X(87)80015-4.

[46] A. Galmonte, A. Soranzo, M. E. Rudd, T. Agostini, The phantom illusion, Vision Research 117 (2015) 49 – 58. doi:https://doi.org/10.1016/j.visres.2015.10.007.

[47] Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli, Image quality assessment: From error visibility to structural similarity, IEEE Transactions on Image Processing 13 (4) (2004) 600–612. arXiv:9411012, doi:10.1109/TIP.2003.819861.

[48] B. Blakeslee, M. E. McCourt, Brightness induction magnitude declines with increasing distance from the inducing field edge, Vision Research 78 (2013) 39 – 45. doi:https://doi.org/10.1016/j.visres.2012.12.007.
URL http://www.sciencedirect.com/science/article/pii/S004269891200404X

Chapter 3

Second Publication: "Dynamic Decorrelation as a unifying principle for explaining a broad range of brightness phenomena"

Chapter abstract: This chapter presents a published study describing a computational model compatible with the early visual system that implements dynamic filtering to predict different kinds of brightness/lightness phenomena.

PLOS COMPUTATIONAL BIOLOGY

RESEARCH ARTICLE

Dynamic decorrelation as a unifying principle for explaining a broad range of brightness phenomena

Alejandro Lerer[1☯], **Hans Supèr**[1,2,3,4], **Matthias S. Keil**[1,2☯*]

1 Departament de Cognició, Desenvolupament i Psicologia de l'Educació, Faculty of Psychology, University of Barcelona, Barcelona, Spain, **2** Institut de Neurociències, Universitat de Barcelona, Barcelona, Spain, **3** Institut de Recerca Pediàtrica Hospital Sant Joan de Déu, Barcelona, Spain, **4** Catalan Institute for Advanced Studies (ICREA), Barcelona, Spain

☯ These authors contributed equally to this work.
* matskeil@ub.edu

Citation: Lerer A, Supèr H, Keil MS (2021) Dynamic decorrelation as a unifying principle for explaining a broad range of brightness phenomena. PLoS Comput Biol 17(4): e1007907. https://doi.org/10.1371/journal.pcbi.1007907

Editor: Michael Rudd, University of Nevada, Reno, UNITED STATES

Received: April 18, 2020

Accepted: April 6, 2021

Published: April 26, 2021

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: https://doi.org/10.1371/journal.pcbi.1007907

Data Availability Statement: All relevant data are within the manuscript and its Supporting Information files.

Funding: AL was supported by the Ministry of Economy and Competitiveness (Spain) PSI2010-

Abstract

The visual system is highly sensitive to spatial context for encoding luminance patterns. Context sensitivity inspired the proposal of many neural mechanisms for explaining the perception of luminance (brightness). Here we propose a novel computational model for estimating the brightness of many visual illusions. We hypothesize that many aspects of brightness can be explained by a dynamic filtering process that reduces the redundancy in edge representations on the one hand, while non-redundant activity is enhanced on the other. The dynamic filter is learned for each input image and implements context sensitivity. Dynamic filtering is applied to the responses of (model) complex cells in order to build a gain control map. The gain control map then acts on simple cell responses before they are used to create a brightness map via activity propagation. Our approach is successful in predicting many challenging visual illusions, including contrast effects, assimilation, and reverse contrast with the same set of model parameters.

Author summary

We hardly notice that what we see is often different from the physical world "outside" of the brain. This means that the visual experience that the brain actively constructs may be different from the actual physical properties of objects in the world. In this work, we propose a hypothesis about how the visual system of the brain may construct a representation for achromatic images. Since this process is not unambiguous, sometimes we notice "errors" in our perception, which cause visual illusions. The challenge for theorists, therefore, is to propose computational principles that recreate a large number of visual illusions and to explain why they occur. Notably, our proposed mechanism explains a broader set of visual illusions than any previously published proposal. We achieved this by trying to suppress predictable information. For example, if an image contains repetitive structures, then these structures are predictable and will be suppressed. In this way, non-predictable structures stand out. Corresponding mechanisms act as early as in the retina (which

18139-P, HS was supported by Ministry of Science and Innovation (Spain) PGC2018-096074-B-100, MSK was supported by Spanish Government Grant PGC2018-099506-B-I00. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

enhances luminance changes but suppresses uniform regions of luminance), and our computational model suggests that such mechanisms also might be used at subsequent stages in the visual system, where representations of perceived luminance (=brightness) are created.

Introduction

Visual perception is relative rather than absolute; the visual system (VS) computes the perceptual attributes of a visual target not only based on its physical properties, but also by considering information from the surrounding region of the target (context). For example, it is possible to induce different kinds of effects by context modification, such that the brightness of a target is contrasted (increasing brightness differences) or assimilated (decreasing brightness differences) with respect to its adjacent surround (e.g. [1]). Variants of these effects give rise to a myriad of visual illusions, which are of great utility for building hypothesis about computational mechanisms or perceptual rules for brightness perception.

At first sight it seems that contrast effects, such as simultaneous brightness contrast (SBC; Fig 1A), can be explained by lateral inhibition between a target (center) and its context (surround). However, activity related to brightness contrast does possibly not occur before V1, albeit the receptive fields of retinal ganglion cells are consistent with lateral inhibition [2].

Unlike brightness contrast effects, brightness assimilation (e.g. Fig 1C and 1D) pulls a target's brightness towards to that of its immediate context, and therefore cannot be explained by mechanisms based on plain lateral inhibition. In fact, the neural mechanisms involved in generating brightness (perceived luminance) and lightness (perceived surface reflectance), respectively, appear to be more intricate (e.g., [3–7]). Intrinsic image approaches are (computational) models which aim at deriving further images based on the characteristics of the *depicted visual scene* of an input image [8]. With respect to lightness computations, corresponding algorithms estimate a (sometimes chromatic) reflectance image and a shading image from an input image (e.g. [9–11]). Further approaches with respect to intrinsic images in a wider sense address surface qualities such as transparency (e.g., [12]) or gloss (e.g., [13]).

The majority of models for computing brightness estimates decompose an input image typically with filters of different orientations and multiple spatial frequencies (image-based decompositions, [8]). In this way, a set of images (=filter response maps) is derived, from which the input can be recovered. The recovered image is considered as brightness prediction. For instance, feature-based approaches classify filter responses into lines and edges (e.g. [14–18]), and build a brightness prediction based on recognized features.

It is particularly challenging for computational proposals to simultaneously explain contrast and assimilation effects with identical model parameters. Blakeslee and McCourt, [19, 20], proposed a highly successful image decomposition with oriented difference of Gaussians (ODOG) filters. In the ODOG-model, a brightness prediction is generated by two steps. First, a weighted sum of filter responses across spatial frequencies is computed for each orientation (orientation channels). Second, each orientation channel is divided by its root mean square level before they are summed to yield the final brightness map. Although the ODOG model predicts SBC and many assimilation displays, in our re-implementation it fails at the Benary Cross (Fig 1D) and Chevreul's Illusion.

Several variants of the multiscale-decomposition approach have been proposed. Dakin and Bex, [21], showed that isotropic filters are sufficient to reproduce the Craik-O'Brien Cornsweet effect and White's effect. Otazu and co-workers, [22], used an invertible wavelet

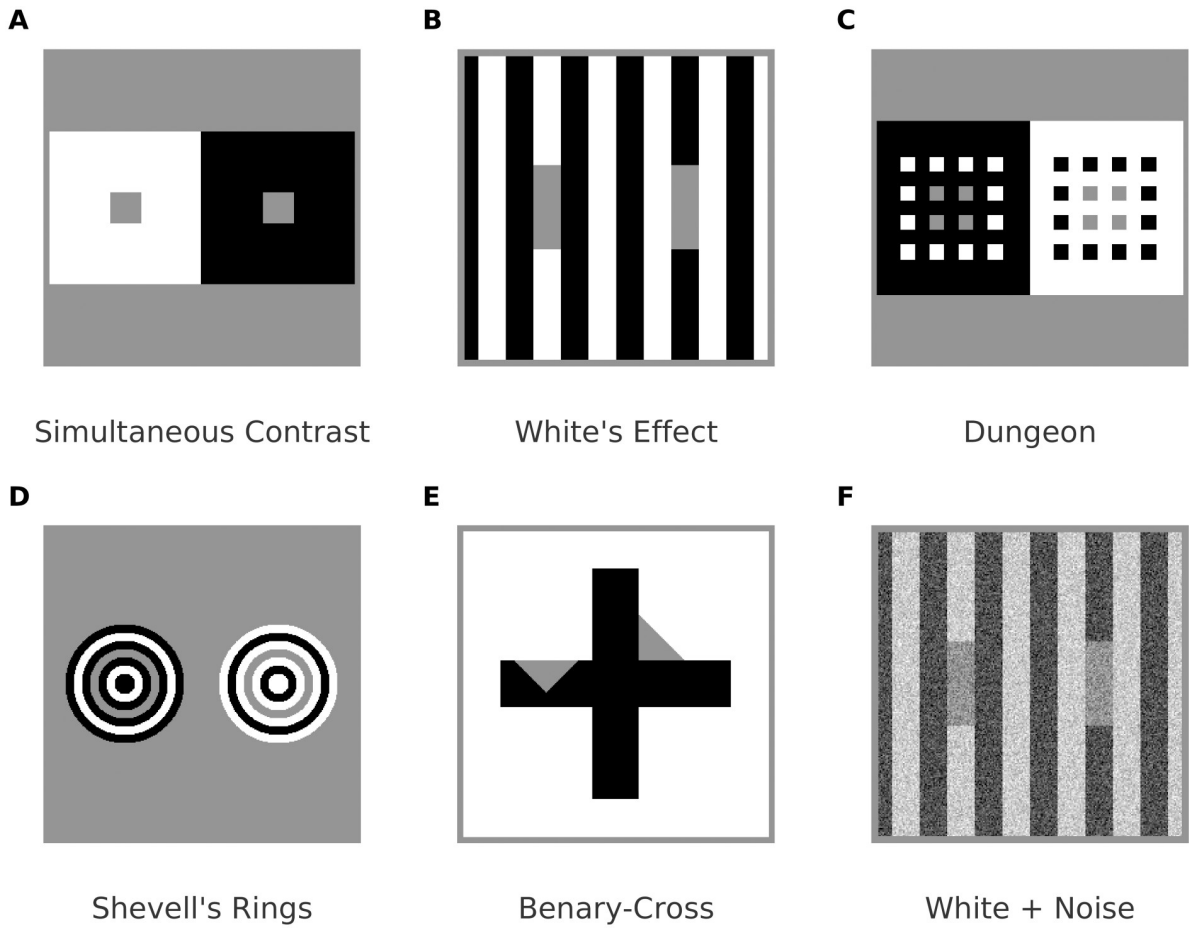


Fig 1. (A) Simultaneous brightness contrast: The two gray patches with identical luminance increase their brightness difference with their respective backgrounds. (B) White's effect is consistent with contrast at the horizontal contours of the bars, and with assimilation along the vertical contours. (C, D) Examples of brightness assimilation: The gray structures with identical luminance decrease their brightness difference with their respective background. (E) Sensitivity to context: The gray triangles have identical luminance and selectively contrast with the cross (left) or the background (right). (F) White's effect is still intact in the presence of noise.

https://doi.org/10.1371/journal.pcbi.1007907.g001

transformation with wavelet coefficients being weighted according to the contrast-sensitivity function [23]. The weights are further adjusted dependent on local surround contrast. However, models based on adjusting the spatial frequency distribution of filter responses appear to fall short of predicting brightness when the input images are contaminated by (band-pass filtered) noise [24], (Fig 1F).

Computationally, the spatial operation of certain types of retinal ganglion cells can be dumbed down to taking the second spatial derivative of the visual input. In this way, sharp contrasts (edges or boundaries) are enhanced. Edge integration models [1, 25–27] try to invert spatial differentiation (=contrast extraction) by thresholding (suppress shallow gradients due to illumination effects), and integration (across edges in order to estimate lightness). This means that edge integration models—unlike multi-scale approaches—propose explicit mechanisms for suppressing illumination effects, and for estimating a reflectance image. Filling-in (FI) models [28, 29] can be considered as a neural implementation of edge integration, where boundary activity is propagated laterally in order to generate perceptible surface properties.

Whether or not activity propagation plays a role in surface perception is subject of a yet ongoing debate [6, 30, 31]; also, where it might occur in the visual system (e.g. [32, 33]). It is often argued that with Grating Induction, FI-mechanisms would incorrectly predict a homogeneous test-field, because activity would "average out". This argument, however, ignores that boundary webs could form across the test field [34], and/or darkness and brightness channels could interact [35, 36]. Furthermore, Grating Induction seems to be perceived instantaneously, what seems incompatible with activity propagation at the neuronal level [6, 37]. However, this argument ignores that the activity to be propagated can be initialized at a coarser scale than the boundaries that contain it. Furthermore, FI may proceed simultaneously in layers with different resolutions [38]. This would render the perceptual effect indistinguishable from multi-scale filtering. Finally, it has been argued the addition of noise would disrupt activity propagation [6, 21]. This argument ignores, however, that texture-like features such as noise (even symmetric simple cells with small receptive fields) and surface boundaries (odd symmetric features) may be represented by different layers [36]. Furthermore, even with a single representation, noise would likely not form closed domains in order to contain activity (e.g. Fig 7 in [35, 39]), or would not fill in at all if the initial activity is computed at a coarser scale.

Typical FI models distinguish two types of contours [40]. The first type represent barriers for activity propagation, and defines the boundary contour system (BCS). It represents the complete 3-D boundary structure of a visual scene, including boundaries from texture and depth. The second type is the feature contour system (FCS) and represents surface properties to be filled-in, such as brightness, lightness, color and depth. FCS processing was suggested to occur in cytochrome-oxidase staining regions in V1 (blobs) and thin stripes in V2. Interaction between FCS and BCS was hypothesized to occur in V4 [41]. The hypothetical FCS/BCS dichotomy is compatible with several experimental findings [42–46].

Whereas the prediction of brightness contrast usually is straightforward with (most) filling-in architectures, assimilation effects remain challenging. On the basis of a one-dimensional luminance profile, Grossberg and Todorovik (in [40]) explained how two (non-)adjacent (luminance) regions could influence each other. If boundaries are sufficiently near in the BCS, then their activity is reduced. Therefore, activity propagation in the FCS may not completely be blocked, causing FCS activity to fill into (non-)adjacent surfaces. In this way, one surface may not just be influenced by the brightness of its immediate surrounding region, but even from further away.

Domijan, [47], extended these ideas to two dimensions. He computed luminance-modulated FCS activity with an unbalanced center-surround kernel, similar to [48], but see also [35] for a different way of luminance encoding. BCS activity is computed by first deriving a local boundary map, where the loss of activity at junctions and corners was corrected. Based on the local boundary map, a global boundary map was computed. In the latter, contours which are parallel or co-linear to another contour were enhanced. Finally, local boundary activity was divided, at each position, by global boundary activity. The division is approximately one at those positions where no contour enhancement took place in the global boundary map (otherwise it is smaller). The final BCS output keeps only those activities that are relatively close to one—boundaries with smaller activity which are parallel to high contrast edges are eliminated. This causes FCS activity to freely diffuse across the eliminated boundaries. In this way, Domijan was able to predict 2-D assimilation displays with a FI-architecture.

Ross and Pessoa, [49], modified FCS activity before filling-in by using an occlusion-sensitive copy of the BCS. The modification of the original boundaries is based on T-junctions: Boundaries along the stem of the "T" are suppressed, while the others are enhanced. The modified boundary map ("context boundaries") is subsequently used for suppressing contrast measurements in the FCS. The original boundaries act as diffusion barriers in the FCS. Although

the model successfully predicts White's effect and the Benary Cross (Fig 1B and 1E), psychophysical evidence suggests that White's effect seems not to be affected significantly if the T-junctions are suppressed [50, 51] (cf. Fig 1D), nor seem to be other illusions [52]. Furthermore, it is not readily clear whether junction rules do represent reliable cues in complex natural scenes: The utility of junctions rules has only been illustrated with relatively simple artificial displays [49, 53].

Barkan [54] used center-surround receptive fields at four resolution levels for edge (or contrast) extraction. At each resolution level, filter response amplitudes ("local contrast") were gain-controlled with a low-pass filtered version of themselves ("remote contrast"). A brightness map was estimated from the gain-controlled contrast map with fixpoint iteration of a Laplacian [55], which implements the filling-in process. The model was successful in simulating assimilation and reverse assimilation effects (mostly centered on challenging variants of White's effect), but failed in predicting Simultaneous Brightness Contrast (SBC).

A completely different approach for explaining visual illusions is based on a statistical analysis of real-world images. This approach suggests that the perception of brightness [56, 57] or lightness [57–59] is related to knowledge about the statistical relationships between visual patterns across space. In particular, [56] proposed that the brightness of a visual target embedded in some context depends on the expected luminance according to a probability distribution function. The probability distribution function integrates all contexts in that what which the target was seen previously. The perception of the target then depends on its expected luminance given its current context: It is perceived as brighter if the expected luminance is lower, and it is perceived as darker otherwise. This approach is successful in predicting contrast and assimilation for several visual illusions, and suggests a statistical relationship between luminance patterns and brightness perception. Unfortunately, no attempt has been made in order to unveil any information processing strategy from the statistical analysis (but see [59]).

As with some of the models reviewed above, our approach also emphasizes the importance of boundaries in brightness perception: We propose to reduce redundancy in the boundary maps. Such encoding strategies usually reduce the overall activity of a representation and thus the expenditure of metabolic energy, [60, 61], and are also known as efficient coding [62], predictive coding [63], whitening [64] or response equalization [65].

With respect to mid- or higher-level processing, White [66] suggested that a pattern-specific inhibition mechanism acts in the visual cortex, which inhibits regularly arranged patterns of a visual stimulus. Our model is related to White's idea: We adjust a boundary map, such that redundant activity is suppressed, while non-redundant activity is enhanced. Since neurons that encode redundant patterns tend to be over-represented, the overall boundary activity is reduced after the adjustment (response equalization). Response equalization is carried out by a dynamic filter.

Fig 2 shows an overview of our model. In the first step an input image is encoded by two sets of Gabor filters, which mimic the spatial response properties of simple cells in V1 [67]. The responses of the high-resolution filters define the Contrast-only channel (similar to the BCS), while responses of the more coarse-grained filters define the Contrast-Luminance channel (similar to FCS). Here the term 'channel' refers to a feature of our model. We do not imply two pathways for encoding luminance and contrast in the visual cortex. From the Contrast-only channel, we compute boundary activity via local energy [16, 17]. Local energy is insensitive to the phase information, and thus resembles complex cell responses. From the local energy map, a decorrelation kernel is learned, and then applied to it, in order to reduce redundancy (=dynamic filtering). The redundancy-reduced energy map then functions as a gain control map for both contrast channels. As a consequence, contrast activity is modified. Subsequently, an iterative procedure is used to recover a brightness map from the two contrast

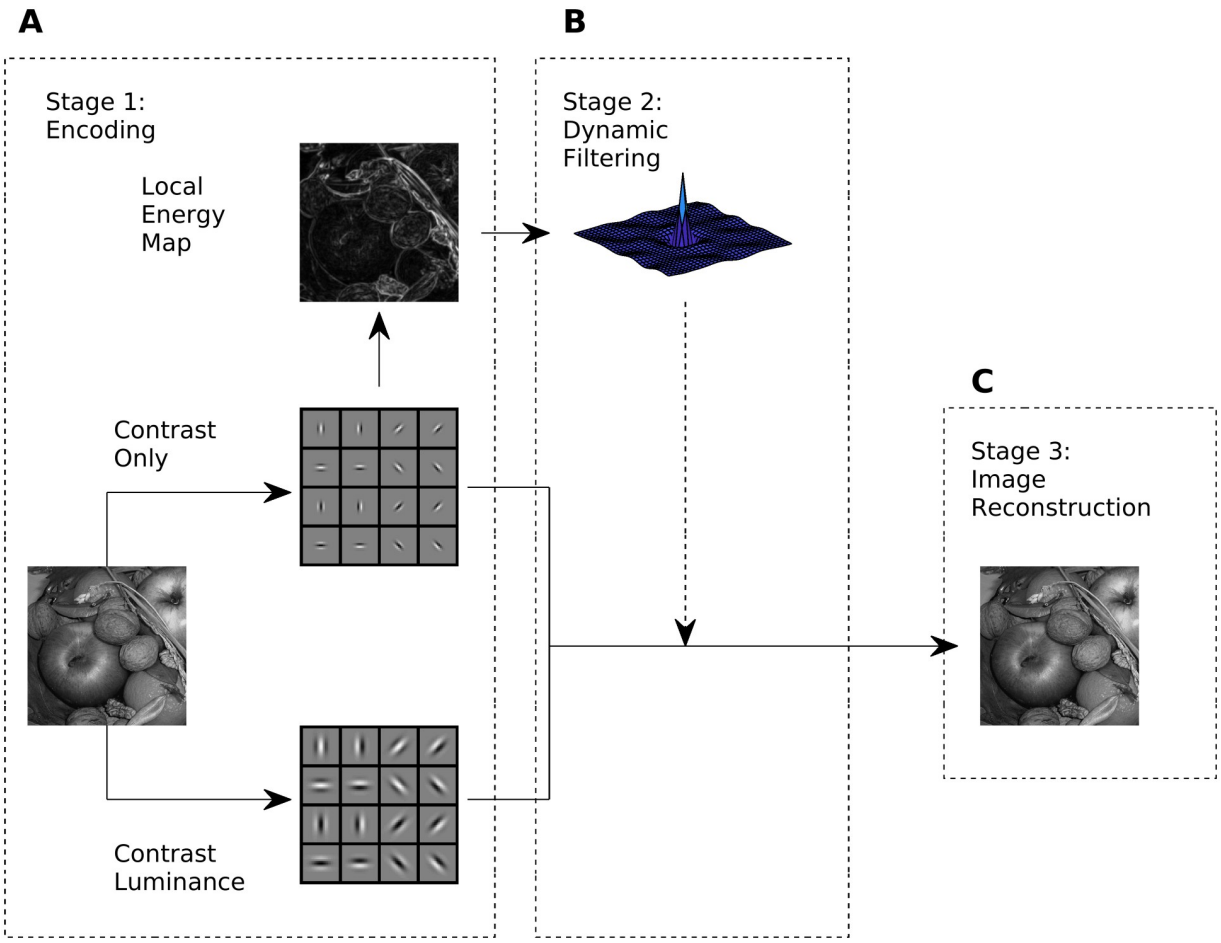


Fig 2. Model overview. Each of the three stages is mathematically specified in the Methods Section. (A) Stage 1: The Contrast-only channel and Contrast-Luminance channel are instantiated by filtering an input image with a corresponding set of Gabor filters with high spatial resolution (0.25 cycles per pixel) and coarse resolution (0.125 cycles per pixels), respectively. The local energy map is computed from the Contrast-only channel. (B) Stage 2: The kernel of the dynamic filter is estimated from the local energy map. Dynamic filtering equalizes the amplitude spectrum of the energy map, reducing redundancy. The decorrelated energy map serves as gain control for both contrast channels. (C) Stage 3: The output of the model is a brightness map that is obtained by solving an inverse problem, that is recovering the image from both contrast channels. Note that the two contrast channels do not interact with each other before Stage 3.

https://doi.org/10.1371/journal.pcbi.1007907.g002

channels. Our iterative procedure resembles a filling-in process. The plausibility of our model is underlined by predicting many challenging visual illusion, including some that were never predicted by any other computational model so far.

Materials and methods

Fig 2 depicts the three stages of our model. Stage 1 encodes the input image into a Contrast-only and a Contrast-Luminance channel by two respective set of Gabor filters. Stage 2 the dynamic filtering process. It consists of equalizing response amplitudes of the local energy map and then gain-controlling both channels. Finally, Stage 3 refers to the filling-in process for estimating a brightness map. The brightness map represents the output of our model. The three stages are detailed in the following subsections.

Stage 1: Encoding

Contrast-only and contrast-luminance channel. We use Gabor filters for encoding Contrast-only and Contrast-Luminance information. In the primary visual cortex, simple cells respond to oriented light-dark bars across a certain spatial frequency range [68] and their receptive fields can be modeled by Gabor filters [67, 69, 70]. Consistent with the properties of Gabor filters, it seems that many simple cells in V1 encode contrast information. Under certain circumstances though, neurons in V1 may respond to surface brightness as well, even without (sharp luminance-)contrasts in their receptive fields [71, 72]. For example, [73] found such neurons in V1 which have large receptive fields, broad orientation tuning, and a preference for low spatial frequencies. These neurons respond to both contrast and luminance.

In our model, the set of Gabor filters for the Contrast-only channel had a spatial frequency of 0.25 cycles/pixel and balanced ON-OFF sub-regions (i.e., the sum across the kernel is zero) In this way they did not respond to homogeneous regions of the input (DC-free). For the Contrast-Luminance channel we used Gabor filters with a lower spatial frequency (0.125 cycles/pixel), and unbalanced ON-OFF sub-regions (i.e., the sum across the kernel is positive) such that they respond to both luminance and contrast (non-zero DC part). Strictly speaking, these filters are actually no longer pure bandpass filters. But since the sum across the kernel is very small, the bandpass properties dominate the response. The use of bigger kernels which are selective for even lower spatial frequencies would not alter significantly our results. However, the computational cost would increase. For this reason, we chose the kernels of the Contrast-Luminance channel with half the spatial frequency of the Contrast-only channel. Fig 2A illustrates the two filter sets. Parameter values and a mathematical description of unbalancing the ON/OFF subregions is provided in section A in S1 Text.

The responses of the Contrast-only and the Contrast-Luminance channel were computed by convolving (symbol "*") a luminance image with the corresponding set of Gabor filters. That is, if g represents a Gabor kernel (either from the Contrast-only or the Contrast-Luminance channel), then R_g represents its activity in response to the input image as:

$$R_g(x, y) = g * Im(x, y) \tag{1}$$

The arguments (x,y) denote 2D spatial coordinates. The contrast channels remain separated until the filling-in process. Fig 3A shows examples of filter responses. Notice that contrast responses dominate the response map of the Contrast-Luminance channel in Fig 3A. Fig 3C shows filter responses of the two channels to a luminance step (contrast) and a uniform region: The DC-response of the Contrast-Luminance filter is small compared with its contrast response.

Local energy map. The local energy map resembles the properties of complex cells in the primary visual cortex [74]. Complex cell responses are similar to those of simple cells in terms of orientation and spatial frequency preference, but tend to be non-linear and shift-invariant with respect to contrast phase [74, 75]. Local energy can be calculated from a pair of Gabor filters in quadrature phase by summing their squared responses and then taking the square root [16, 17, 76, 77]. A quadrature pair are two Gabor filters with 90 degree phase difference, but with identical preference in orientation and spatial frequency. Finally, the local energy map E was computed through averaging the activity of our model complex cells across all orientations. Fig 2A shows an example local energy map. It essentially corresponds to the contours of the input image. Mathematical details are given in section B in S1 Text.

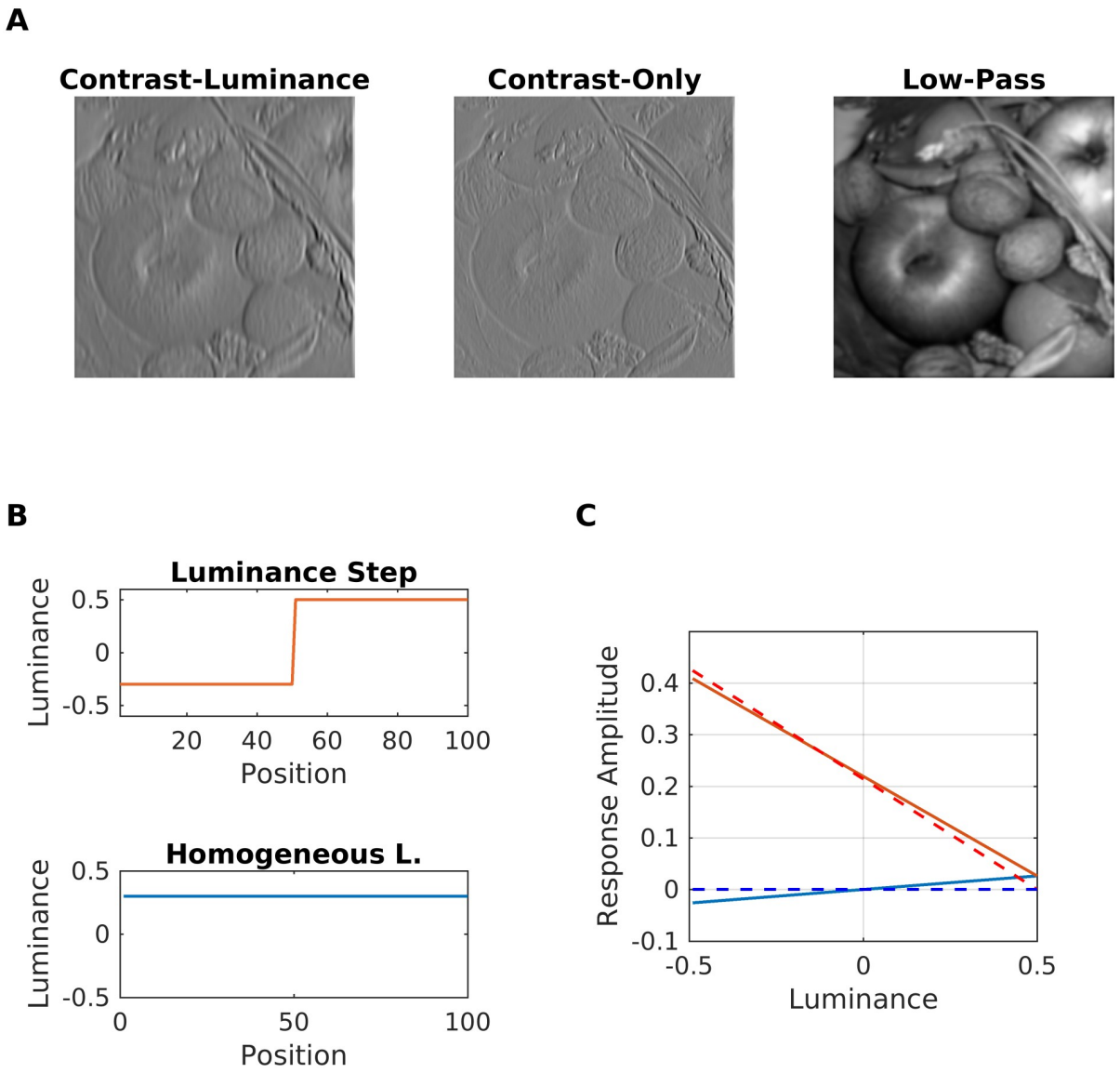


Fig 3. Sensitivity to contrast and luminance. (A) (left) Fruits filtered by a Contrast-Luminance filter g (Eq 1). (Middle) Fruits filtered by a Contrast-only filter g with the same orientation (Eq 1). (Right) Fruits filtered by a low-pass filter. (B) Example luminance step (values -0.3 to 0.5) and an example for a homogeneous region with luminance 0.3. (C) Responses (maximum) of Contrast-Luminance filters (solid lines) and Contrast-only filters (dashed lines) to the homogeneous region (red color) and the luminance step (blue color). For the luminance step, the lower luminance was increased from -0.5 to 0.5. Negative filter responses result from the use of negative input values.

https://doi.org/10.1371/journal.pcbi.1007907.g003

Stage 2: Decorrelation

In this section we describe stage 2 of our model (Fig 2B). The first subsection describes how the dynamic filter is computed with zero-phase component analysis (ZCA: [78]). ZCA is a decorrelation method that whitens the covariances of the original data while preserving their original direction [79]. With the dynamic filter we equalize the amplitude spectrum of the energy map. In fact, it produces very similar results to the "Whitening-by-Diffusion" method proposed in [65]. In the second and third subsections we detail the computation of the gain control map and how it interacts with the whitened energy map, respectively.

Dynamic filter. The purpose of the dynamic filter is to equalize the amplitude spectrum of the energy map. It is computed with zero-phase whitening (ZCA), a technique which has been used for learning the receptive fields of retinal ganglion cells [78]. ZCA resembles principal component analysis (PCA), and signal decorrelation can be achieved with both of the latter. However, the components are constrained to be symmetrical with ZCA. This "symmetry constraint" guarantees that the principal components are localized in the spatial domain [78], and therefore can be used as filter kernels. We nevertheless introduced a couple of modifications to the original ZCA (see section C in S1 Text). As a result of the modifications, we obtained a spatial filter that adapts to the spatial structure of the local energy map of an image. It is called "dynamical" because a different filter is learned from each image. After filtering, the amplitude spectrum of the energy map is more uniform (see Fig 4). By the Wiener–Khinchin theorem, a more uniform power or amplitude spectrum implies that the original signal is more decorrelated [80, 81]. For the decorrelated energy map, this means that spatial patterns with low redundancy tend to be intensified, while patterns with high redundancy tend to be attenuated. This is illustrated with Fig 4, where after filtering, horizontal edges are intensified in the energy map as compared to vertical ones.

Gain control map. The gain control map G is computed in two steps. First, the dynamic filter F is used as a convolution kernel for the energy map E:

$$\tilde{E}(x,y) = \begin{cases} F * (E(x,y) - \mathrm{mean}(E)) & if\ \ E(x,y) \geq \omega \\ 0 & if\ \ E(x,y) < \omega \end{cases} \quad (2)$$

Here, the symbol "*" indicates convolution. We set the threshold to 10 percent of the maximum activity as $\omega = 0.1\max(E)$. We observed that without thresholding, artifacts and noise tend to accumulate in the brightness prediction. In addition, thresholding increases sharpness. The exact percentage value is not critical. Our results would not change significantly when using, for example, 0.15, 0.2, or 0.3 times the maximum. In the second step we normalized the activity of the gain control map with a sigmoid function $S(x, a, b) = 1/(1 + e^{-ax-b})$ as

$$G(x,y) = \begin{cases} 2\left[S\left(\frac{\tilde{E}(x,y)}{max|\tilde{E}|}, a, b\right)\right] - 1 & if\ \ \tilde{E}(x,y) \neq 0 \\ 0 & if \tilde{E}(x,y) = 0 \end{cases} \quad (3)$$

The parameters were fixed as $a = 5$ and $b = \min(3\mathrm{mean}(\tilde{E}), 0.3)$. Notice that the gain control map G is normalized to [−1, 1]. The sigmoid's slope a controls the intensity and smoothness of G. For bigger values of a the soft-thresholding turns into hard thresholding. Even with hard thresholding, the results will be not significantly affected. For small values ($a < 1$), the Gain control map was practically useless due to the small values of G. The inflection point b (if $a \gg 1$ it would be the threshold) is, however, important for estimating brightness, as it determines when the values of G change from negative to positive: If $G(x, y)>0$, then the brightness contrast at (x, y) is increased, while contrast will be reduced (=assimilation) if G is negative. We found b by manual optimization.

Feedback interaction and channel gain control. The Contrast-only and the Contrast-Luminance channel were subjected to gain control using the decorrelated energy map as

$$R_g^*(x,y) = R_g(x,y)\left[\frac{\tau + \tau G(x,y)}{\tau + |R_g(x,y)|G(x,y)}\right] \quad (4)$$

Here, G is the gain control map, τ is a control parameter, R_g represents the activity of a Gabor filter g of the corresponding filter set (i.e., Contrast-only or Contrast-Luminance), (x,y)

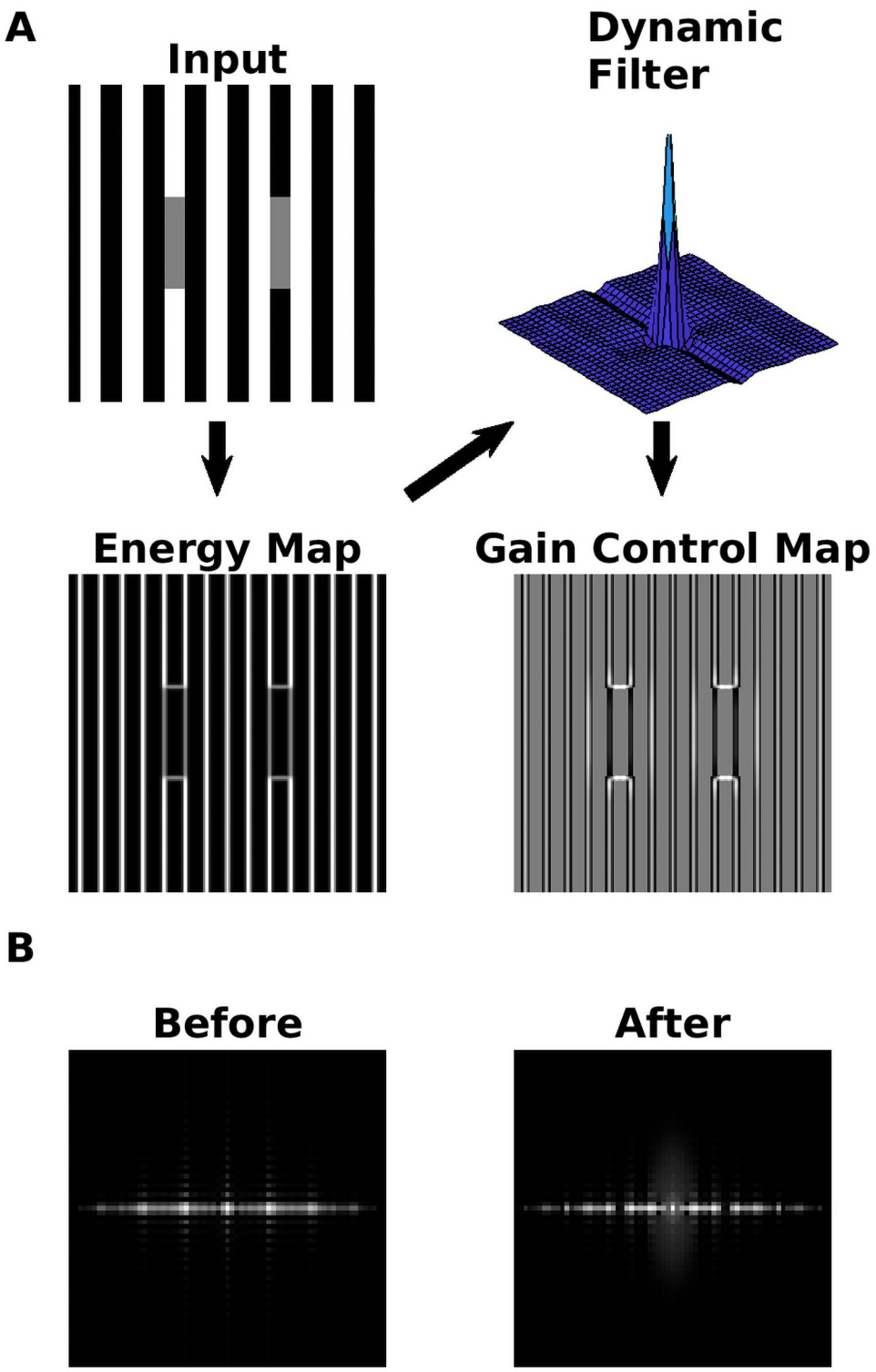


Fig 4. Action of the dynamic filter. (A) The arrows indicate the steps in order to obtain the gain control map: (i) A local energy map is computed; (ii) a dynamical filter is constructed with a customized zero-phase whitening procedure (ZCA, see section C in S1 Text); (iii) a gain control map is obtained by filtering the energy map with the dynamical filter (see subsection Gain Control Map). (B) The power spectrum (= square of amplitude spectrum) before and after of applying the dynamic filter on the local energy map.

https://doi.org/10.1371/journal.pcbi.1007907.g004

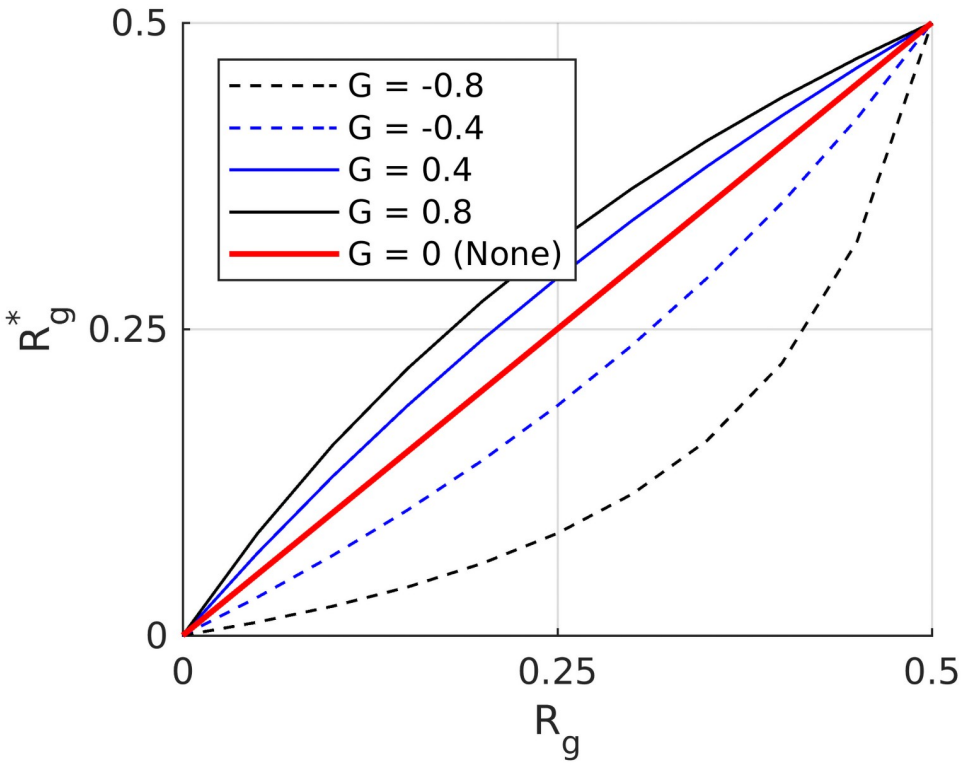


Fig 5. Filter responses before and after applying Eq 4. Each curve represents a different scalar value for the gain control G (notice that in Eq 4, G is two-dimensional). The values of G corresponding to each curve are indicated in the figure legend.

https://doi.org/10.1371/journal.pcbi.1007907.g005

denote pixel coordinates, and τ is a control parameter. The control parameter τ acts as an upper bound for the maximum activity that any filter can reach when it encodes a luminance grating (min=-0,5, max = 0.5) that matches the spatial frequency of the filter.

Fig 5 illustrates the behavior of Eq 4 for different values of the gain control map G, where G was set to the (scalar) values indicated in the figure legend. Gabor filter responses are amplified for $G > 0$. For $G < 0$, they are attenuated. The gain control map G is furthermore modulated by the denominator of Eq 4. The modulation is weak or absent when Gabor filter responses R_g are equal to τ or 0, (Fig 5), while it is strong between the latter values. The modulation is crucial for explaining Chevreul's illusion and Mach Bands, but it is not relevant for all other brightness displays. Finally, after applying Eq 4 to each channel, the results were lowpass filtered with a Gaussian kernel (standard deviation 1 pixel) in order to reduce possible artifacts.

Stage 3: Brightness estimation as a filling in process

The brightness map $\hat{z}$ is the output of our model. It is estimated by minimizing an objective function $E(z)$, which optimizes the trade-off between the reconstruction error (first term in the sum of Eq 5) using the gain-controlled contrast channels R^*_g and a smoothness constraint (second term):

$$\hat{z} = \text{argmin}_z E(z) := \text{argmin}_z \sum_g \| R^*_g - g * z \|^2 + \mu \| \nabla^2 z \|^2 \quad (5)$$

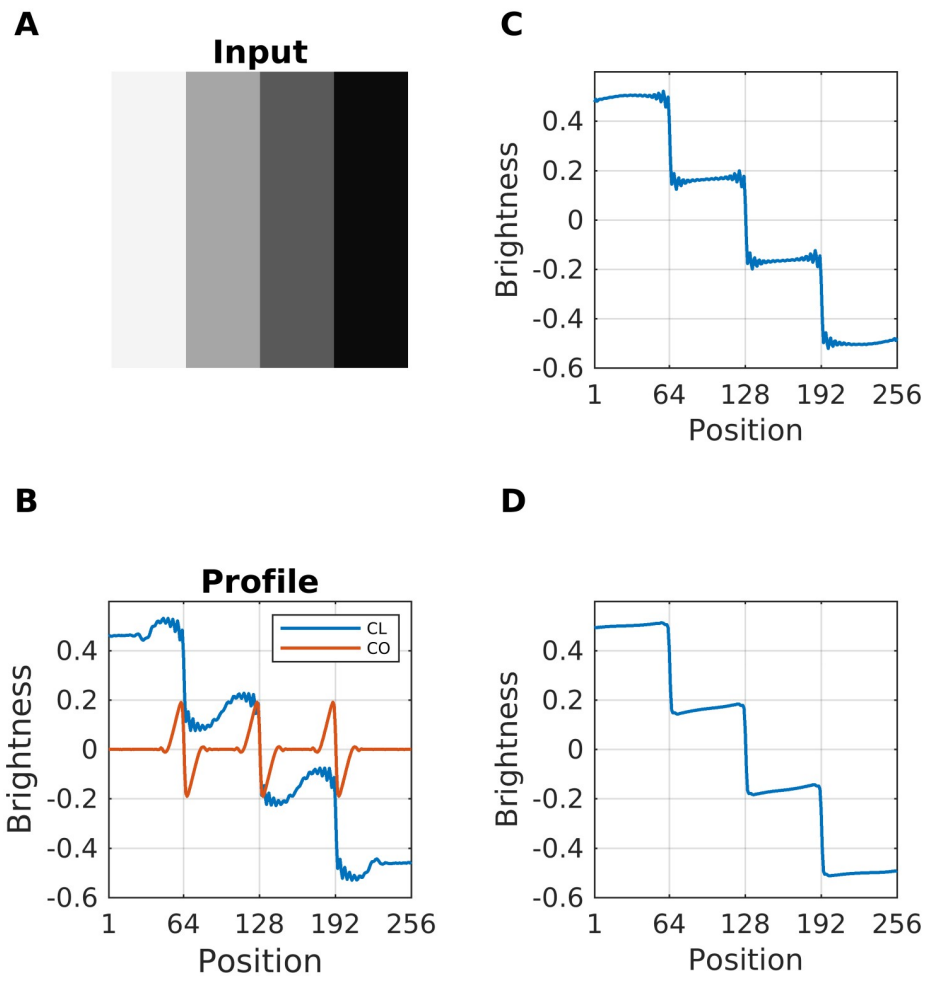


Fig 6. Contribution of each channel to brightness estimation. (A) A luminance staircase (giving rise to Chevreul's Illusion) served as input. (B) The resulting brightness profiles are estimated at 10 iterations. The blue curve (legend label "CL") uses only the Contrast-Luminance channel in Eq 5. The red curve ("CO") uses only the Contrast-Only channel without responses to luminance. (C) Resulting brightness profile at the stop criterion for the Contrast-Luminance channel. Aliasing artifacts appear due to undersampling. (D) Brightness estimation with both channels. The Contrast-only channel eliminates undersampling.

https://doi.org/10.1371/journal.pcbi.1007907.g006

Notice that the sum involves the Gabor filters g of both contrast channels (i.e., Contrast-only and Contrast-Luminance). The regularization parameter controls the smoothness constraint and is set to 0.01. The Laplacian is denoted by ∇^2. The smoothness term and the Contrast-only channel serve to reduce artifacts produced at discontinuities (Fig 6). The equation is solved iteratively with the conjugate gradient method. The method starts with an image z_0 of random values. The weights of z_k are updated such that the value of the cost function $E(z_k)$ decreases with each iteration step. This gradient descent continues until a maximum of 100 iterations is reached, or until an error criterion is satisfied (see section D in S1 Text for more details). The filter responses R_g determine how the weights z_k are updated. Because the R_g are higher at the edges, activity is iteratively propagated from the edges. Direction and orientation of activity spreading are determined by the odd-symmetrical Gabor filters that are used for encoding the image. Thus, image reconstruction (i.e., brightness estimation) proceeds according to a filling-in process.

It is instructive to highlight the role of both channels for estimating brightness. The bulk of activity propagation depends on the Contrast-Luminance channel. Since it has bigger filter kernels than the Contrast-only channel, it propagates activity across greater distances per unit time. The time to convergence is further reduced by the residual luminance responses of the kernels in the Contrast-Luminance channel. Fig 6 compares the reconstruction of a luminance staircase at 10 iterations based on a single channel. Although both channels are used in Eq 5, the major contribution to brightness estimation comes from the Contrast-Luminance channel, while the Contrast-only channel serves to encode edge-information. As a consequence, the Contrast-only channel could be excluded from the reconstruction without significantly affecting brightness predictions. Nevertheless, the participation of the high-frequency filters (Contrast-only channel) eliminates undersampling (or sub-sampling), which would cause an accumulation of oscillatory artifacts (due to aliasing) close to the edges (see Fig 6C versus Fig 6D).

Classification of model predictions according to three scenarios

The effect of the Gain Control Map on estimated brightness (=model output) is as follows. If a Gabor response amplitude after gain control has increased (decreased), it would produce a major (minor) contrast in the reconstructed image (=estimated brightness). This means that brightness estimates as generated by our model depend critically on Eq 4 (see also Fig 5). The purpose of this section is to evaluate the influence of dynamic filtering and Eq 4 on predicted brightness. To this end, we identified three prominent scenarios for explaining corresponding classes of brightness illusions (Fig 7).

The luminance pattern giving rise to Scenario 1 differs only in its spatial layout (Fig 7A), as all structures have the same intensity value. In this case, the patterns with high spatial correlations are attenuated by the dynamic filter, while patterns with lower spatial correlation are somewhat increased such that a brightness contrast effect in predicted for the central disk. Scenario 2 is defined by luminance patterns with similar spatial structure but different intensity range (see Fig 7B). Here the effect is limited by the size of the dynamic filter. We observed that the dynamic filter will not only reduce the spatial correlations, but it will also act as a contrast filter, if the redundant activity is in a sufficiently small spatial region. As a result, redundant activity with higher (lower) intensity than the other patterns would be increased (decreased). If this increment (or decrement) is sufficiently big, it will produce a major (minor) brightness contrast effect.

In Scenario 3, the major contribution to predicted brightness is caused by the Contrast-Luminance channel and modulation (denominator of Eq 4, see Fig 5). Edges in the Contrast-Luminance channel might be enhanced via the Gain Control Map. The enhanced edges eventually produce a boost in (estimated) brightness contrast. The degree of boosting depends on the ratio between the activity (after boosting) and the control parameter (upper bound) in Eq 4. An example of this effect can be observed by comparing both input images and their profiles (at the edges) in Fig 8. It is essential to highlight that Scenario 3 serves just to explain Chevreul's illusion and Mach Bands (see results section), but is not relevant for all other brightness displays.

Results

All of the following simulation results are based on the same set of model parameters. This means that model parameters were never changed. This section presents simulation results (i.e. brightness predictions) from our model. The first subsection focuses on contrast effect: Simultaneous Brightness Contrast, Benary Cross and Reverse Contrast. The second subsection is

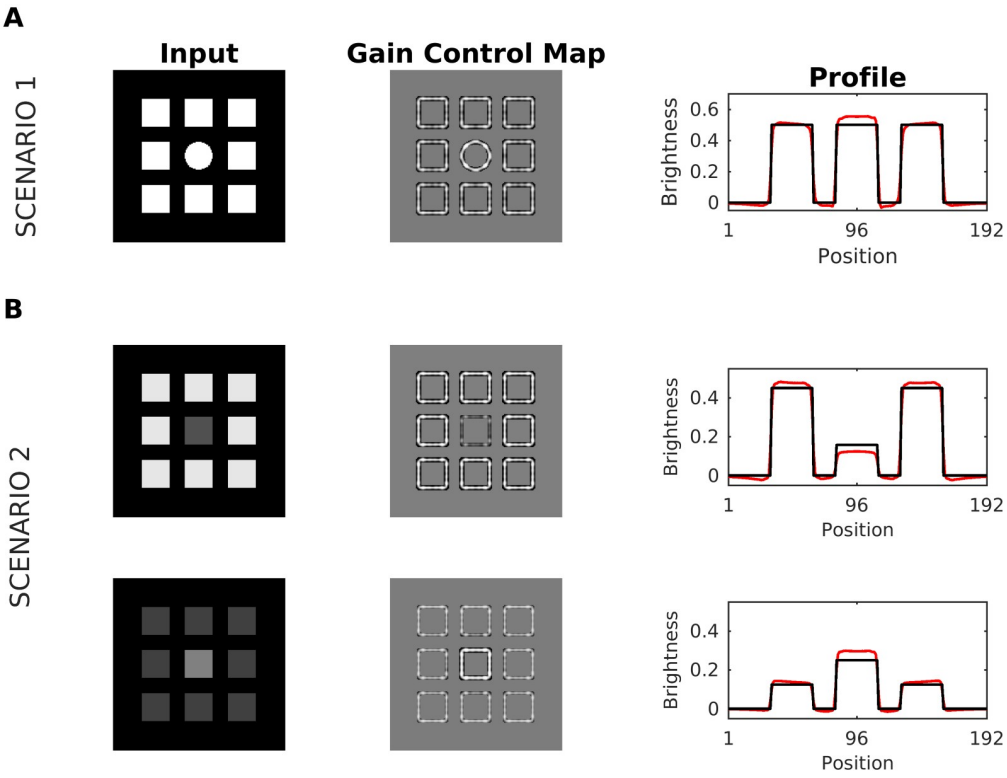


Fig 7. Scenario 1 and scenario 2. (A) Scenario 1: A disk embedded in a redundant pattern of eight squares served as input (first column). The middle column depicts the corresponding gain control map G, and the right column the profiles of input (black line) and brightness estimation (red line). The brightness of the center disk ts enhanced with respect to luminance, meaning that a brightness contrast effect is predicted merely based on redundancy (but not on grounds of luminance—note that all features have the same luminance). (B) Scenario 2. (Top) The input consists of a series of nine squares arranged in a spatially redundant pattern, where the middle square has a different luminance. The profile plot suggests an overall increase in brightness contrast: Brightness of the middle square is further reduced, while the brightness of the surrounding squares is enhanced. (Bottom) While the brightness contrast also increased in the display with the bright middle square, this increase in contrast is caused nearly exclusively by the middle square.

https://doi.org/10.1371/journal.pcbi.1007907.g007

dedicated to assimilation effects: White's effect, Todorović's Illusion, Dungeon illusion, Checkerboard illusion, and Shevell's Rings. The third subsection shows our brightness predictions for the Craik-O'Brien-Cornsweet effect, Hermann/Hering grid, Chevreul's illusion (including the luminance pyramid), Mach Bands, and Grating induction. Finally, the last subsection shows how our model deals with Real-World Images and Noise. It is essential to highlight that all input images were normalized such that pixel intensity ranged between −0.5 and 0.5.

Brightness contrast effects

Simultaneous brightness contrast (SBC). The SBC display consists of two gray patches with identical luminance which are embedded in a dark and bright background, respectively. The patch on the bright background is perceived as darker than the patch on the dark background (Fig 7A). SBC can be attributed to low-level processing. For example, retinal ganglion cells may enhance patch contrast by lateral inhibition. However, other studies suggest that SBC may involve higher-level processing as well [82]: The apparent brightness of the patches can be modulated by the region surrounding the patches (=spatial context). In fact, psychophysical studies report that the contrast effect is perceived more intense for smaller patches [83–87]. Fig 9 shows the estimated brightness for SBC. In our model, the effect conforms to

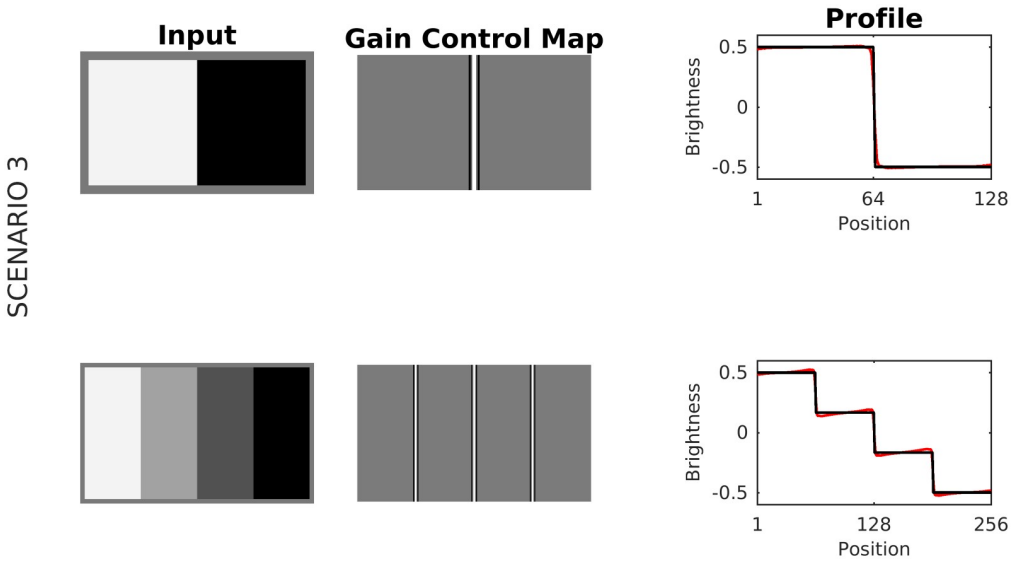


Fig 8. Scenario 3. A luminance step and a luminance staircase, respectively, served as input images. Activity in response to the luminance step is close to the control parameter of Eq 4 ($\tau = 0.5$), producing barely changes in the corresponding brightness estimation at the edges. In contrast, for the luminance staircase, the activity at the edges is relatively far from the control parameter, inducing a boost (an increment of brightness contrast) in the corresponding brightness estimation.

https://doi.org/10.1371/journal.pcbi.1007907.g008

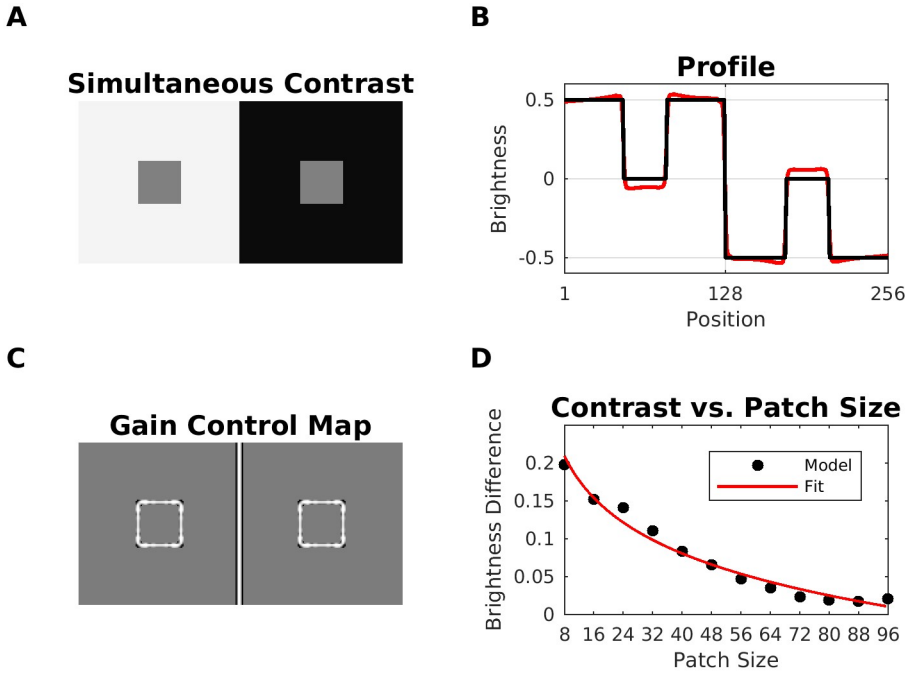


Fig 9. Model prediction for simultaneous brightness contrast (SBC). (A) Simultaneous brightness contrast display (model input). (B) The corresponding Gain Control Map. (C) Profile plot of the estimated brightness map (red line) and the input (black line). (D) Mean absolute brightness difference between the left and the right patch as predicted by our model (filled circles). The continuous (red) lines show the fit of $y = a + b\log(x)$ to the model data. The fit was carried out by linear regression with fitting parameters: intercept $a = -0.3851$, slope $b = -0.0823$, $R^2 = 0.9831$, and $RMSE = 0.0086$.

https://doi.org/10.1371/journal.pcbi.1007907.g009

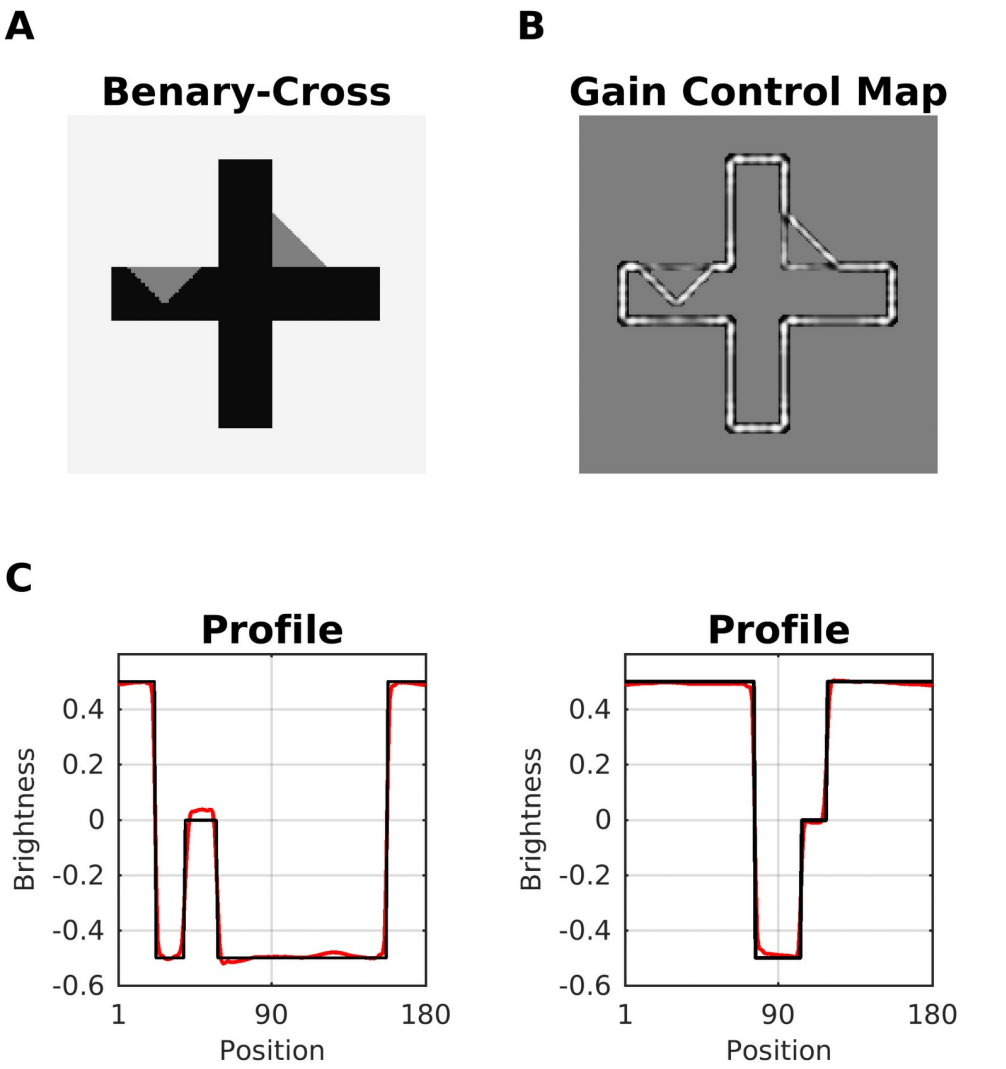


Fig 10. Prediction for Benary Cross illusion. (A) Benary Cross (input). Both triangles have the same intensity, but the triangle embedded in the cross is perceived as brighter. (B) In the gain control map, redundant edges (=aligned with the cross) of the triangle are weakened. (C) Profile plots of predicted brightness (red line) versus luminance (black line). The left profile plot shows the left triangle.

https://doi.org/10.1371/journal.pcbi.1007907.g010

Scenario 1. In SBC, the patterns with low spatial correlation are the patch edges with equal intensity, what causes an enhancement of their contrast after gain control (Gain Control Map: Fig 9C). This translates to an increased contrast in predicted brightness (profile plot in Fig 9B). We also studied the relation between patch size and their predicted brightness. In agreement with previous studies, Fig 9D shows a logarithmic relationship between patch size and our brightness estimation [88].

Benary cross. The Benary Cross [89] is composed of a black cross and two gray triangles with the same luminance (Fig 10). The triangle embedded in the cross is perceived as brighter than the other. Notice that both triangles have identical contrast edges—one white to gray and two black to gray. This effect cannot be explained by lateral inhibition and is usually attributed to "belongingness theory", where the region in which the triangle appears to belong to induces a contrast effect [89]. Noise masking experiments support the idea that the effect is caused by

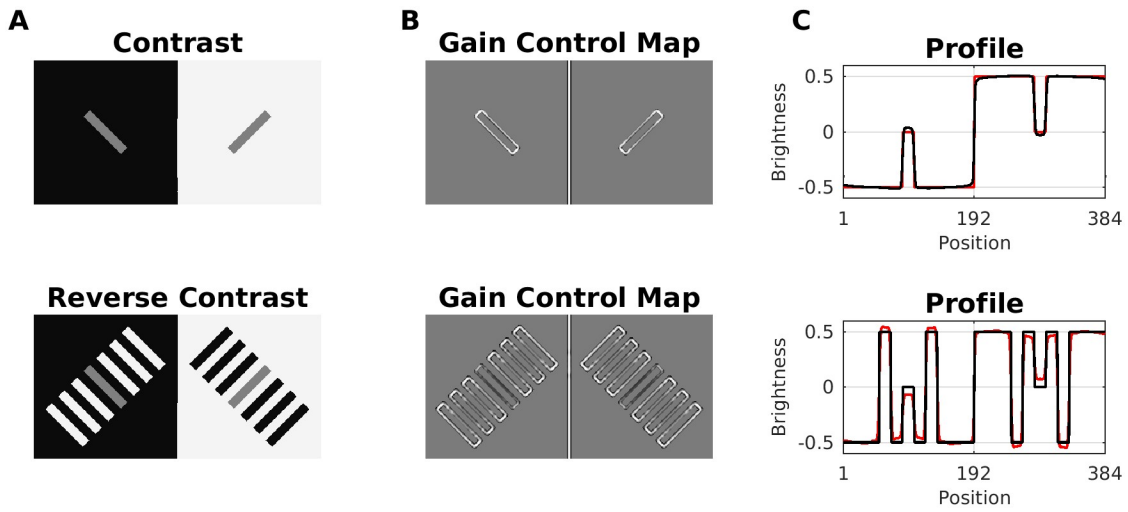


Fig 11. Model prediction for reverse contrast effect. (A) Simultaneous Brightness Contrast (SBC, top) and Reverse Contrast (bottom) which is constructed by adding flanking bars to the SBC configuration. All gray patches have the same luminance. Reverse contrast can be explained either by assimilation with the in-between bars that have the same intensity as the background, or as contrast with the flanking bars that have the opposite luminance to the background. (B) Gain Control Maps obtained by dynamic filtering. Notice the suppression of parallel edges corresponding to flanking patches with the same intensity. (C) Profile plots of predicted brightness (red line) versus luminance (black line).

https://doi.org/10.1371/journal.pcbi.1007907.g011

low-level mechanisms [90]. Our model predicted the brightness difference in the triangles according to two scenarios. According to Scenario 1 (but also 2, considering the intensity differences), the redundant patterns correspond to those edges of the triangles which are aligned with the cross. These are attenuated, while the non-redundant edges are enhanced. The Gain Control Map suggests (Fig 10) a rather balanced effect, which is confirmed by the model's predicted brightness map. Notice, however, that the length of the non-suppressed edges is bigger for the left triangle.

Reverse contrast. Gilchrist & Annan (in [91]) suggested that simultaneous brightness contrast (SBC) can be reversed (e.g. by overcoming lateral inhibition) by adding more structures to the original SBC display. This is Reverse Contrast (Fig 11). The purported mechanism acts on grounds of perceptual grouping of these structures.

Our model predicted the reverse contrast effect according to Scenario 1 and 2, respectively. In case of SBC, dynamic filtering increments the activity of the non-redundant edges that outline the two patches (Gain Control Map: Fig 11B). In the case of reverse contrast, the redundant activity depends on both edge orientation (Scenario 1) and contrast polarity (Scenario 2). Accordingly, all parallel edges of the (flanking) patches with equal intensity are weakened by the dynamic filter. However, the central patch has a different intensity than the flanking patches, and its edges are enhanced. In order to better understand how our model predicted reverse contrast, we probed it with further configurations (see Fig 11). We observed that the change in brightness of the gray patches increases as a function of the number of flanking bars (Fig 12A). On the other hand, if the flanking bars were misaligned to various degrees (disrupting the good continuation principle of perceptual organization), the effect was considerably reduced (Fig 12B). Both results stand in agreement with psychophysical experiments [92]. However, in the latter study the authors examined displays with even more configurations that our model cannot predict (results not shown).

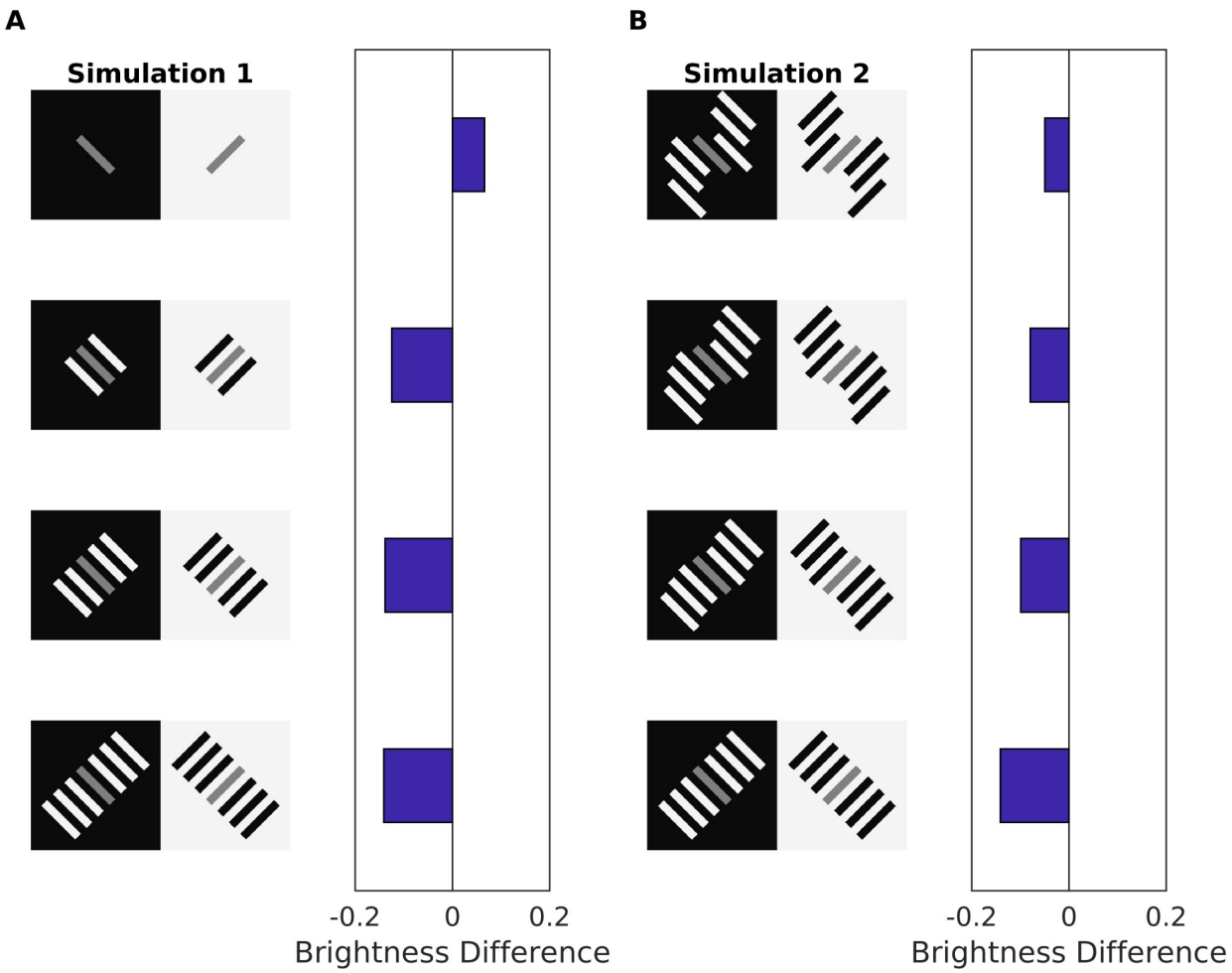


Fig 12. Model prediction for reverse contrast effect for displays with different configurations. (A) Reverse contrast with a varying number of adjacent bars to the gray patch. The bar plots show the predicted brightness difference between the gray patches for the corresponding display (a positive value indicates contrast, while negative values "reverse contrast"). (B) Reverse contrast where the good continuation of the end points is varied. This in turn affect the suppression of redundant edges, which increases with the alignment of the flanking bars.

https://doi.org/10.1371/journal.pcbi.1007907.g012

Brightness assimilation effects

White's effect. Fig 13A shows the White's effect, where two gray bars with identical luminance are embedded in alternating black and white stripes. The bar on the black stripe is perceived as brighter as the other one. Lateral inhibition cannot account for this effect, and it has been suggested that the effect is caused by assimilation [66, 93, 94]. Assimilation means that the brightness of the flanking stripes averages with the gray bars, and therefore one expects that reducing the bar height would also reduce the strength of assimilation. However, experimental data indicate that the perceived difference between the bars increases with smaller heights [95], and that bandpass-filtered noise with the same orientation as the stripes enhanced the effect, while with perpendicular orientation the effect was diminished [90]. Therefore, White's effect seems to be principally generated by contrast at the horizontal edges of the bars (Fig 13B and 13C), and to a lesser extend by assimilation from the flanking stripes [96]. In fact, a mainly contrast-based account is supported by the Gain Control Maps of Fig 13A and 13B. Because the vertical edges (assimilation) are highly redundant, their activity is reduced (Scenario 1 & 2). The brightness estimation is dominated by the horizontal edges of the bars

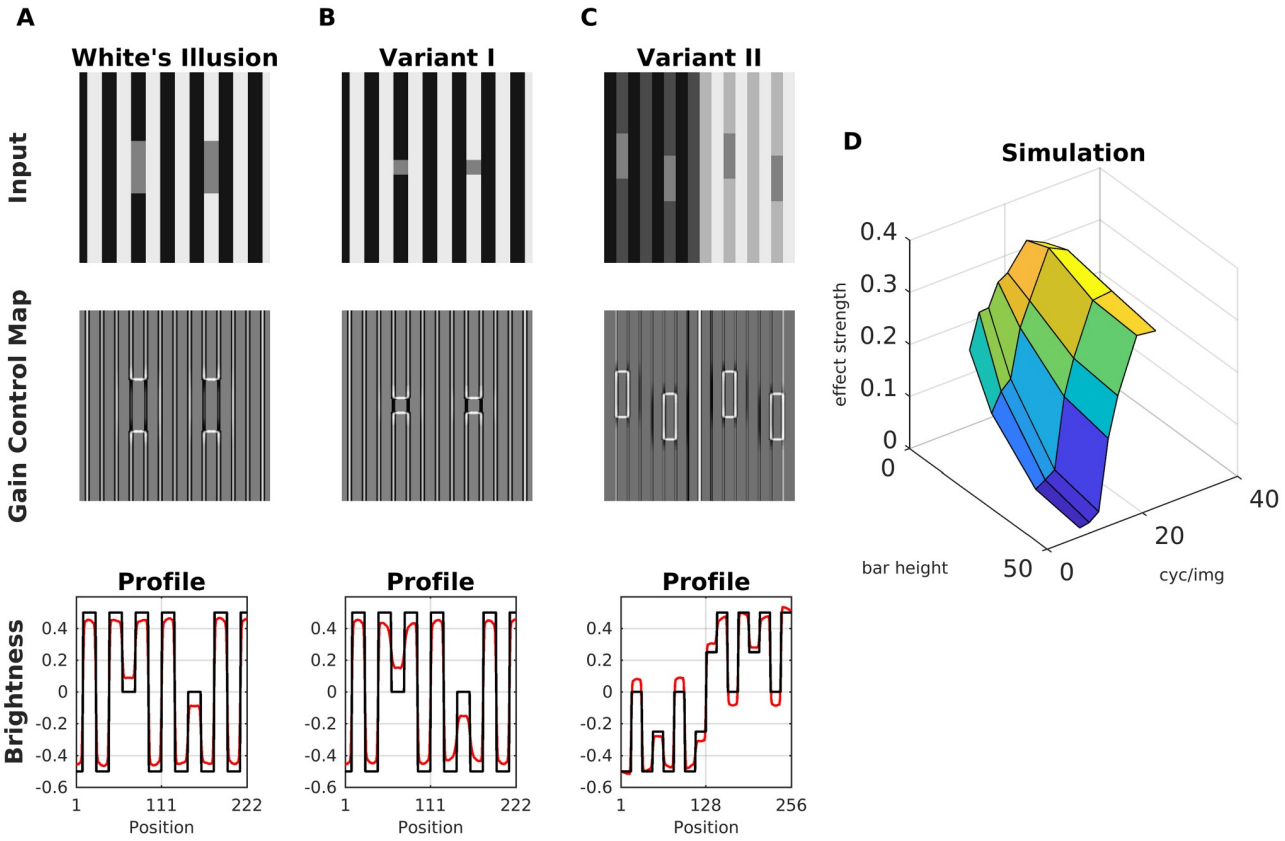


Fig 13. Model prediction for White's illusion. (A) Top: White's illusion; middle: the corresponding gain control map; bottom: profile plot of estimated brightness (red line) and luminance (black line). (B) With smaller bar height, the brightness difference between the bars increases. (C) Modification of White's illusion which produces a strong contrast effect. (D) Surface plot of the estimated brightness difference (effect strength) between the bars as a function of bar height (in units of pixels) and spatial frequency of the background stripes (in units of cycles/image). Image size was 256 x 256 pixels.

https://doi.org/10.1371/journal.pcbi.1007907.g013

(contrast), which are enhanced. The vertical edges nevertheless account for a residual assimilation, but the effect on estimated brightness is less for Fig 13B than for Fig 13A. Therefore, the display with the smaller bars (Fig 13B) has a higher predicted brightness difference between the patches, because less activity from the vertical edges "mixes" with that from the horizontal edges during filling-in.

Despite of the presence of stripes (but with low contrast), the display of Fig 13C shows a clear contrast effect of the bars according to Scenario 2. We also studied the relation between the target size and brightness estimation. We observed that the predicted brightness of the bars could be modified as a function of bar height and spatial frequency of the background (Fig 13D). Specifically, the predicted brightness difference between the bars increases both with decreasing bar height and with increasing spatial frequency. These model predictions are in agreement with previous studies [20].

Todorovic's illusion. Todorovic's illusion consists of a display with two luminance disks with identical intensity and two sets (black and white) of four squares. The original illusion is designated as Context B in Fig 14, where the disks are occluded by the squares. The test patch occluded by the white squares appears to be brighter than the other. This illusion was originally explained in terms of T-junctions [53]. However, Yazdanbakhsh et al. [50] showed that the effect persisted without T-junctions. Later, [97] studied different variations of Todorovic's

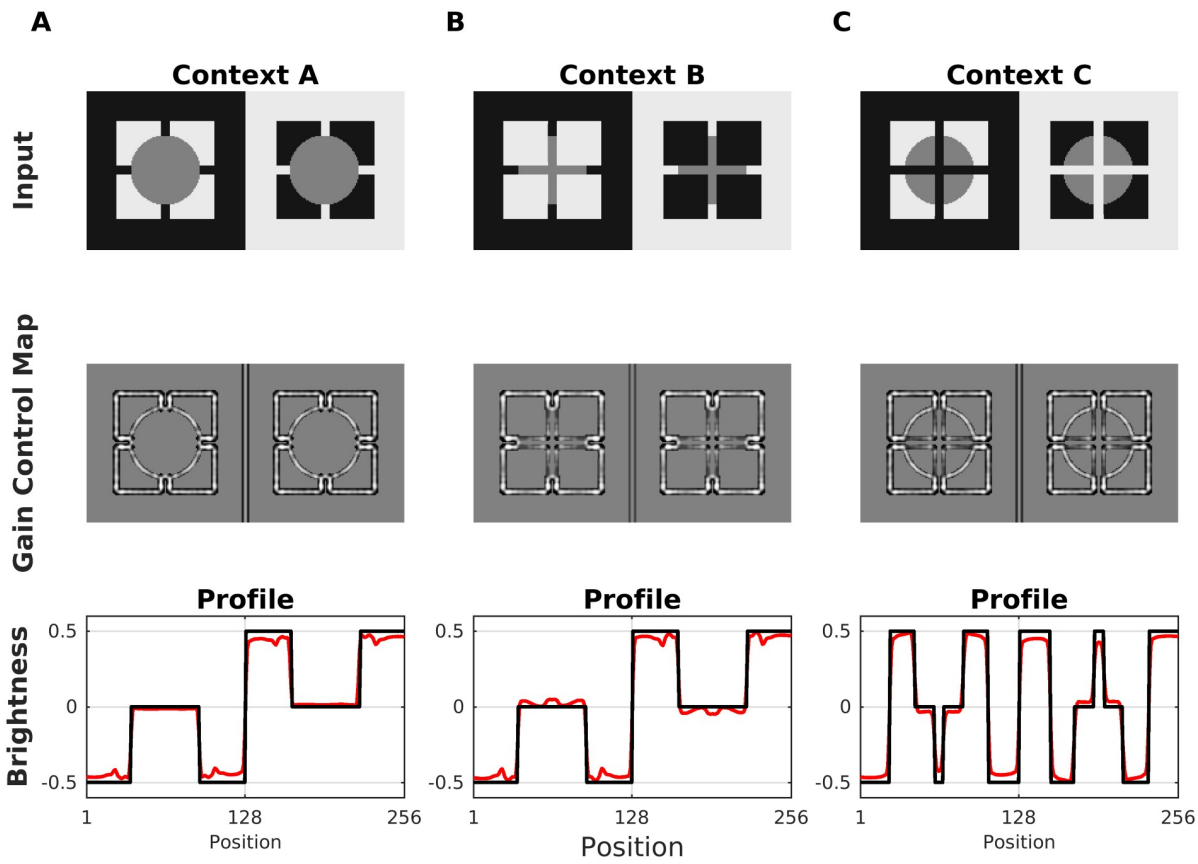


Fig 14. Model prediction for Todorovic's illusion. (A) Top: A variation of Todorovic's original illusion with the gray disks in the foreground (Context A). An effect is hardly perceivable. Middle: The corresponding Gain Control Map. Bottom: Profile plots of estimated brightness (red line) and luminance (black line). The model predicted at most a very weak effect. (B) Original Todorovic Illusion (Context B), where the occluded left disk is perceived as being brighter. (C) Reversed Todorovic Illusion (Context C). Now it looks like viewing the disks on a single square background through a window cross, and the left disk is perceived as being darker.

https://doi.org/10.1371/journal.pcbi.1007907.g014

Illusion (labeled Context A and Context C in Fig 14). They found that the target size interacted with the strength of assimilation. The original effect can be reversed according to Context C (Fig 14), which looks like looking through a window cross. It can also be abolished by moving the disk into the foreground (Fig 14, Context A). The bottom row of Fig 14 shows the profile plots of the brightness maps produced by our model. The results could be understood by analyzing the corresponding gain control maps. For Context B, the activity of the occluding edges (disk with squares) is reduced by dynamic filtering. As a consequence, less contrast is produced in the brightness estimation. On the other hand, the edges between the disk and the (black or white) background are enhanced, which produces more contrast in the brightness estimation. This double effect combined to the finally predicted brightness. For Context C, the effect is analogous to Context B, but with opposite disk brightness. With Context A, edge activity along the disks' circumferences is enhanced. Note that the edge covers the squares as well as the background. Although this enhancement of activity produced locally more contrast, it is approximately the same for the two disks, thus producing almost no effect (profile plot of Context B). Fig 15 shows simulation results for the three Contexts as a function of disk size, where we observed qualitatively similar results to previous studies [97].

More assimilation displays: Dungeon, checkerboard and shevell. The predictions of our model generalize well to further assimilation displays. Fig 16 shows the Dungeon illusion [52],

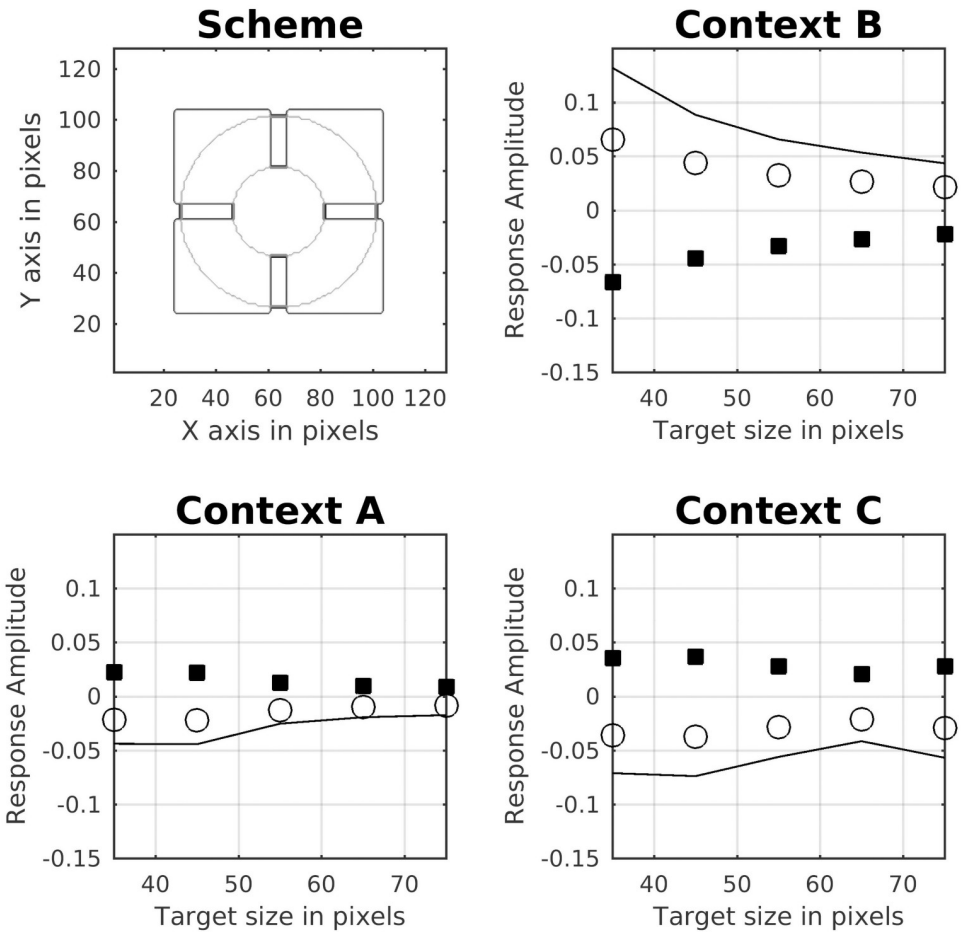


Fig 15. Brightness dependence on target size for Todorovic's display. The Scheme shows the smallest and biggest disk size that was used with respect to the squares in order to generate the plots. Each plot indicates the brightness effect for each of the three Contexts shown in Fig 14. The empty circles indicate the predicted brightness of the disk with the white squares. The filled symbols show disk brightness with the black squares. The continuous lines show the estimated brightness difference between both disks and indicate the predicted strength of the illusion.

https://doi.org/10.1371/journal.pcbi.1007907.g015

the Checkerboard illusion [23] and Shevell's Ring [98]. Although all three reveal an assimilation effect on the gray areas, they are different with respect to their spatial configuration. Notice in particular the absence of T-junctions in the Shevell's Ring display. Fig 16 also shows the corresponding simulation results. All three illusions can be explained according to Scenario 2 since the edges corresponding to the gray areas represent redundant patterns with low intensity.

Further visual illusions

Craik-O'Brien-Cornsweet effect (COCE). Fig 17A shows the COCE, which consists of regions separated by opposing luminance gradients ("cusps") starting at the edges. The cusps drop quickly to a homogeneous gray level and thus the regions between the edges have the same luminance. Nevertheless, especially at low contrast, the gradients seem to fill into the intermediate regions, such that the display is perceived as a low-contrast rectangular wave. The perception of a rectangular wave is less pronounced with high contrast cusps.

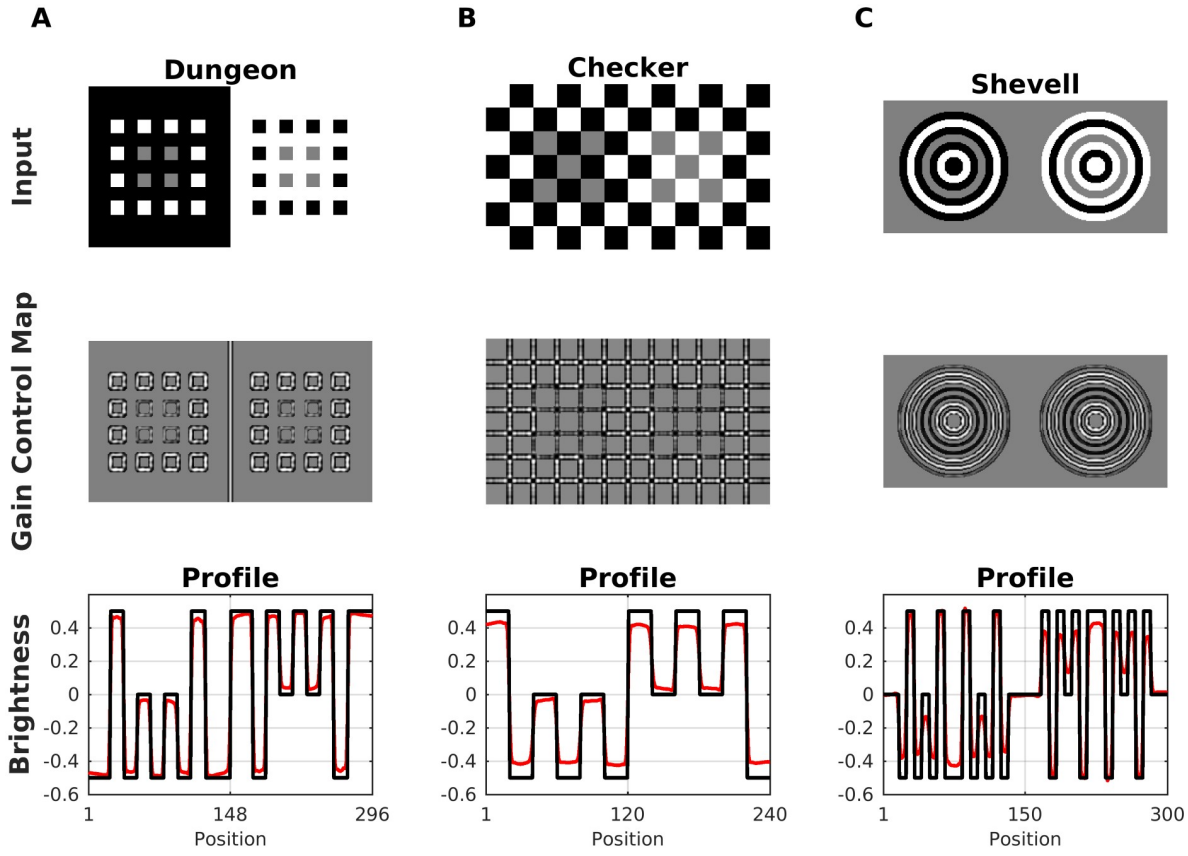


Fig 16. Model predictions for Dungeon, Checkerboard and Shevell. In each display, the gray areas have the same luminance, yet they are perceived differently because of assimilation with the adjacent structures. (A) Top: Dungeon illusion. Middle, corresponding gain control map. Bottom, profile plot of the estimated brightness (red line) compared with the input (black line). (B) Checkerboard illusion. (C) Shevell's Rings. Notice that this illusion cannot be explained with T-junctions.

https://doi.org/10.1371/journal.pcbi.1007907.g016

Our model predicted the COCE, but the explanation is more intricate. At first sight, being a filling-in type effect, it should be predicted in a straightforward way by our model. We verified that without gain control (Eq 4), the COCE cannot be predicted. Could it be then that the effect is produced by the low-pass filtering which is applied after the gain control mechanism (see method section). This is not the case, since the removal of low-pass filtering did not affect the prediction of the COCE (see profile plot 2 in Fig 17A). Therefore, the gain control mechanism contributes to producing the effect. Indeed, the luminance gradients cause negative values (indicated by black lines) in the Gain Control Map around the edges (cf. Fig 17A). As a consequence, activity corresponding to the luminance gradients is suppressed by the gain control mechanism, which furthermore reduces the peak activity at the edges. After all, the gradients are "ignored", and our model generates the COCE as a result of assimilation of the edges. This explanation is also consistent with the cow-skin illusion, which is a variant of the COCE without luminance gradients. It is composed exclusively of adjacent black and white lines, and the empty regions are randomly arranged. Fig 17B shows the brightness prediction for the cow-skin illusion.

Hermann/Hering grid. Fig 18A (in the top) shows the Hermann/Hering Grid (HG). Although the luminance between the black squares is constant, illusory gray dots appear at the (white) intersections. The textbook explanation deems the center-surround receptive fields of retinal ganglion cells as the principal acting mechanism [99]: Assume a circular receptive field

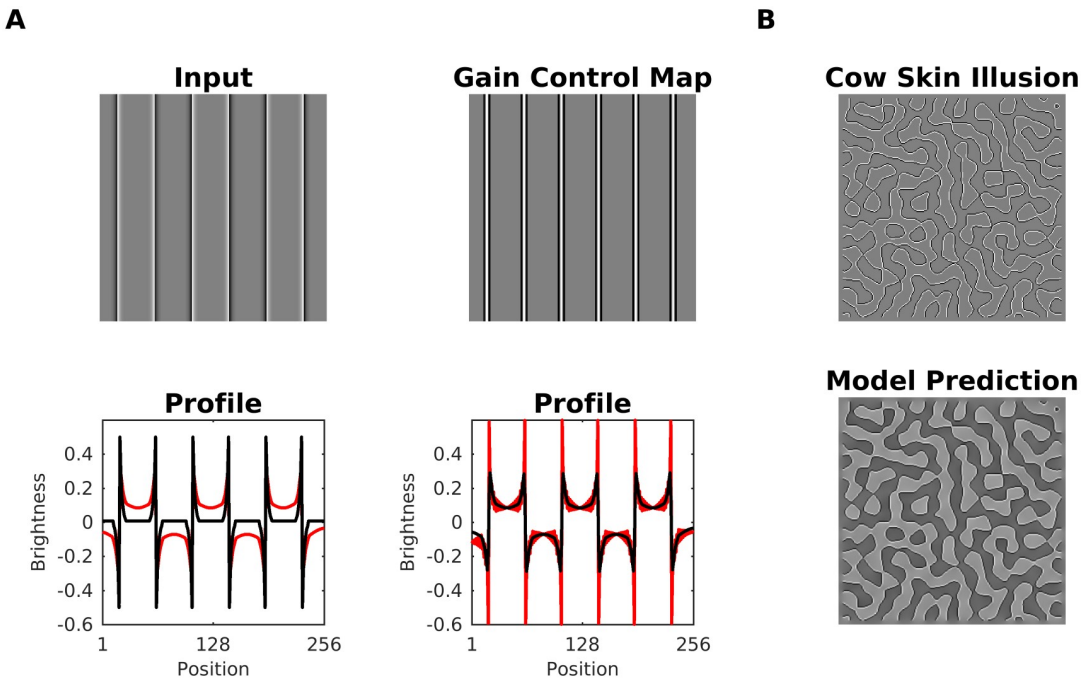


Fig 17. Prediction for Craik-O'Brien-Cornsweet Effect (COCE). (A) Top: COCE along with the Gain Control Map. Notice that black lines (adjacent to the edges) in the Gain Control Map, which indicate negative values. Bottom: The first profile plot shows the predicted brightness (red line) along with input luminance (black line). The second profile plot shows the predicted brightness without low-pass filtering after the gain control mechanism (Eq 4). (B) The cow-skin illusion is a variant of the COCE without luminance gradients. It consists only of adjacent black and white lines. Our model consistently predicted this illusion: The brightness map generated by our model is shown at the bottom right.

https://doi.org/10.1371/journal.pcbi.1007907.g017

with an excitatory center which has the same width as the white grid lines. The inhibitory surround covers in addition the black squares. If the center is located right at an intersection, it receives more inhibition from the surround (from two white lines) than when the center is positioned between two intersections (inhibition from one line). This translates to a brightness reduction at the intersections, but not in between. This mechanism, though, is insufficient to explain why the effect is considerably reduced (or even removed) if the bars are slightly corrugated (Fig 18A, bottom). It is also reduced as a function of the ratio between grid line width and block width, where no effect is produced for a ratio of one [100]. Our model predicted the darkening of the intersections for the HG, and the absence of darkening for the corrugated HG. The HG adheres to Scenario 1, where redundant activity is inhibited. Inhibition is especially strong at the intersections of the Gain Control Map. In this way, an assimilation effect is induced. As to the corrugated HG, the corners also represent the redundant patterns, but because the spatial structure are less regular, the inhibition is correspondingly weaker (compare the Gain Control Maps shown in Fig 18B). Consequently, the brightness reduction at the intersections is considerably weaker for the corrugated HG.

Fig 18D shows the dependence of the darkening effect on the ratio between grid line width and block width. In agreement with the results from [100], we find that the darkening effect decreases while the ratio approaches one. We were unable to predict further results with the HG that were presented in [100].

Luminance staircase and pyramid (Chevreul's illusion). Chevreul's illusion consists of increasing levels of luminance, arranged as a staircase or as a pyramid. Although luminance is constant at each step, one perceives an illusory brightening on the side of each step where the

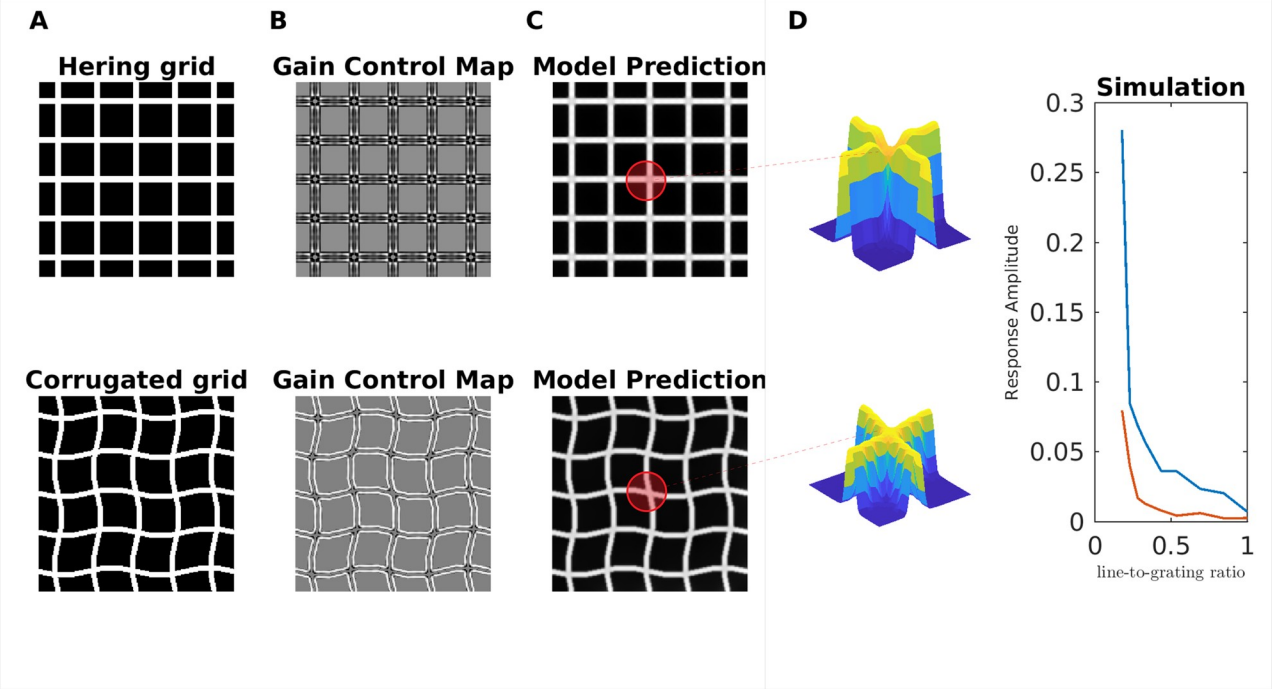


Fig 18. Model predictions for Hermann/Hering grid and corrugated grid. (A) Hermann/Hering (HG) illusion and a corrugated version of it. At the intersections of the white grid lines, illusory gray spots are perceived in the HG, but not in the corrugated grid. (B) The corresponding gain control maps of the input images of A. (C) The brightness estimation from our model. The surfaces plots (insets) illustrate the 3D profile of the brightness estimation corresponding to regions highlighted with red. (D) The predicted brightness magnitude at the intersections as a function of the ratio $\frac{\text{line width}}{\text{grating width}}$, where the red curve corresponds to the corrugated grid, and the blue curve to the HG.

https://doi.org/10.1371/journal.pcbi.1007907.g018

adjacent step is darker, and an illusory darkening on the other. In the pyramid version, one perceives in addition (illusory) glowing diagonals (Fig 19). The effect is absent at the lowest (black) and highest (white) luminance level, and is considerably reduced on the middle step for a staircase made up of three steps.

All aspects of Chevreul's illusion are consistently predicted by the gradient system, which is a computational model for representing luminance gradients [10, 101]. The idea behind gradient representations is to capture the smooth variations of luminance (illumination effects) in order to help to disentangle reflectance from the illumination component in luminance (since luminance is the product of reflectance with illumination).

Brightness predictions from our model for the luminance staircase and the pyramid are shown in Fig 19. The illusory whitening and darkening at the stairs can be explained according to Scenario 3: On the one hand, the gain control map increases the activity of the Contrast-Luminance channel. On the other hand, the increase in excitation is offset by the modulation mechanism of Eq 4, thereby producing non-uniform brightness activity at the stairs. The glowing diagonals of the Pyramid Illusion are produced according to Scenario 1, where the activity of non-redundant spatial patterns—especially at the corners—is enhanced. On the other hand, the edges of the staircase represent a redundant pattern, the activity of which is decreased. Consequently, more (less) contrast at the corners (at edges) is generated in the brightness estimation (Fig 19C, bottom). Finally, it is important to emphasize a limitation of our model in this context. We observed that for a big number of steps (i.e., very narrow steps) the dynamic filter "collapses" and the model could not longer predict the illusion nor the glowing diagonals. This is a consequence of the scale-sensitivity of the dynamic filter (i.e., the size of the sampling

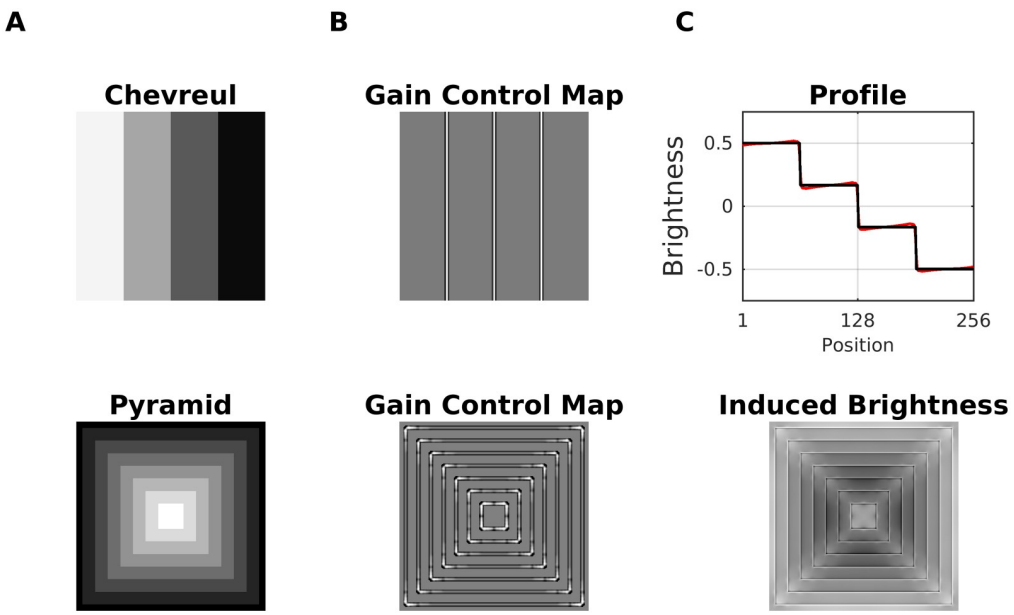


Fig 19. Model predictions for luminance staircase and pyramid (Chevreul's illusion). (A) Top: Luminance staircase. Bottom: luminance pyramid. (B) The corresponding gain control maps as a result of dynamic filtering. (C) Top: Profile plot of the estimated brightness (red line) of the luminance staircase (black line). Bottom: The induced brightness, which consists of the difference map between estimated brightness and input luminance (i.e., brighter gray level mean positive values, and darker gray levels mean negative values).

https://doi.org/10.1371/journal.pcbi.1007907.g019

patches), since with decreasing step size, the staircase eventually approaches a linear luminance gradient, and the filter cannot resolve anymore individual steps.

Mach bands. Mach Bands [102] are illusory glowing stripes that are perceived adjacent to knee points that are connected with a luminance ramp, where the bright (dark) band is attached to the plateau with high (low) luminance (Fig 20). Notice that Mach bands do not cause Chevreul's illusion. The perceived strength of Mach bands decreases when the ramp gets steeper and eventually approaches a luminance step. Also, for very shallow ramps, the perceived strength decreases. The perceived strength has thus a maximum at intermediate ramp widths [103, 104]. The textbook explanation based on lateral inhibition is insufficient to explain the variation of strength with ramp width—it would wrongly predict maximum perceived strength at a luminance step [101, 105–107]. The perceived strength of Mach bands is also modulated by the proximity, contrast and sharpness of an adjacently placed stimuli [105, 107].

The only computational model published so far that quantitatively predicted all published data about Mach Bands is the gradient system [55, 101]. The gradient system suggests that Mach bands are also perceived at the peaks and troughs of a triangular wave. The gradient system furthermore predicts that bright Mach bands are key for the perception of light-emitting surfaces [10]. Our model predicts the Mach bands, as well as the absence of them at steps (see profile plots in Fig 20). It furthermore succeeds in predicting the inverted-U curve of the perceived strength of Mach Bands as a function of the ramp width (Fig 20C). The inverted-U behavior replicates the trend for measured threshold contrasts for perceiving Mach bands [104]. The threshold is assumed to be minimal where the model predicts the maximum brightness. The measured threshold contrasts for the bright Mach band are also shown in Fig 20C.

The inverted-U curve could be explained by two mechanisms which act in opposite ways. (i) If the ramp width decreases, then the activity at the knee points reaches a maximum that

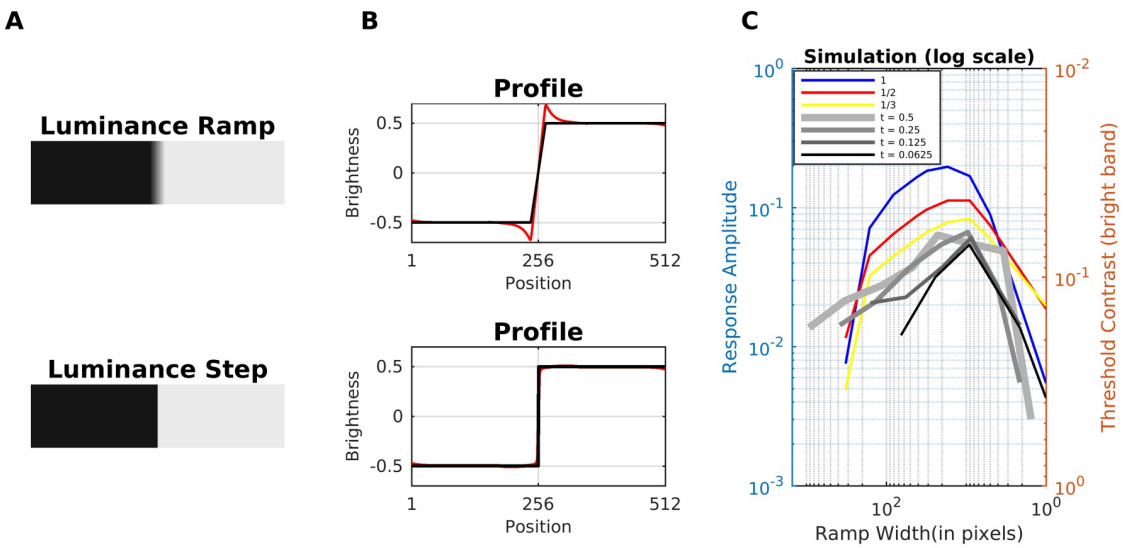


Fig 20. Model prediction for mach bands. (A) Top, A luminance ramp that leads to the perception of Mach bands close to the knee points of the ramp. Bottom, a luminance step (no Mach bands are perceived). (B) Profiles plots of estimated brightness (red line) compared with the corresponding input (black line) of A. (C) Brightness magnitude (at the inflection point of the ramp) as a function of ramp width. The plots show the predictions of our model on the perceived strength of the bright Mach band. The colored curves (left axis label: response amplitude) represent model predictions for different dynamic ranges (i.e., differences between luminance values of the upper and the lower plateau, see legend). The gray curves are the threshold contrasts (axis label on the right) for seeing the bright Mach bands at trapezoidal waveforms according to [104]. The trapezoidal waves are characterized by a shape parameter t (see legend; t = 0.5 corresponds to a triangular wave).

https://doi.org/10.1371/journal.pcbi.1007907.g020

renders the modulation mechanism (denominator of Eq 4) of the gain control mechanism ineffective (Scenario 3; ideal step luminance). If the ramp width increases, then the luminance transition between the plateaus is more gradual, which is associated with less activity at the edge locations. In this way, the edge activity becomes more susceptible to the gain control mechanism towards the maximum of the perceived strength. However, this effect does not remain constant. After a certain ramp width, the activity across the ramp gets comparable to the activity at the knee points, which produces less variability in the energy map E. As a result, (ii) the dynamic filter has less effect in Eq 2, reducing gradually the perceived strength induced by the dynamic filtering.

Grating induction (GI). Fig 21 shows the grating induction (GI) display [108], which consists of two sinusoidal gratings (inducers) separated by a gap (test field). Although it has a constant luminance, an illusory brightness modulation is perceived across the test field if the two inducers are in-phase. The phase of the brightness modulation is opposite to the induction wave. The effect decreases when shifting the phase of the inducer gratings relative to each other, being minimum when the gratings are in anti-phase. The illusory modulation is furthermore attenuated with increasing distance between the inducer gratings and with increasing spatial frequency. The GI can be explained in terms of multi-scale filtering [19, 83]), but also by filling-in models [36]. Notice that a common misconception with diffusion-based approaches (=filling-in models) is that the illusory brightness modulation across the test field would average out. This, however, is usually not the case. The exact explanation depends on the model under consideration. For instance, a mechanisms that counteracts "averaging out" are boundary webs from the boundary contour system (BCS) that extend across the test field

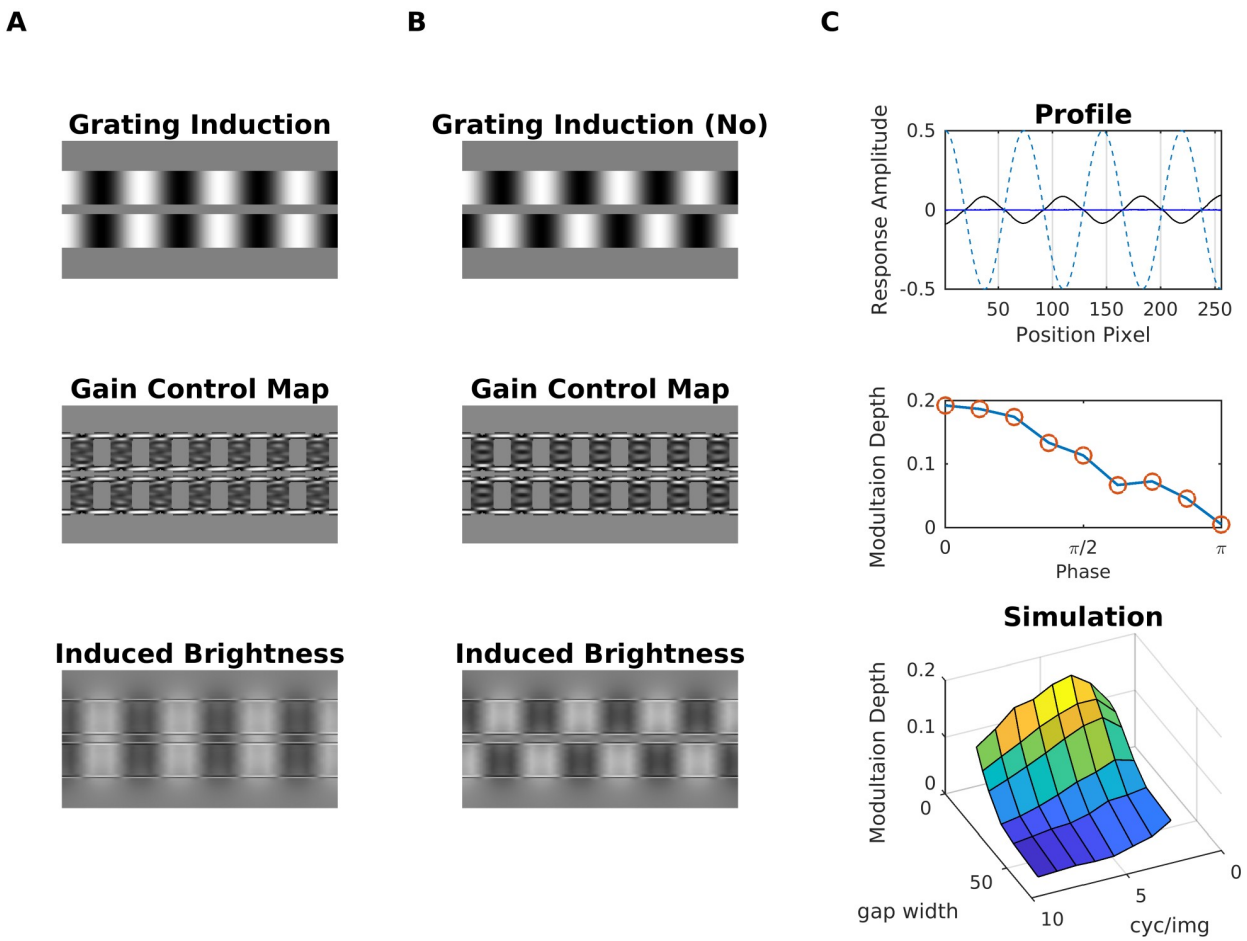


Fig 21. Model prediction for grating induction. (A) The Grating Induction refers to the illusory perception of a brightness modulation across the gap (=test field) between the inducer gratings. The brightness modulation is perceived in opposite phase to the inducer gratings. (B) When the inducer gratings stand in opposite phase to each other, then brightness modulation is considerably reduced. The corresponding gain control maps are shown in the middle row, and the last row shows the induced brightness, which consists of the difference map between estimated brightness and input luminance (i.e., brighter gray levels mean positive values, and darker gray levels mean negative values). (C) Top: Profile plot of brightness estimation for display A (black line) and display B (blue line). The dashed blue line shows the luminance profile of the inducer grating A. Middle, modulation depth as a function of the phase difference between the two inducer gratings. Bottom: Surface plot that shows how modulation depth depends on test field width and spatial frequency of the inducer grating.

https://doi.org/10.1371/journal.pcbi.1007907.g021

and trap feature contour activity (FCS) [34]. Other mechanisms include cross-channel inhibition between brightness and darkness activity during filling-in [109].

Fig 21 shows that our model consistently predicted the illusory modulation of brightness across the test field. The strength of the effect decreases in an approximately linear fashion with increasing phase difference between the inducer gratings (Fig 21C). We also observed from a specific spatial frequency on (4 cycles/image) that the brightness modulation decreases with increasing separation and spatial frequency, respectively, of the inducer gratings (surface plot in Fig 21C). Unlike the rest of the illusions the GI effect was produced mainly at the filling-in stage (Eq 5), and to a lesser degree by dynamic filtering. Dynamic filtering increased the activity at the boundaries of the inducer gratings (see gain control maps in Fig 21); this increment produces a significant contrast in estimated brightness between the inducer gratings and the test field, which eventually propagated (by filling-in) across the test field.

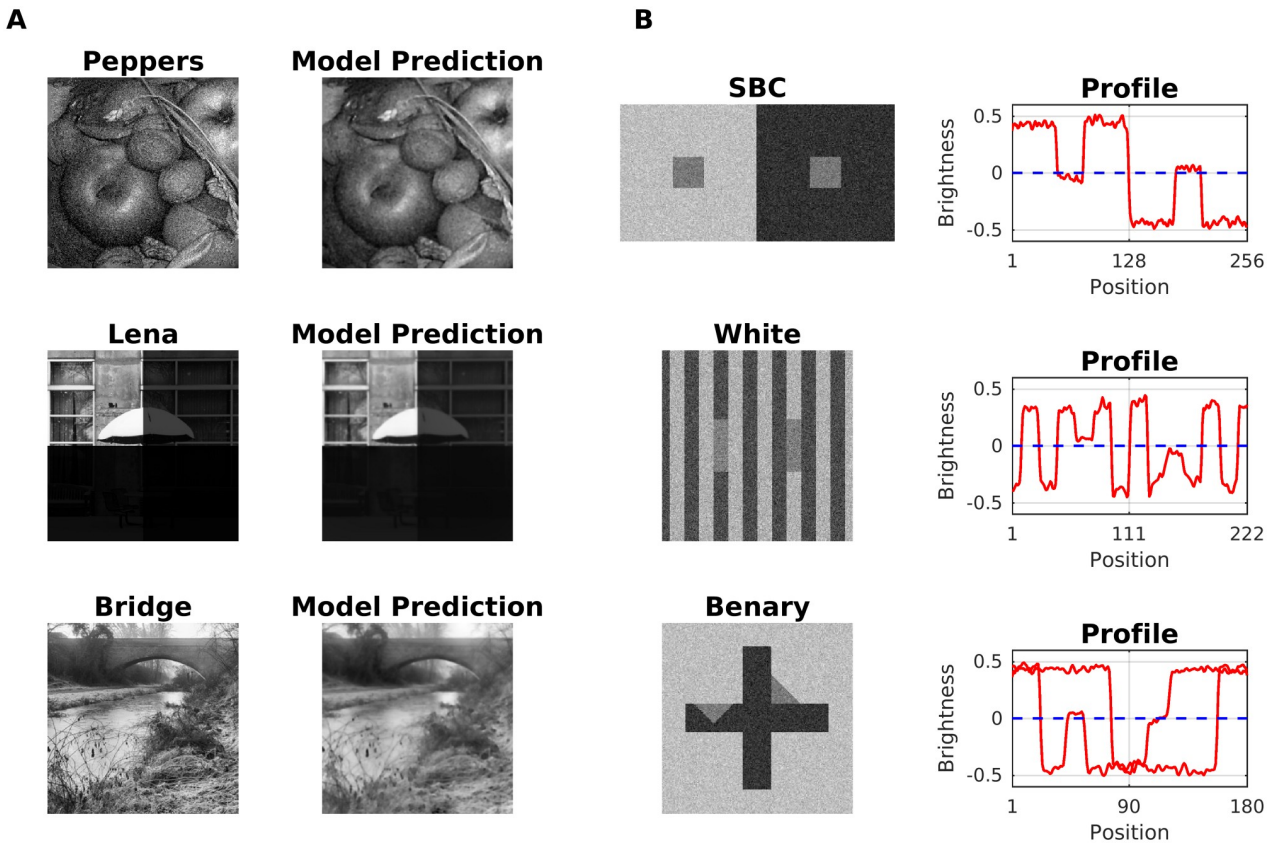


Fig 22. Real-world image processing. (A) Top: Fruits image with additive white noise (SNR = 2.6266dB and PSNR = 8.9813dB) along with the corresponding model output (SNR = 5.4682db and PSNR = 11.8228dB); The models capacity for noise removal is worse than an algorithm based on Tikhonov regularization (SNR = 8.30dB and PSNR = 14.65dB; [111]. Middle: A high-dynamic-range version of a real image (where dynamic range of each quadrant decreases clockwise by one order of magnitude) and model output; the dynamic range of the input is 1, and that of the output is 0.9596. Bottom: Bridge image alongside with corresponding model output (SNR = 6.1084dB and PSNR = 13.6201dB). (B) Top: Simultaneous Brightness Contrast display with additive white noise and corresponding brightness profile (red line) as predicted by the model. The dashed line indicates the gray level of the gray squares. Middle: White Effect. Bottom: Benary Cross.

https://doi.org/10.1371/journal.pcbi.1007907.g022

Real world images and noise

Although synthetic images are a valuable tool for the study of certain aspects of the visual system, it nevertheless evolved to the processing of real-world images. Real-world images provide, therefore, a test of robustness for any model of the visual system. Fig 22 demonstrates that our model is capable of real-world image processing, as well as its ability to handle noisy versions of visual illusion displays (Benary Cross, Simultaneous Brightness Contrast, and White's Illusion). Previously we showed that by using a set derivative filters that cover all orientations (similar to simple cells), Eq 5 globally converges to a stable solution [110]. The convergence is robust against adding noise to the input, or using a high dynamic range of luminance values (Fig 22A, top, middle). Fig 22 suggests that our model's noise reduction performance is worse than a control model based on Tikhonov regularization [111]. The robustness of our model extends to the consistent prediction of visual illusions in the presence of additive noise. In particular, we noted that simultaneous brightness contrast was more sensitive to the presence of uncorrelated noise than assimilation displays. The robustness against noise relates to dynamic filtering, which reduces the correlated (or redundant) spatial information in the edge map. The redundancy of the edges would not be affected by spatially uncorrelated noise. Finally, we

note that dynamic filtering has a couple of limitations with respect to spatially correlated noise, such as band-pass-limited additive noise (see discussion). We did not study this issue in more depth, as it would go beyond the scope of the present paper.

Discussion

The perceived luminance (brightness) of target structure is highly sensitive to its spatial context. Despite of many modeling attempts for brightness, we still have not arrived at a detailed understanding of the corresponding neuronal information processing principles. With our model, we emphasize the role of decorrelation and response equalization, respectively, in brightness perception. Response equalization is implemented by a dynamic filter that adapts to the spatial structure of luminance patterns in each image in order to reduce the redundancy of boundary maps. Coding strategies that aim at reducing activity and thus energy expenditure in organisms are consistent, for example, with efficient coding [60–62, 64, 112], predictive coding [63], whitening [64] or response equalization [65]. In this sense, we propose that brightness perception is the consequence of suppressing redundant (i.e., predictable) information. Our model is build on the latter idea(s), and apart from being able to process real-world images, it predicts a bigger set of visual illusions than any other previously published model.

Our focus is thereby on low-level vision, as our model simulates the activity of simple and complex cells of the primary visual cortex. For each input, the model learns a filter kernel by identifying redundant patterns in (simulated) complex cell responses (i.e., the edge map), and subsequently uses the filter kernel to suppress redundant information (dynamic filtering). Dynamic filtering amounts to response equalization of simulated complex cell responses, much like the previously proposed "Whitening-by-Diffusion" method which directly acts on the (Fourier) amplitude spectrum [65]. The equalized responses are subsequently used for creating a representation of the sensory input by filling-in (brightness estimation). Nevertheless, dynamic filtering is a global mechanism, which was adopted for the ease of implementation. In the primary visual cortex, we expect that dynamic filtering acts in a more local fashion, but still on a spatial scale that exceeds the typical receptive field sizes of V1 neurons. Such non-local mechanisms could be biologically implemented by feedback from mid-level visual neurons with sufficiently big receptive fields for detecting non-local correlations in activity.

We believe that our success in predicting a relatively large number of visual illusions lends some support to our proposed computational principle. Without changing any of our model's parameter values, we are able to predict Simultaneous Brightness Contrast (SBC), White's Effect, Reverse Contrast, Benary's Cross, Todorovic's illusion (with variations), the Dungeon Illusion, the Checkerboard Illusion, Shevell's Ring, the Craik-O'Brien-Cornsweet effect (COCE), the Hermann/Hering grid, the corrugated grid, Chevreul's illusion (including the luminance pyramid), Grating Induction (GI), and Mach Bands. Additionally, for some of the illusions, we were able to reproduce the trend for corresponding psychophysical data (SBC, White, Reverse Contrast, Hermann/Hering grid, Todorovic, GI, and Mach Bands). Despite all of these successes, we must not forget to mention some of the limitations of our model. We cannot predict illusions—without modifying the current parameters—such as achromatic neon "color" spreading, the Ehrenstein illusion, Chubb's Illusion and some variations of the Hermann/Hering grid; Reverse Contrast with different grouping factors, SBC with articulated noise, and Mach Bands with an adjacently placed stimuli.

A further limitation is handling visual illusions where a target patch is surrounded by an articulation pattern. An articulation pattern can be created from a region with uniform luminance. The region is subdivided into small square patches. The luminance of each patch is randomly modulated according to a Gaussian random variable, with the mean value being the

original luminance value, and the standard deviation being the modulation depth. The average luminance across the articulated pattern has to be identical to the luminance of the original uniform region. The articulation patterns would introduce additional spatial redundancy into a luminance display, and the kernel would eventually learn this excess redundancy for dynamic filtering. As a consequence, dynamic filtering may modify the target's edge representation in an unpredictable way. For brightness estimation, this could mean, for instance, that assimilation effects turn into contrast effects or vice versa. This behavior appears to be inconsistent with current psychophysical observations [113], and may hint at additional mechanisms that need to be considered. It cannot be ruled out that additional mechanisms reduce the redundancy along other stimulus dimensions as well, for example luminance, relative contrast, or auto-correlation. The global nature of the dynamic filter represents another trade-off. In order to learn the kernel for dynamic filtering, we sample patches randomly across the input. As a consequence, local statistical information between (unrelated) patches that are far away from each other could be intertwined, affecting their brightness predictions. A possible solution would be to introduce local constrains upon sampling, or as well to introduce a local normalization function, which takes into account local spatial auto-correlations.

Comparison with other models that predict contrast and assimilation

Our approach produces contrast effects by enhancing non-redundant edges through dynamic filtering according to Scenario 1. Assimilation effects are generated by suppressing redundant edges as a function of their relative intensity with respect to the other edges (Scenario 2). We thus do not make any explicit assumptions about how a visual target is related to its context in terms of segmentation, or belongingness and perceptual frameworks, respectively. We do not require the categorization of image features either. In this sense, our approach is more general than previous computational proposals and theories [47, 49, 114], which purport that a stimulus is divided into perceptual frameworks based on anchors [114] or T-junctions [49]. However, it is not clear whether anchors or T-junctions are sufficiently robust cues in real-world images, and actually few previously published models demonstrated the processing of real-world images.

Dynamic filtering is sensitive to the correlation structure of spatial patterns in order to generate contrast and assimilation effects. In this way, the output of our model would not be significantly affected if uncorrelated noise was added to the input. Yet multi-scale models are highly sensitive to additive noise [19, 20, 22], because their predictions depend on a careful readjustment of filter responses across spatial frequencies. Thus, if noise was added to contrast and assimilation displays, then corresponding predictions would be altered, because of corresponding changes in the spatial frequency spectrum [24].

Our model adapts to the statistical structure of each input image. This is to say that we do not evaluate each input image in a previously learned long-term statistical context. A long-term statistical context usually is learned from a big number of input samples in order to derive feature-specific probability distributions. In connection with brightness, a relationship between occurrence frequency of certain types of natural images and brightness perception has been proposed [56–58]. The main limitation of such models is that they require an enormous amount of data, and that visual illusions act much like an associative trigger or they are perceived according to humdrum Bayesian inference.

Conclusion

One might ask whether the range of illusions that we successfully predict with our model can be attributed to a common mechanism. The answer is yes, and the underlying mechanism is

dynamic filtering. Dynamic filtering acts to equalize the amplitude spectrum of a boundary map. In the spatial domain, thus, dynamic filtering depends on pattern redundancy (but not on activity). In this way, non-redundant patterns are enhanced (i.e., contrasted) compared to to redundant patterns (which are assimilated). In conclusion, this study provides a proof of concept of a hypothetical information processing strategy for visual system, based on economizing edge representations. Our predictions are reliant on the self-structure of the visual input, but not on accumulated visual experience, spatial frequency representations, or predefined detectors. Our proposed mechanism does not exclude information processing principles like accumulating visual experience or spatial frequency representations, and should be considered as being complementary to these. Finally, future work should address the understanding of how the statistical structure of the context surrounding a target patch influences its appearance. We also plan to study how different noise structures (as narrow-band, oriented, or correlated) influences the predictions of our model. Our redundancy-reduction hypothesis should be compatible with all levels of information processing. This means that redundancy reduction likely might apply to higher-order patterns and shapes that form the primitives for object recognition.

Supporting information

S1 Text. A. Gabor filters. In this section are described the parameter values and a mathematical description of unbalancing the ON/OFF subregions for the filters used in the Contrast-Luminance channel and the Contrast-only channel. **B. Energy Map**. In this section are included the mathematical details corresponding to the local energy map. **C. Dynamic filtering with zero-phase whitening (ZCA)**. In this section are described the mathematichal details to perform dynamic filtering of our model. **D. Solving** Eq 5. In this section, a solution for Eq 5 is derived, which is used to estimate the output (brightness map) of our model.
(PDF)

Acknowledgments

We thank the Visca Group for laboratory space and Gonçalo Correia for useful discussions.

Author Contributions

Conceptualization: Alejandro Lerer, Matthias S. Keil.

Formal analysis: Alejandro Lerer, Matthias S. Keil.

Investigation: Alejandro Lerer, Matthias S. Keil.

Methodology: Alejandro Lerer.

Project administration: Hans Supèr.

Resources: Hans Supèr.

Software: Alejandro Lerer.

Supervision: Hans Supèr, Matthias S. Keil.

Validation: Hans Supèr, Matthias S. Keil.

Writing – original draft: Alejandro Lerer, Matthias S. Keil.

Writing – review & editing: Alejandro Lerer, Matthias S. Keil.

References

1. Rudd Michael E. Lightness computation by the human visual system. Journal of Electronic Imaging 2017, 26(3), 031209. https://doi.org/10.1117/1.JEI.26.3.031209
2. Rossi A. F., Paradiso M. A. Neural correlates of perceived brightness in the retina, lateral geniculate nucleus, and striate cortex. Journal of Neuroscience 1999, 19(14), 6145–6156. https://doi.org/10.1523/JNEUROSCI.19-14-06145.1999 PMID: 10407050
3. Grossberg A. The visual world as illusion: The ones we know and the ones we don't. In Oxford Compendium of Visual Illusions (Shapiro A. and Todorovic D., Eds.) Oxford, United Kingdom: Oxford University press 2017, Chapter 7, pp. 90–118.
4. Schmid A. C. & Anderson B. L. Perceptual dimensions underlying lightness perception in homogeneous center-surround displays. Journal of Vision 2017, 17(2):6, https://doi.org/10.1167/17.2.6 PMID: 28245501
5. Anderson BL, Whitbread M, de Silva C. Lightness, brightness, and anchoring. Journal of Vision 2014, Aug 2; 14(9):7–7. https://doi.org/10.1167/14.9.7 PMID: 25104828
6. Kingdom F.A.A. Lightness, brightness and transparency: A quarter century of new ideas, captivating demonstrations and unrelenting controversy. Vision Research 2011, 51(7):652–673. https://doi.org/10.1016/j.visres.2010.09.012 PMID: 20858514
7. Anderson B. L. & Winawer J. Image segmentation and lightness perception. Nature 2005, 434:79–83. https://doi.org/10.1038/nature03271 PMID: 15744303
8. Tappen M.F.Image Decompositions. In: Ikeuchi K. (ed) Computer Vision—A Reference Guide. Springer, 2014 Boston.
9. Rother C., Kiefel M. Zhang L., Schölkopf B. & Gehler P.V. Recovering Intrinsic Images with a Global Sparsity Prior on Reflectance. Advances in Neural Information Processing Systems 2011, 24 (NIPS 2011).
10. Keil MS. Gradient representations and the perception of luminosity. Vision Res. 2007 Dec 1; 47(27):3360–72. https://doi.org/10.1016/j.visres.2007.09.018 PMID: 17998141
11. Tappen M.F., Freeman W.T. & Adelson E.H. Recovering Intrinsic Images from a Single Image. IEEE PAMI 2005, 27(9):1459–1472. https://doi.org/10.1109/TPAMI.2005.185 PMID: 16173188
12. Faul Franz & Ekroll Vebjørn. On the filter approach to perceptual transparency. Journal of vision 2011, 11. https://doi.org/10.1167/11.7.7 PMID: 21659428
13. Motoyoshi I., Nishida S., Sharan L., & Adelson E.H. Image statistics and the perception of surface qualities. Nature 2007, 447:206–209. https://doi.org/10.1038/nature05724 PMID: 17443193
14. du Buf J.M.H. and Fischer S. Modeling brightness perception and syntactical image coding. Optical Engineering 1995, 34(7):1900–1911. https://doi.org/10.1117/12.200602
15. Kingdom F.A.A. & Moulden B. A multi-channel approach to brightness coding. Vision Research 1992, 32(8):1565–1582. https://doi.org/10.1016/0042-6989(92)90212-2 PMID: 1455729
16. Morrone MC, Owens RA. Feature detection from local energy. Pattern Recognit. Lett. 6 (1987): 303–313. https://doi.org/10.1016/0167-8655(87)90013-4
17. Morrone MC, Burr DC. Feature detection in human vision: a phase-dependent energy model. Proc R Soc Lond B Biol Sci. 1988; 235(1280):221–245. https://doi.org/10.1098/rspb.1988.0073 PMID: 2907382
18. Watt R.J. & Morgan M. A theory of the primitive spatial code in human vision. Vision Research 1985, 25:1661–1674. https://doi.org/10.1016/0042-6989(85)90138-5 PMID: 3832590
19. Blakeslee B, McCourt ME. A multiscale spatial filtering account of the White effect, simultaneous brightness contrast and grating induction. Vision Research. 1999; Volume 39, Issue 26, Pages 4361–4377, ISSN 0042–6989, https://doi.org/10.1016/S0042-6989(99)00119-4 PMID: 10789430
20. Blakeslee B, McCourt ME. A unified theory of brightness contrast and assimilation incorporating oriented multiscale spatial filtering and contrast normalization. Vision Research. 2004; Volume 44, Issue 21, Pages 2483–2503, ISSN 0042-6989, https://doi.org/10.1016/j.visres.2004.05.015 PMID: 15358084
21. Dakin S.C. & Bex P.J. Natural image statistics mediate brightness'filling in'. Proceedings of the Royal Society B 2003, 270(1531):2341–2348. https://doi.org/10.1098/rspb.2003.2528 PMID: 14667349
22. Otazu X, Vanrell M, Alejandro Párraga C. Multiresolution wavelet framework models brightness induction effects. Vision Research. 2008, Volume 48, Issue 5, Pages 733–751, ISSN 0042-6989, https://doi.org/10.1016/j.visres.2007.12.008 PMID: 18241909
23. DeValois RL, DeValois KK. Spatial Vision. Annual Review of Psychology. 1988. 31(1) 309–341. https://doi.org/10.1146/annurev.ps.31.020180.001521

24. Betz T, Shapley R, Wichmann FA, Maertens M. Noise masking of White's illusion exposes the weakness of current spatial filtering models of lightness perception. Journal of Vision. 2015; 15(14):1. http://jov.arvojournals.org/article.aspx?doi=10.1167/15.14.1 PMID: 26426914

25. Berthold Horn. Determining lightness from an image. Computer Graphics and Image Processing. 1974 3; 277–299. https://doi.org/10.1016/0146-664X(74)90022-7

26. Land Edwin & McCann John. Lightness and retinex theory. Journal of the Optical Society of America. 1971; 61. 1–11. https://doi.org/10.1364/JOSA.61.000001 PMID: 5541571

27. McCann J.J. Retinex at 50: color theory and spatial algorithms, a review. Journal of Electronic Imaging 2017, 26(3):031204 https://doi.org/10.1117/1.JEI.26.3.031204

28. Cohen MA, Grossberg S. Neural dynamics of brightness perception: Features, boundaries, diffusion, and resonance. Advances in Psychology. 1987; Volume 43,Pages 212–270, ISSN 0166-4115, ISBN 9780444701183, https://doi.org/10.1016/S0166-4115(08)61762-8

29. Gerrits HJM, Vendrik AJH. Simultaneous contrast, filling-in process and information processing in man's visual system. Exp Brain Res. 1970; 11(4):411–430. https://doi.org/10.1007/BF00237914 PMID: 5496938

30. Pessoa L, Thompson E, Noe A. Finding out about filling-in: A guide to perceptual completion for visual science and the philosophy of perception. Behav Brain Sci. 1998; 21(6):723–802. https://doi.org/10.1017/S0140525X98001757 PMID: 10191878

31. Komatsu H. The neural mechanisms of perceptual filling-in. Vol. 7, Nature Reviews Neuroscience. Nature Publishing Group; 2006. p. 220–31. https://doi.org/10.1038/nrn1869 PMID: 16495943

32. Huang X. & Paradiso M.A. V1 response timing and surface filling-in. Journal of Neurophysiology 2008, 100(1):539–547. https://doi.org/10.1152/jn.00997.2007 PMID: 18509081

33. Hong S.W. & Tong F. Neural representation of form-contingent color filling-in in the early visual cortex. Journal of Vision 2017, 17(13):10. https://doi.org/10.1167/17.13.10 PMID: 29136409

34. Grossberg S. Cortical dynamics of three-dimensional form, color, and brightness perception: I. Monocular theory. Percept Psychophys. 1987 Mar; 41(2):87–116. https://doi.org/10.3758/BF03204875 PMID: 3822755

35. Keil MS, Cristóbal G, Hansen T, Neumann H. Recovering real-world images from single-scale boundaries with a novel filling-in architecture. Neural Netw. 2005; 18(10):1319–1331. https://doi.org/10.1016/j.neunet.2005.04.003 PMID: 16039097

36. Keil MS. From Neuronal Models to Neuronal Dynamics and Image Processing. Biologically inspired Computer Vision: Fundamentals and Applications. wiley; 2015. p. 221–44. Available from: http://arxiv.org/abs/1801.08585.

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

38. Sepp W & Neumann H. A multi-resolution filling-in model for brightness perception. 9th International Conference on Artificial Neural Networks 1999: ICANN'99, Edinburgh, GB, Conference Publication 470, Proceedings of the ICANN Volume 1, pp. 461–466, https://doi.org/10.1049/cp:19991152.

39. Grossberg S. Cortical Dynamics of Figure-Ground Separation in Response to 2D Pictures and 3D Scenes: How V2 Combines Border Ownership, Stereoscopic Cues, and Gestalt Grouping Rules. Frontiers in Psychology 2016, 6:2054.

40. Grossberg S, Todorovic D. Neural dynamics of 1-D and 2-D brightness perception: A unified model of classical and recent phenomena. Percept Psychophys. 1988 May; 43(3):241–77. https://doi.org/10.3758/BF03207869 PMID: 3347487

41. Grossberg S. 3-D vision and figure-ground separation by visual cortex. Percept Psychophys. 1994; 55 (1):48–121. https://doi.org/10.3758/BF03206880 PMID: 8036093

42. Chen S and Glasauer S and Mueller HJ and Conci M. Surface filling-in and contour interpolation contribute independently to Kanizsa figure formation. Journal of Experimental Psychology. Human Perception and Performance 2018, 44(9):1399–1413.

43. Breitmeyer B. G., Kafaligönül H., Öğmen H., Mardon L., Todd S., & Ziegler R. Meta-and paracontrast reveal differences between contour-and brightness-processing mechanisms. Vision Research 2006, 46(17), 2645–2658.

44. Hung Chou P., Ramsden Benjamin M., Chen Li Min and Roe Anna Wang. Building surfaces from borders in Areas 17 and 18 of the cat. Vision Research 2001, 41(10–11):1389–1407.

45. Lamme V.A.F., Rodriguez-Rodriguez V., Spekreijse H. Separate Processing Dynamics for Texture Elements, Boundaries and Surfaces in Primary Visual Cortex of the Macaque Monkey. Cerebral Cortex 1999, 9:406–413.

46. Diane C., Rogers-Ramachandran D.C. & Ramachandran V.S. Psychophysical evidence for boundary and surface systems in human vision. Vision Research 1998, 38:71–77. https://doi.org/10.1016/S0042-6989(97)00131-4

47. Domijan D. A Neurocomputational account of the role of contour facilitation in brightness perception. Front Hum Neurosci. 2015; 9:93. Published 2015 Feb 19. https://doi.org/10.3389/fnhum.2015.00093 PMID: 25745396

48. Neumann H. Mechanisms of neural architecture for visual contrast and brightness perception. Neural Netw. 1996; 9(6):921–936. https://doi.org/10.1016/0893-6080(96)00023-8 PMID: 12662572

49. Ross WD, Pessoa L. Lightness from contrast: A selective integration model. Percept Psychophys. 2000; 62(6):1160–81. https://doi.org/10.3758/BF03212120 PMID: 11019614

50. Yazdanbakhsh A, Arabzadeh E, Babadi B, Fazl A. Munker-White-like illusions without T-junctions. Perception. 2002; 31(6):711–5. https://doi.org/10.1068/p3348 PMID: 12092797

51. Howe PD. White's effect: Removing the junctions but preserving the strength of the illusion. Perception. 2005; 34(5):557–564. https://doi.org/10.1068/p5414 PMID: 15991692

52. Bressan P. Explaining lightness illusions. [published correction appears in Perception 2002;31 (4):516]. Perception. 2001; 30(9):1031–1046. https://doi.org/10.1068/p3109 PMID: 11694081

53. Todorović D. Lightness and junctions. Perception. 1997; 26(4):379–394. https://doi.org/10.1068/p260379 PMID: 9404489

54. Barkan Y, Spitzer H, Einav S. Brightness contrast-contrast induction model predicts assimilation and inverted assimilation effects. Journal of Vision 2008; 8(7):27. https://doi.org/10.1167/8.7.27 PMID: 19146233

55. Keil MS, Cristóbal G, Neumann H. Gradient representation and perception in the early visual system-A novel account of Mach band formation. Vision Res. 2006; 46(17):2659–2674. https://doi.org/10.1016/j.visres.2006.01.038 PMID: 16603218

56. Yang Z, Purves D. The statistical structure of natural light patterns determines perceived light intensity. Proc Natl Acad Sci. 2004 Jun 8; 101(23):8745–50. Available from: https://www.pnas.org/content/101/23/8745. https://doi.org/10.1073/pnas.0402192101 PMID: 15152077

57. Purves D, Williams SM, Nundy S, Lotto RB. Perceiving the Intensity of Light. Psychol Rev. 2004 Jan; 111(1):142–58. Available from: http://www.ncbi.nlm.nih.gov/pubmed/14756591

58. Corney D, Lotto RB. What Are Lightness Illusions and Why Do We See Them? PLoS Comput Biol. 2007; 3(9):1790–1800. https://doi.org/10.1371/journal.pcbi.0030180 PMID: 17907795

59. Morgenstern Y, Rukmini DV., Monson BB, Purves D. Properties of artificial neurons that report lightness based on accumulated experience with luminance. Front Comput Neurosci. 2014; 8:134. Published 2014 Nov 3. https://doi.org/10.3389/fncom.2014.00134 PMID: 25404912

60. Vincent BT, Baddeley RJ, Troscianko T, Gilchrist ID. Is the early visual system optimised to be energy efficient? In: Network: Computation in Neural Systems. 2005; 16. 175–90. https://doi.org/10.1080/09548980500290047 PMID: 16411495

61. Levy William & Baxter Robert. Energy Efficient Neural Codes. Neural Comput. 1996; 8. 531–43. https://doi.org/10.1162/neco.1996.8.3.531 PMID: 8868566

62. Barlow HB. Possible principles underlying the transformation of sensory messages. Sensory Communication. 1961; 1.

63. Srinivasan MV., Laughlin SB, Dubs A. Predictive coding: A fresh view of inhibition in the retina. Proc R Soc Lond B Biol Sci. 1982; 216(1205):427–459. https://doi.org/10.1098/rspb.1982.0085

64. Atick JJ, Redlich AN. What Does the Retina Know about Natural Scenes? Neural Comput. 1992; vol. 4, no. 2, pp. 196–210. https://doi.org/10.1162/neco.1992.4.2.196

65. Keil MS. Does face image statistics predict a preferred spatial frequency for human face processing? Proc Biol Sci. 2008; 275(1647):2095–2100. https://doi.org/10.1098/rspb.2008.0486 PMID: 18544506

66. White M. The effect of the nature of the surround on the perceived lightness of grey bars within square wave test gratings. Perception. 1981; 10, no. 2 215–30. https://doi.org/10.1068/p100215 PMID: 7279550

67. Jones JP, Palmer LA. An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex. J Neurophysiol. 1987; 58(6):1233–1258.

68. Hubel DH, Wiesel TN. Receptive fields of single neurones in the cat's striate cortex. Physiol. 1959; 148(3):574–591. https://doi.org/10.1113/jphysiol.1959.sp006308 PMID: 14403679

69. Daugman JG. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. J Opt Soc Am A. 1985; 2(7):1160–1169. https://doi.org/10.1364/JOSAA.2.001160 PMID: 4020513

70. Marcelja S. Mathematical description of the responses of simple cortical cells. Journal of the Optical Society of America. 1980; 70. 1297–300. https://doi.org/10.1364/JOSA.70.001297 PMID: 7463179

71. Komatsu H, Murakami I, Kinoshita M. Surface representation in the visual system. Brain Res Cogn Brain Res. 1996; 5(1–2):97–104. https://doi.org/10.1016/S0926-6410(96)00045-6 PMID: 9049075

72. Rossi AF, Rittenhouse CD, Paradiso MA. The representation of brightness in primary visual cortex. Science. 1996; 273(5278):1104–1107. https://doi.org/10.1126/science.273.5278.1104 PMID: 8688096

73. Dai J, Wang Y. Representation of surface luminance and contrast in primary visual cortex. Cereb Cortex. 2012; 22(4):776–787. https://doi.org/10.1093/cercor/bhr133 PMID: 21693782

74. Hubel DH, Wiesel TN. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. J Physiol. 1962; 160(1):106–154. https://doi.org/10.1113/jphysiol.1962.sp006837 PMID: 14449617

75. De Valois RL, Albrecht DG, Thorell LG. Spatial frequency selectivity of cells in macaque visual cortex. Vision Res. 1982; 22(5):545–59. https://doi.org/10.1016/0042-6989(82)90113-4 PMID: 7112954

76. Adelson EH, Bergen JR. Spatiotemporal energy models for the perception of motion. J. Opt. Soc. Am. 1985; A 2, 284–299. https://doi.org/10.1364/JOSAA.2.000284 PMID: 3973762

77. Pollen DA, Ronner SF. Visual Cortical Neurons as Localized Spatial Frequency Filters. IEEE Transactions on Systems, Man, & Cybernetics. 1983; 13(5), 907–916. https://doi.org/10.1109/TSMC.1983.6313086

78. Bell AJ, Sejnowski TJ. The "independent components" of natural scenes are edge filters. Vision Res. 1997; 37(23):3327–3338. https://doi.org/10.1016/S0042-6989(97)00121-1 PMID: 9425547

79. Kessy A., Lewin A. & Strimme K. Optimal Whitening and Decorrelation. The American Statistician 2018, 72(4):309–314 https://doi.org/10.1080/00031305.2016.1277159.

80. Khintchine A. Korrelationstheorie der stationären stochastischen Prozesse. Math Ann. 1934; Dec, 109(1):604–15. Available from: http://link.springer.com/10.1007/BF01449156.

81. Wiener N. Generalized harmonic analysis. Acta Math. 1930; 55(0):117–258. Available from: http://projecteuclid.org/euclid.acta/1485887877. https://doi.org/10.1007/BF02546511

82. Logvinenko AD, Kane J. Hering's and Helmholtz's types of simultaneous lightness contrast. J Vis. 2004; 4(12):1102–1110. Published 2004 Dec 24. https://doi.org/10.1167/4.12.9 PMID: 15669914

83. Blakeslee B, McCourt ME. Similar mechanisms underlie simultaneous brightness contrast and grating induction. Vision Res. 1997; Volume 37, Issue 20, Pages 2849–2869, ISSN 0042-6989, https://doi.org/10.1016/S0042-6989(97)00086-2 PMID: 9415365

84. Diamond AL. Foveal simultaneous brightness contrast as a function of inducing, and test-field luminances. J Exp Psychol. 1953; 45(5):304–314. https://doi.org/10.1037/h0060230 PMID: 13052866

85. Kitterle FL. The effects of simultaneous and successive contrast on perceived brightness. Vision Res. 1972; 12(11):1923–1931. https://doi.org/10.1016/0042-6989(72)90080-6 PMID: 5079257

86. Stevens JC. Brightness inhibition re size of surround. Percept & Psychophys. 1967; 189–192. https://doi.org/10.3758/BF03213048

87. William Yund E, Armington JC. Color and brightness contrast effects as a function of spatial variables. Vision research. 1975; 15. 917–29. https://doi.org/10.1016/0042-6989(75)90231-X

88. Shi V, Cui J, Troncoso XG, Macknik SL, Martinez-Conde S. Effect of stimulus width on simultaneous contrast. PeerJ. 2013; 1:e146. Published 2013 Sep 5. https://doi.org/10.7717/peerj.146 PMID: 24032092

89. Benary W. Beobachtungen zu einem Experiment über Helligkeitskontrast. Psychol Forsch. 1924; 5, 131–142. https://doi.org/10.1007/BF00402398

90. Salmela VR, Laurinen PI. Low-level features determine brightness in White's and Benary's illusions. Vision Res. 2009 Apr 29; 49(7):682–90. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19200439. https://doi.org/10.1016/j.visres.2009.01.006

91. Gilchrist A.L. & Annan V. Articulation effects in lightness: Historical background and theoretical implications. Perception 2002, 31:141–150. https://doi.org/10.1068/p04sp PMID: 11922128

92. Economou E, Zdravković S, Gilchrist A. Grouping Factors and the Reverse Contrast Illusion. Perception. 2015; 44(12):1383–1399. https://doi.org/10.1177/0301006615607118 PMID: 26562863

93. Taya R, Ehrenstein WH, Cavonius CR. Varying the strength of the Munker-White effect by stereoscopic viewing. Perception. 1995; 24(6):685–694. https://doi.org/10.1068/p240685 PMID: 7478908

94. Ripamonti C, Gerbino W. Classical and inverted White's effects. Perception. 2001; 30(4):467–488. https://doi.org/10.1068/p3108 PMID: 11383193

95. Kingdom F, Moulden B. White's effect and assimilation. Vision Res. 1991; 31(1):151–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/2006548. https://doi.org/10.1016/0042-6989(91)90082-G

96. Betz T, Shapley R, Wichmann FA, Maertens M. Testing the role of luminance edges in White's illusion with contour adaptation. J Vis. 2015 Aug 1; 15(11):14. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26305862. https://doi.org/10.1167/15.11.14

97. Güçlü B, Farell B. Influence of target size and luminance on the White-Todorović effect. Vision Res. 2005; 45(9):1165–1176. https://doi.org/10.1016/j.visres.2004.10.025 PMID: 15707925

98. Hong SW, Shevell SK. Brightness contrast and assimilation from patterned inducing backgrounds. Vision Res. 2004; Volume 44, Issue 1, Pages 35–43, ISSN 0042-6989, https://doi.org/10.1016/j.visres.2003.07.010

99. Baumgartner G. Indirekte Größenbestimmung der rezeptiven Felder der Retina beim Menschen mittels der Hermannschen Gittertäuschung. Pflugers Arch Gesamte Physiol Menschen Tiere. 1960;. https://doi.org/10.1007/BF00680926

100. Schiller PH, Carvey CE. The Hermann grid illusion revisited. Perception. 2005; 34(11):1375–1397. https://doi.org/10.1068/p5447 PMID: 16355743

101. Keil MS. Smooth Gradient Representations as a Unifying Account of Chevreul's Illusion, Mach Bands, and a Variant of the Ehrenstein Disk. Neural Comput. 2006 Feb 12; 18(4):871–903. https://doi.org/10.1162/neco.2006.18.4.871 PMID: 16494694

102. Mach E. Über die Wirkung der räumlichen Verteilung des Lichtreizes auf die Netzhaut [On the effect of the spatial distribution of the light stimulus on the retina]. Sitzungsberichte der mathematisch-naturwissenschaftlichen Classe der kaiserlichen Akademie der Wissenschaften 1865, 52:303–322.

103. Ratliff F. Mach bands: quantitative studies on neural networks in the retina. Holden-Day, INC. 1965.

104. Ross J, Concetta Morrone M, Burr DC. The conditions under which Mach bands are visible. Vision Res. 1989 Jan 1; 29(6):699–715. Available from: https://www.sciencedirect.com/science/article/pii/0042698989900333. https://doi.org/10.1016/0042-6989(89)90033-3 PMID: 2626828

105. Pessoa L. Mach bands: how many models are possible? Recent experimental findings and modeling attempts. Vision Res. 1996 Oct; 36(19):3205–27. Available from: http://www.ncbi.nlm.nih.gov/pubmed/8917780. https://doi.org/10.1016/0042-6989(95)00341-X

106. Kingdom FAA. Mach bands explained by response normalization. Frontiers in human neuroscience. 2014; 8. 843. https://doi.org/10.3389/fnhum.2014.00843 PMID: 25408643

107. Ratliff F, Milkman N, Rennert N. Attenuation of Mach bands by adjacent stimuli. Proc Natl Acad Sci U S A. 1983 Jul; 80(14):4554–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/6576350. https://doi.org/10.1073/pnas.80.14.4554

108. McCourt ME. A spatial frequency dependent grating-induction effect. Vision Res. 1982; 22(1):119–134. https://doi.org/10.1016/0042-6989(82)90173-0 PMID: 7101736

109. Keil MS. Neural architectures for unifying brightness perception and image processing [dissetation] Universität Ulm; 2003. Available from: https://www.researchgate.net/publication/35660056_Neural_architectures_for_unifying_brightness_perception_and_image_processing.

110. Lerer A, Supèr H, Keil MS. Luminance gradients and non-gradients as a cue for distinguishing reflectance and illumination in achromatic images: A computational approach. Neural Networks. 2019; 110. https://doi.org/10.1016/j.neunet.2018.11.001 PMID: 30496916

111. Hearn T.A., Reichel L. Application of denoising methods to regularization of ill-posed problems. 2014, Numer Algor 66, 761–777. https://doi.org/10.1007/s11075-013-9760-5

112. Simoncelli EP, Olshausen BA. Natural Image Statistics and Neural Representation. Annu Rev Neurosci. 2001; 24:1193–1216. https://doi.org/10.1146/annurev.neuro.24.1.1193 PMID: 11520932

113. Bressan P, Actis-Grosso R. Simultaneous Lightness Contrast on Plain and Articulated Surrounds. Perception. 2006 Apr 25; 35(4):445–52. Available from: http://journals.sagepub.com/doi/10.1068/p5247. https://doi.org/10.1068/p5247 PMID: 16700287

114. Gilchrist A, Kossyfidis C, Agostini T, Li X, Bonato F, Cataliotti J, et al. An anchoring theory of lightness perception. Psychological Review. 1999. 106(4), 795–834. https://doi.org/10.1037/0033-295X.106.4.795 PMID: 10560329

Chapter 4

Global results

Chapter abstract: This chapter describes the overall results of the present study. Subchapters 4.1 and 4.2 summarize the most significant results presented in Chapter 2 and Chapter 3, respectively. Subchapter 4.3, as a proof of concept, describes an extended model version that unifies the hypothetical mechanisms presented in Chapter 2 and Chapter 3; it also includes some simulations that combine both mechanisms in predicting (or explaining) brightness and lightness phenomena.

4.1 Summary First Publication

This study builds on the idea that reflectance changes are often associated with fast luminance variations (e.g., edges) and that shading is often associated with slow variations (.e.g., gradients (Keil et al., 2006; Keil, 2006). In achromatic images, we hypothesize that the visual system could exploit this cue (fast versus slow luminance variations) in order to estimate the apparent reflectance of a surface (i.e., lightness) and the apparent shading (or illumination) by segregating low vs. high neural variability responses of encoded visual input. Then, the final estimation of apparent reflectance and perceived illumination (shading) relies on the segregation and interaction between non-gradients and gradients.

To test the above idea, we proposed a computational approach. Initially, the model encodes an achromatic visual image using a set of filters. Further, since the luminance spatial frequency difference, in general, facilitate the segmentation of superimposed textures and perhaps gradients and non-gradients cues, we first learned two classes of filter kernels from natural (or real-world) gray-level images constrained to low spatial frequencies (LSF) and high spatial frequencies (HSF). The filter kernels served to eventually segregate gradients from non-gradients through a gating map based on a response variability mechanism. The gating map was computed from the variability over a set of filter kernels. The high variability estimates non-gradients, while the low variability set estimates smooth luminance gradients. Finally, in order to reconstruct both segregated representations (gradient and non-gradient layers), we implemented an optimization method based on conjugate gradient descent; it served to recover both layers through an iterative procedure that resembles a filling-in process (Gerrits & Vendrik 1970; Grossberg 1988, Rossi 1996, Komatsu 2006).

Computationally, our model successfully segregated the luminance of any achromatic input image (e.g., real-world images) into two layers. The segregated layers correlated with various perceptual attributes as lightness and perceived illumination (e.g., shadings) and successfully predicted lightness phenomena of visual illusions such as Adelsons's checker-shadow display or grating induction. More specifically, the non-gradient layer captured the estimated lightness. In particular, the direction - increasing or decreasing - of the estimated non-gradients was "balanced" in the opposite direction of the luminance gradients according to our initial hypothesis. Figure 4.1 shows an example of commented

results.

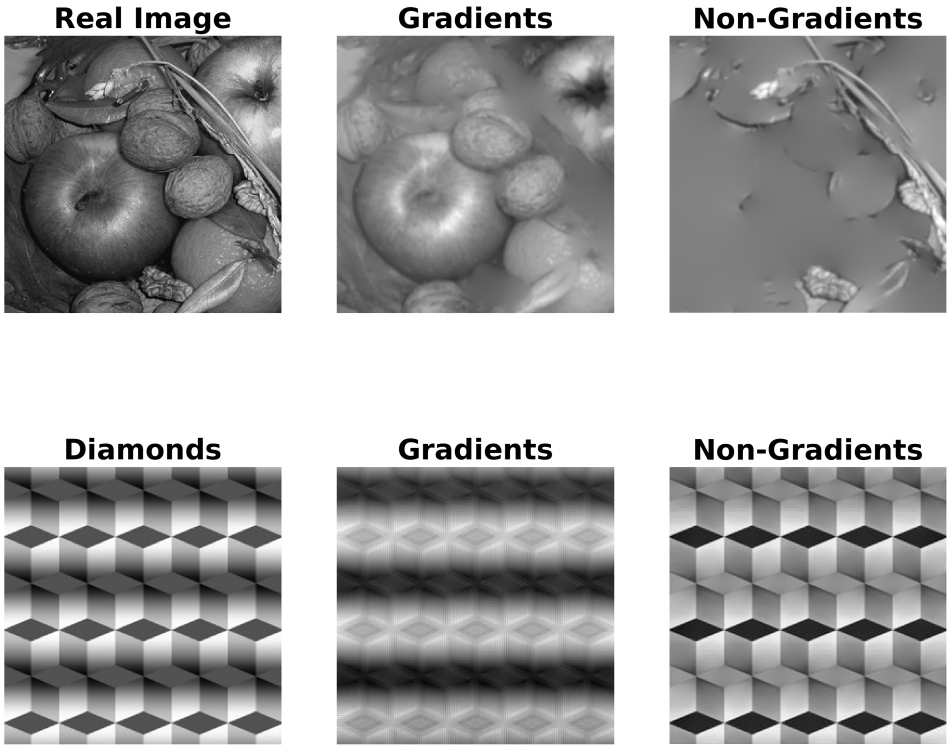


Figure 4.1: Example of segregation into gradients and non-gradients of Real Image and Diamond's Illusion. Although in Diamond's Illusion, all diamonds have identical intensity values, we perceive them as light-gray and dark-gray, respectively, dependent on the amplitude of the superimposed sine-wave grating. The image corresponding to Real Image was extracted from https://search.creativecommons.org/ with license CC BY.

Regarding the segregation mechanism, we identified the variability response and not the choice of a specific spatial frequency as a vital step to luminance's segregation. Indeed, the segregation of gradient and non-gradient we noted could not be only based on spatial frequencies (e.i., lowpass and highpass filtering) since the gradient and non-gradient tend to be mixed across the spatial scale (see Figure 4.2). The gradient layer captured only slow (or smooth) variations in luminance suppressed in the non-gradient layer. In contrast, non-gradient contained only the abrupt luminance changes that often

correspond to changes in reflectance. In addition, we also identified a plausible functional role in the shape of the learned filter kernels in computing (reconstruction) both layers: Odd symmetrical filters responded the best to smooth luminance gradients being LSF odd the set of filters that propagated activity across greater distances per unit time (iteration) than their HSF odd counterparts. In contrast, the even symmetrical filters (independent of the frequency range, i.e., LSF and HSF) responded poorly to gradient features.

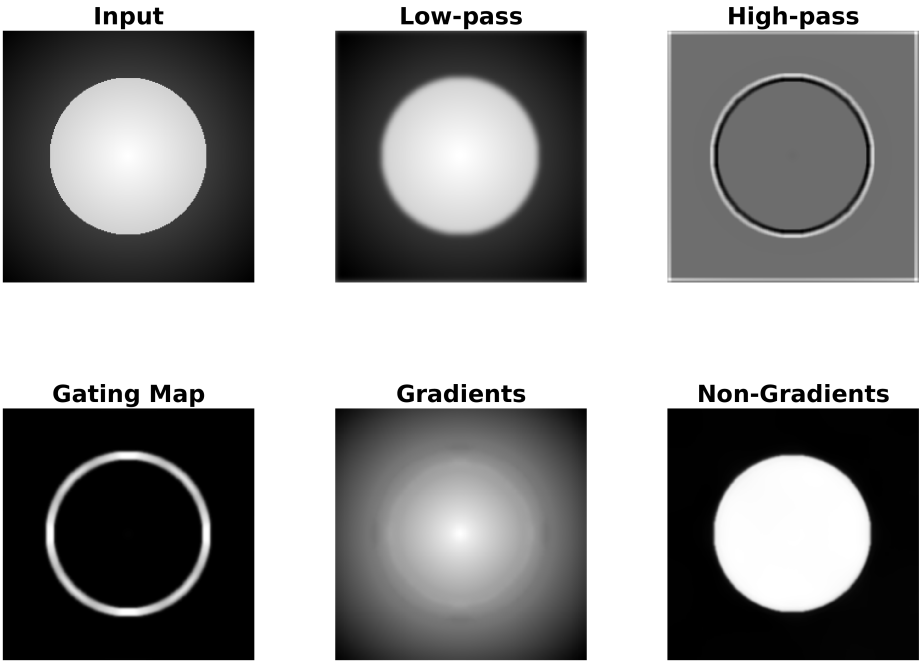


Figure 4.2: Illustration of segregating an input image into gradients and non-gradients. *First row:* A homogeneous disk with an overlaid gradient corresponds to the Input image with its corresponding lowpass and high-pass filtering. *Second row:* The corresponding gating map based on a response variability computed with the actual model. The result of segregating the input image into Gradients and Non-Gradients. Notice gradient activity is not just equal to lowpass information, and similarly, non-gradient information is not the same as high-pass filtering of the input image.

4.2 Summary Second Publication

Here we proposed a novel computational model for estimating the brightness/lightness of many visual illusions. Context sensitivity inspired the proposal of many neural mechanisms for explaining the perception of luminance (brightness). In this approach, we hypothesize that many aspects of brightness can be explained by a dynamic filtering process that reduces context redundancy edge representations on the one hand, while non-redundant context activity is enhanced on the other. The dynamic filter is learned for each input image and implements context sensitivity. Dynamic filtering is applied to the responses of (model) complex cells in order to build a gain control map. The gain control map then acts on simple cell responses before they are used to create a brightness map via activity propagation.

To test the above idea, we developed a computational model in which a dynamic filter carries out equalization. In the first step, an input image is encoded by two sets of Gabor filters, which mimic the spatial response properties of simple cells in V1 (Jones & Palmer, 1987a). The high-resolution filters' responses encoded the contrast-only information, while the more coarse-grained filters encoded the contrast-luminance information. From contrast-only information filters, we computed boundary activity via local energy (Morrone & Burr, 1988; Morrone & Owens, 1987), which is insensitive to the phase of simple cells and resembles complex cell responses. A decorrelation kernel (dynamic filtering) was then learned from the local energy map and applied to the latter to reduce (equalize) its redundancy. The redundancy-reduced energy map then functions as a 'Gain control map' for both sets of filters, which served to recover a brightness map through an iterative procedure that resembles a filling-in process (Gerrits & Vendrik 1970; Grossberg 1988, Rossi 1996, Komatsu 2006).

Without changing any of our model's parameter values, the model successfully predicted different "styles" of brightness phenomena for a great number of visual illusions: Simultaneous Brightness Contrast (SBC), White's Effect, Reverse Contrast, Benary's Cross, Todorovic's illusion (with variations), the Dungeon Illusion, the Checkerboard Illusion, Shevell's Ring, the Craik-O'Brien-Cornsweet effect (COCE), the Hermann/Hering grid, the corrugated grid, Chevreul's illusion (including the luminance pyramid), Grating Induction (GI), and Mach Bands. Figure 4.3 shows some of these predictions. Additionally, we could reproduce the trend of corresponding psychophysical data (SBC, White,

Reverse Contrast, Hermann/Hering grid, Todorovic, GI, and Mach Bands). Furthermore, our model overcame some inconsistencies of traditional explanations for these illusions (e.g., T-junctions, variants of White's illusion). Finally, the model was capable of real-world image processing and handling noisy versions of visual illusion displays with uncorrelated additive noise (see Figure 4.4).

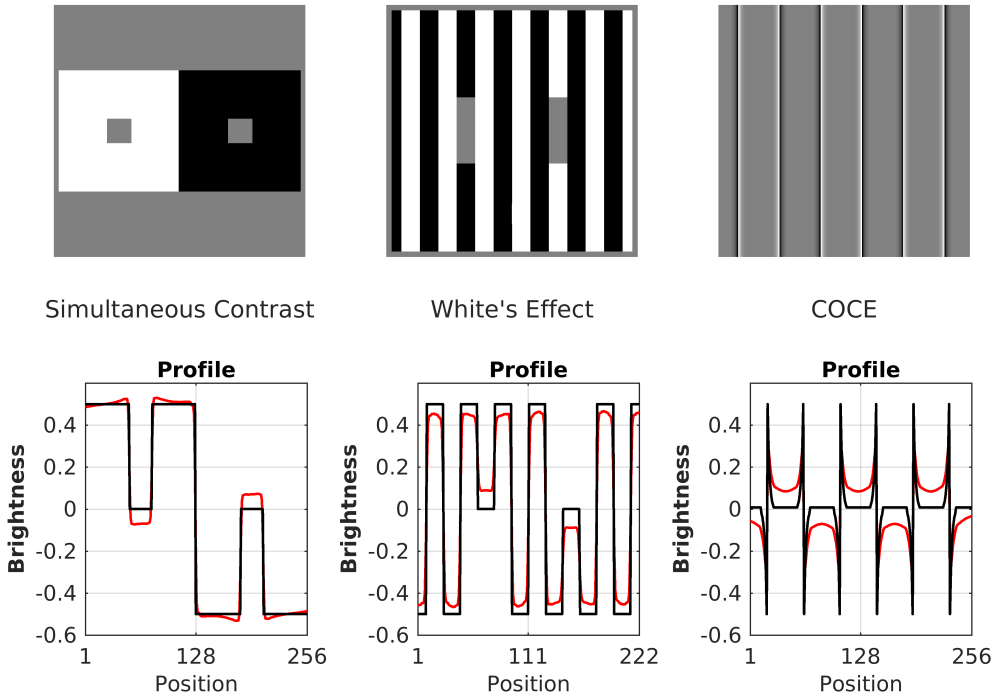


Figure 4.3: Prediction of Simultaneous Contrast, White's illusion and Craik-O'Brien-Cornsweet Effect and their corresponding profiles. Each profile plot shows the predicted brightness (red line) along with input luminance (black line)

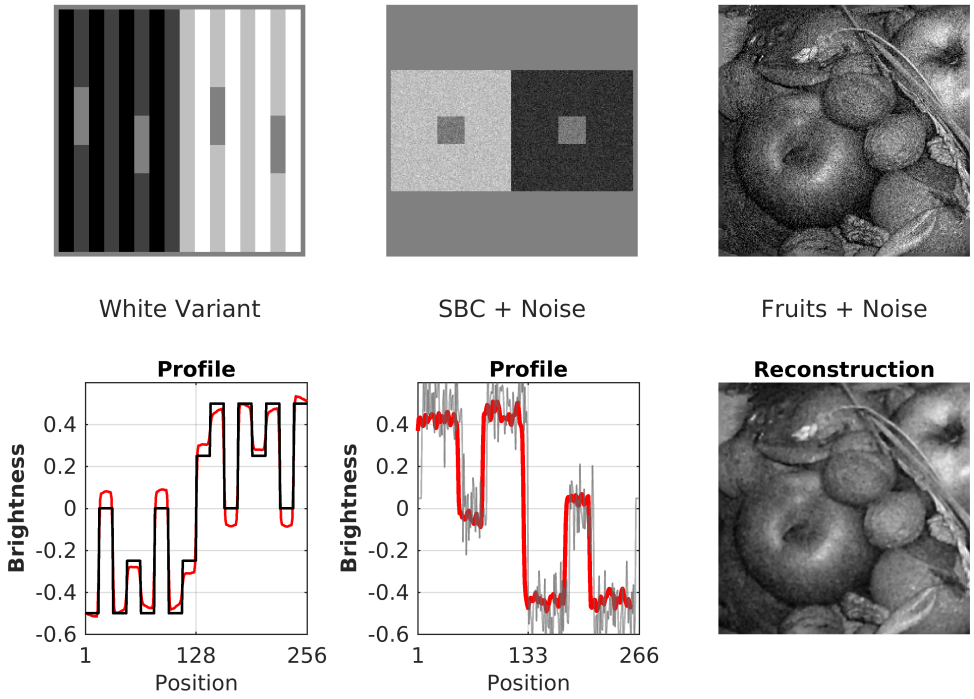


Figure 4.4: Prediction of Variant of White's Effect, Simultaneous Brightness Contrast (SBC) masked with uncorrelated white Gaussian noise, and a real-world image (Bridge). Bottom, their corresponding profiles for Variant of White's Effect, SBC and image reconstruction of real-world image. Each profile plot shows the predicted brightness (red line) and input luminance (black line or gray line). The image corresponding to bridge was extracted from https://search.creativecommons.org/ with license CC BY.

4.3 Unifying Mechanisms: An extension of both models

As a proof of concept of a unifying mechanism, we have developed an extended model unifying the mechanisms of our previous studies (see Chapter 2 and Chapter 3). Figure 4.5 illustrates a scheme of the extended model. Overall, in the first stage, the extended model initially encodes through a set of Gabor Filters the contrast-only and the contrast-luminance information of a given input image similar to the model presented in Chapter 3. Subsequently, the extended model segregates gradient responses and non-gradient responses through a gating map based on a response variability mechanism similar to the model presented in Chapter 2. After that, in the second stage, the extended model

implements dynamic filtering on the gradient and non-gradient responses, respectively. Finally, in the third stage, after dynamic filtering, both segregated representations – gradient and non-gradient responses – served to recover both layers through an iterative procedure that resembles a filling-in process similar to the models presented in Chapter 2 and Chapter 3. Importantly, although the unifying of both mechanisms was computationally feasible, it was necessary to introduce certain modifications in the parameter values (see Appendix A).

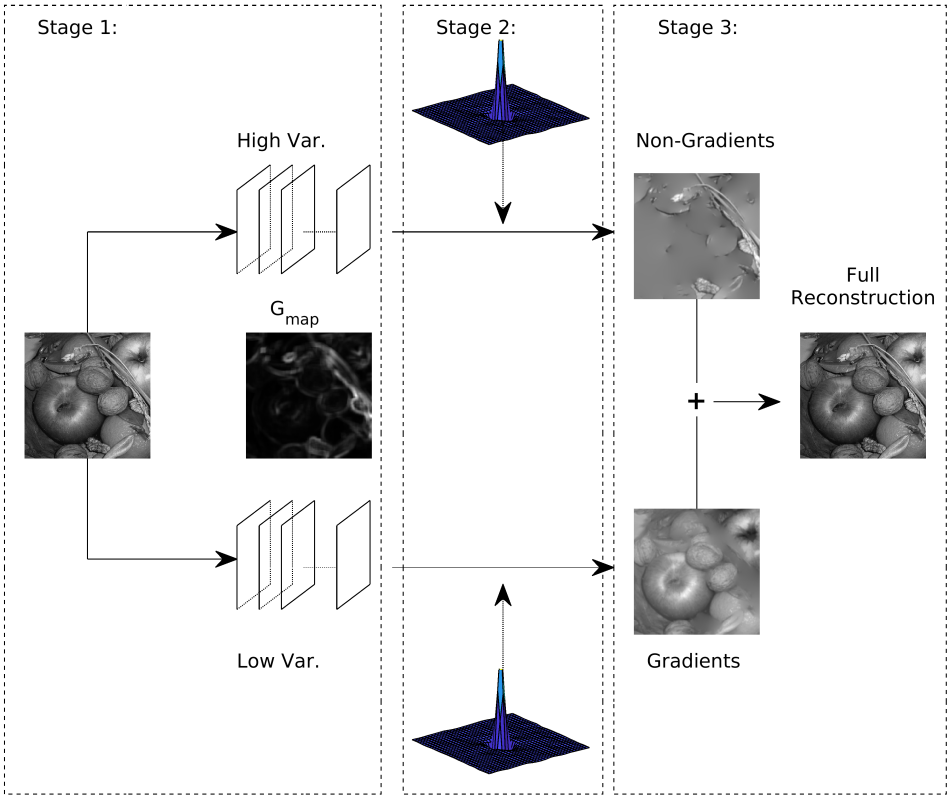


Figure 4.5: Extended Model Scheme. Firstly, in the first stage, the input image is encoded by a set of Gabor filters that mimic the responses of simple cells in V1. Then, in the same way as described in Chapter 2, we used a variability measure in order to generate two sets of maps from the corresponding response maps - one for high variation (i.e., high variability map) and another one for low variation (i.e., low variability map). Subsequently, in the second stage, we implemented dynamic filtering in both low and high variability responses in the same way as described in Chapter 3. Finally, in the third stage, the gradient and non-gradient layers are obtained by solving an inverse problem described in Chapter 2 and Chapter 3. The image example was extracted from https://search.creativecommons.org/ with license CC BY.

We noted two significant variations in the output results concerning previous studies. First, unlike our first study described in Chapter 2, the extended model involves filters responses of contrast-luminance information similar to the second study described in Chapter 3. Consequently, it may produce partial segregation where the non-gradient layer contains residual gradient components and vice versa for certain configurations (see Appendix C). Second, the brightness/lightness phenomena induced by the dynamic filtering were predicted in the full reconstruction (summing both layers), being dominated by the non-gradient layer. Figure 4.6 and Figure 4.7 illustrate simulations to understand these two points better. Notice in the case of luminance step, the response to luminance information led to the apparition of "artificial gradients," while the full reconstruction remains barely changed (see Figure 4.6). In contrast, in the case of luminance ramp (see Figure 4.7), the gradient layer captured almost all the luminance information, whereas the non-gradient layer correlated the surface appearance with three luminance levels. Notice, however, that the full reconstruction still predicts the tendency of brightness phenomena.

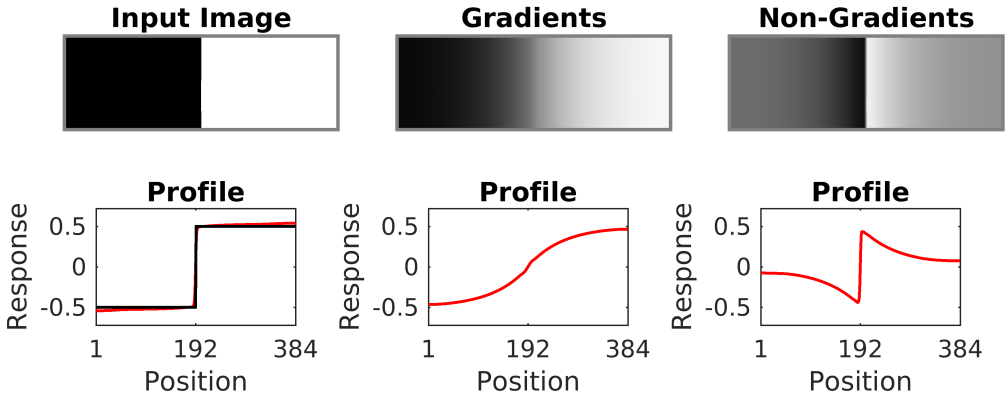


Figure 4.6: Simulations using the extended model for luminance step. A luminance step (no Mach bands are perceived) and corresponding segregation into gradients and non-gradients layers. Under each display, the corresponding profiles (left to right): the input image (black line) with full reconstruction (red line), gradients, and non-gradients.

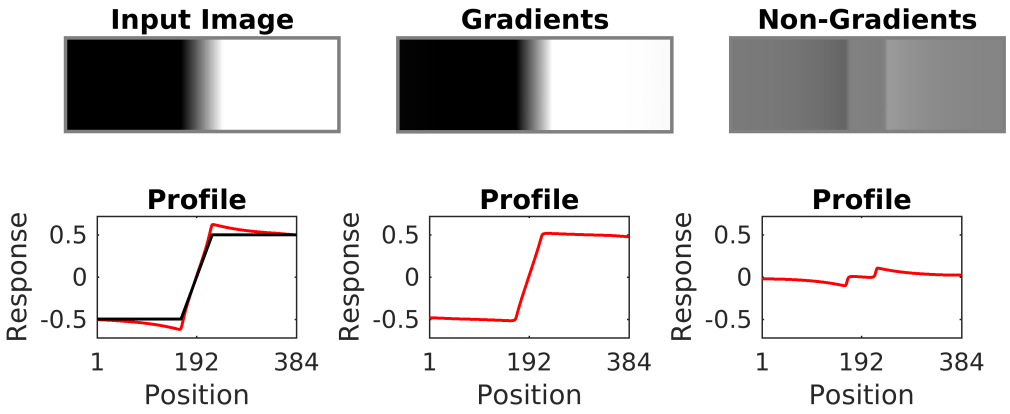


Figure 4.7: Simulations using the extended model for luminance ramp. A luminance ramp leads to the perception of Mach bands close to the knee points of the ramp and its corresponding segregation into gradients and non-gradients layers. Under each display, the corresponding profiles (left to right): the input image (black line) with full reconstruction (red line), gradients, and non-gradients.

4.3.1 Simulation I: Simultaneous Contrast with adjacent luminance gradient

To test our extended model, we focused on a phenomenon of lightness and brightness presented by (Zavango & Daneyko, 2012; Agostini & Galmonte, 2002). They studied how the apparent luminosity of different configurations modified the contrast effect despite the local stimulation remaining identical (see Figure 4.8). The configurations involved luminance gradients (with positive and negative ramps) and solid squares. The luminance gradient originated an appearance of illumination (e.g., halo or glow) despite the absence of physical illuminants, while a sensation of homogeneous illumination for solid squares. The appearance of illumination induced a strong contrast enhancement (or minor reduction) with respect to "original." A possible explanation of these phenomena suggests that luminance gradients are encoded as perceptual features, involving a sensation of self-luminosity, affecting brightness, and it may, in turn, affect lightness computations (Keil, 2008; Correani et al., 2005). Notice that this phenomenon is not necessarily a simultaneous contrast since the tendency remains even if the backgrounds are isolated (see Figure 4.9A).

For the isolated version, our extended model successfully segregated the luminance of

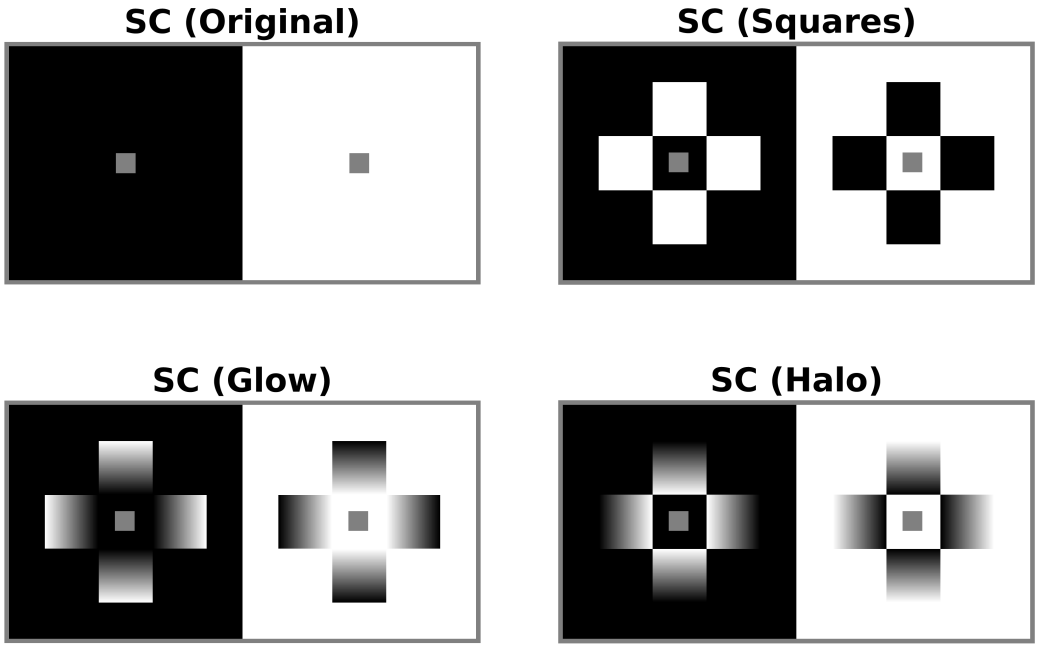


Figure 4.8: Simultaneous Contrast with different adjacent configurations. The contrast effect varies for each configuration even though local stimulation remained identical across displays.

each configuration (see Figure 4.9B and Figure 4.9C). Notice, in the positive and negative ramp configuration, that the gradient layer captured the "illumination effects" as "glow" and "halo." Figure 4.9D shows the tendency of the target test for each configuration in the full reconstruction profiles: We noted that the most intense contrast with respect to the background is produced by positive ramp configuration, followed (in decreasing order) for original configuration, negative ramps configuration, and solid configuration.

This simulation illustrates very well how both mechanisms are combined in the extended model. In the case of solid squares configuration, the contrast response reduction – compared with "the original configuration" – was due to the spatial sensitivity of the dynamic filtering mechanism previously discussed in Chapter 3. Since the solid squares increment the spatial redundancy of edges, the dynamic filtering process led to less excitability on the target test in the square configuration than the original configuration. In the case of positive and negative ramps configurations, the tendency of the appearance of target tests was balanced in the "opposite direction" of the gradient luminance background, similar to results presented in Chapter 2. Consequently, an increment (or reduction) in

the target test of positive (or negative) ramps compared to the original configuration, respectively.

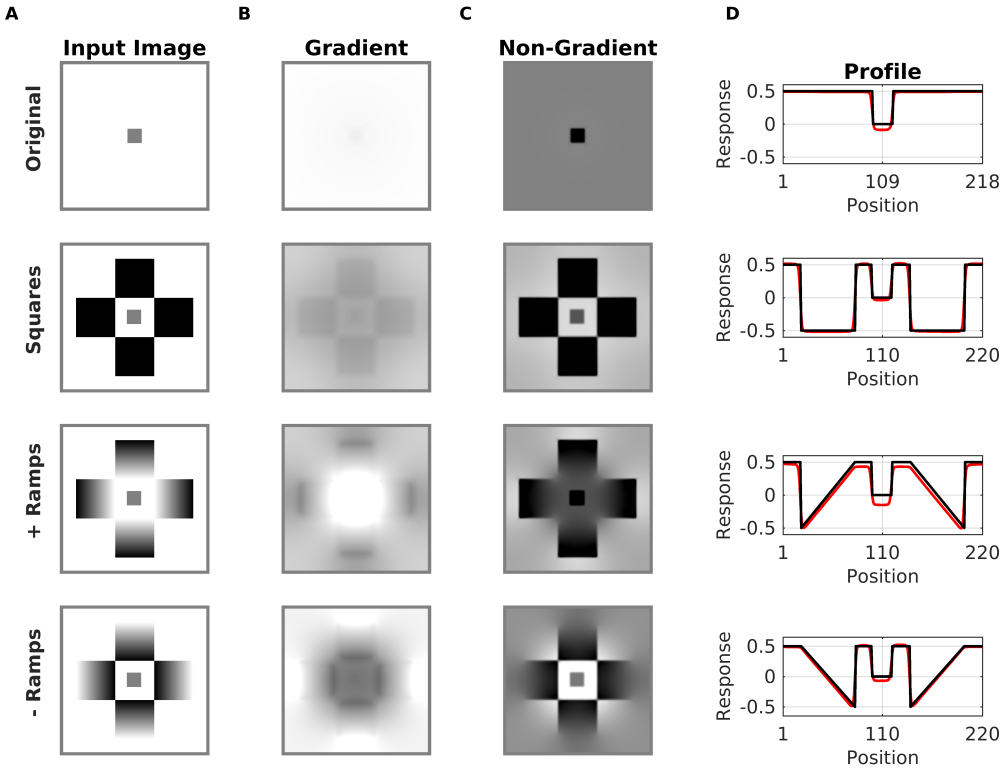


Figure 4.9: Predictions of Simultaneous Contrast with different adjacent configurations. A. First column from top to bottom: Classic configuration (or lightness) contrast; solid squares configuration; squares with positive ramps (halo); and squares with negative ramps (glow). **B.** The corresponding estimated gradient layer of each configuration input image of A. **C.** The corresponding estimated non-gradient layer of each configuration input image of A. **D.** The corresponding profiles of the input image (in black) and the full reconstruction (in red).

4.3.2 Simulation II: Staircase Gelb Effect

Another interesting stimulus to test our extended model is the staircase Gelb effect. Figure 4.10 illustrates the classical and a modified version of the staircase Gelb effect. The classical version of the Gelb effect consists of a row of squares ranging from black to white which is successively placed from left to right on a constant darker space (or background). Initially, the first square appears to be gray or white, and suddenly it appears to be darker when the subsequent whiter (in reflectance) square is placed in

the visual field, inducing a lightness rescaling phenomenon (Cattaliotti & Gilchrist, 1995). As a clarification, do not confuse with lightness compression, which in addition to lightness rescaling, the gamut of perceived values of the presented five squares is compressed relative to the actual values. A modified version is proposed since it is difficult to notice the psychophysical effect of lightness rescaling on a computer screen. In the modified version, the row of squares is successively placed on a gradient luminance background that changes from "higher" to "lower" luminance. Although all squares have identical luminance ranges, the newly introduced square is perceived as brighter than the original square. Each new square that previously appeared brighter (left) now appears a shade of gray upon the addition of a square (right), inducing the lightness rescaling. Notice this phenomenon in G.E version II by comparing the rightest square of the display (n = 3) that is perceived as white while the same square is perceived as gray for display (n = 5).

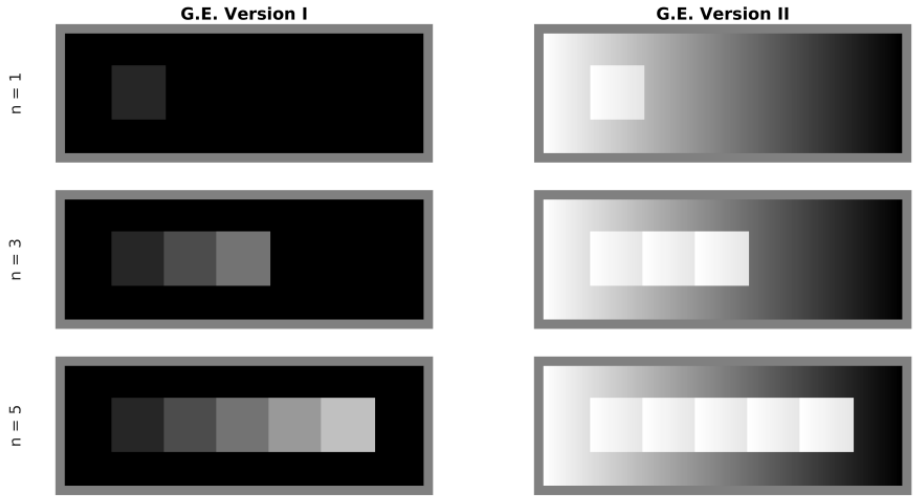


Figure 4.10: Gelb effect. **A.** Computational representation of staircase Gelb effect (G.E version I). The luminance level of the squares (from left to right) are fixed increasingly. **B.** The modified Staircase Gelb effect. All squares have the same luminance ramps. Notice, the ramp induced a slight contrast between adjacent squares.

Our extended model successfully segregated the luminance of both G.E versions (see Figure 4.11). In addition, Figure 4.12 shows, in the full reconstruction, the predicting of the tendency of lightness rescaling of both Gelb's versions: The reduction response

of the first square as a function of subsequent ones. Specifically, the phenomenon was dominated by the non-gradient layer (data not shown). Here, the lightness rescaling could be explained in terms of redundancy reduction by the dynamic filtering mechanism: The greater the number of squares, the greater the redundancy in the edges which separate them, involving progressively more assimilation between the squares and the background as the number of square increase (i.e., a lightness rescaling).

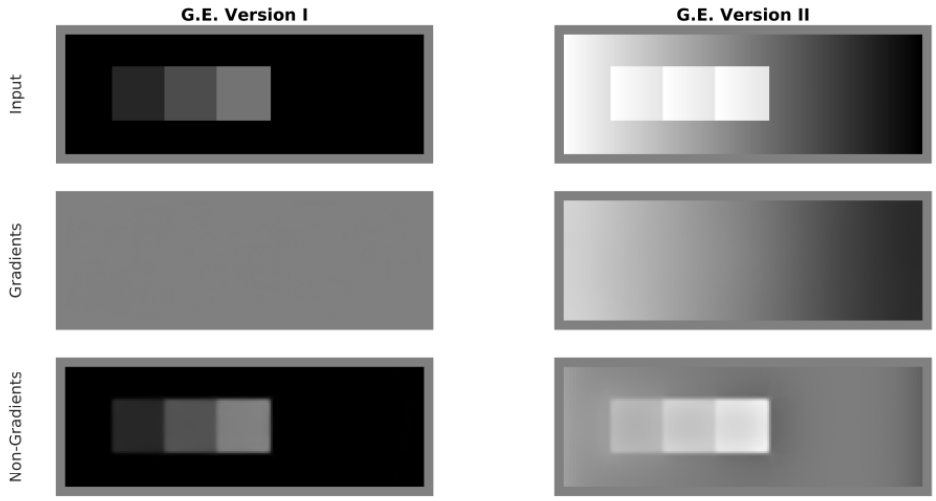


Figure 4.11: Example of segregation of model predictions for G.E version I and G.E version II. Top, input for both versions with exactly (n = 3) successive squares. Middle the corresponding gradient layer. Bottom, the corresponding non-gradient layer.

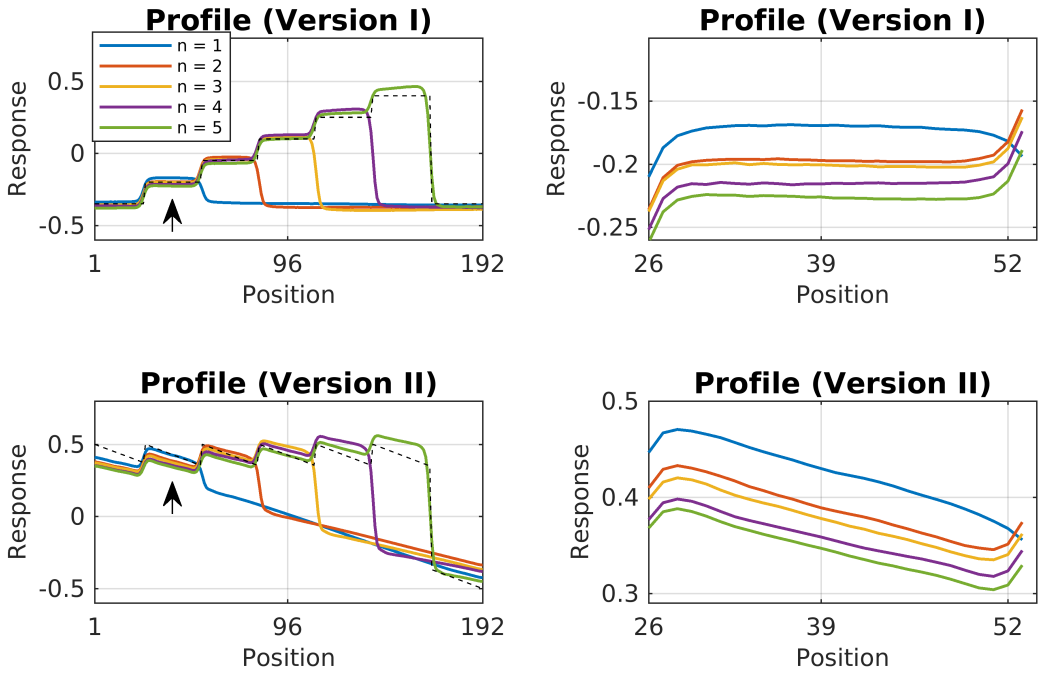


Figure 4.12: Profiles of full reconstruction (gradient + non-gradient) for G.E version I and G.E version II. *Left.* The profiles indicate the full reconstruction responses for n = 1, 2,3,4, and 5 squares for Gelb version I and Gelb version II. The legend indicates the color line corresponding to the number of squares in the sequence (for all profiles). *Right.* The amplified profiles show the section indicated by the arrow in the left. Notice the tendency of rescaled lightness when the number of squares increases in sequence.

Chapter 5

General Discussion

Chapter abstract: This chapter describes an overall discussion of the scope, viability, and limitations of the presented mechanisms in this study. These mechanisms consisted of a segregation mechanism (Chapter 2), dynamic filtering (Chapter 3), and a reconstruction stage (Chapter 2 and 3). The output of our model is the visual input (luminance) segregated in two layers: gradient and non-gradient representations – which correlated to lightness and the perceived illumination, respectively, while full reconstruction (summing both layers) correlated to brightness. In the context of this segregation hypothesis, we also discuss the linearity between brightness and lightness computations achieved by our model.

Segregation Mechanism

The segregation mechanisms proposed in Chapter 2 build on the idea that surface reflectance changes are often associated with fast luminance variations (e.g., surface edges) and shading is often associated with slow variations (i.e., gradients). Although it is clear that edges surface and gradient representations have to interact at some level in the object recognition hierarchy (e.g., in order to derive shape from shading), how such interactions could be implemented - if they exist - at an early level in the visual system is not clear.

At first sight, multi-scale models based only on frequency range may encode luminance gradients through large-scale filters that respond to a low spatial frequency range. However, as noted, the low-spatial filters respond to gradient information (smooth transitions) and surface changes (see Figure 4.2 in Chapter 4). being the gradient and non-gradient information is mixed across the frequency range, and an additional mechanism is necessary to distinguish gradients vs. non-gradients. A proposed mechanism that successfully distinguishes gradients vs. non-gradients is the 'gradient system' proposed by Keil (Keil, 2008). The 'gradient system' consists of a hypothetical neuronal circuit in the retina based on ON and OFF' channels' interaction. Gradients are detected according to the activity peaks of ON and OFF cells occurring in close or remote. This mechanism indirectly amounts to a boundary detection that serves as a source or sinks to generate representations of linear luminance gradients through lateral propagation of activity.

Inspired by Keil's model, we implemented an alternative mechanism based on filter responses' variability. The mechanism distinguishes the gradient and non-gradient responses according to low vs. high variability, which indirectly amounts to boundary detection. Although any contour detector could compute segregation, we noted some interesting aspects and advantages of the variability response mechanism (see Chapter 2). In particular, the variability response was irrespective of each filter's contrast amplitude or preference response. It means that different shapes of receptive fields would be feasible to compute variability response "without the need to use contour detectors" as simple cells in V1. Since other shapes of receptive fields would be compatible, we propose the segregation mechanism – if it exists – may be located in other areas (e.g., the retina, the lateral geniculate nucleus (LGN), or higher visual areas).

Further, considering that variability response was computed locally, the variability-based segregation mechanisms could be feasibly implemented via population coding (Averbeck et al., 2006). It suggests variability-based encoding strategies may be a promising mechanism for conveying information about visual attributes. Indeed, recently, response variability in individual neurons of oriented-selective neurons has been modeled from neural recordings to encode the uncertainty of orientation and texture (Hénaff et al., 2020). However, at present, we are not aware of any neurophysiological study that connects response variability to gradient and non-gradient cues, and the approach of luminance gradients remains only for psychophysical studies and computational models.

Reconstruction Stage

Unlike the segregation mechanism, the shape of filters was essential for computing gradients and non-gradients layers in the reconstruction stage. We identified that the odd-symmetrical filters responded best to smooth luminance gradients, whereas the even-symmetrical filters responded poorly to gradient features (see Chapter 2). Indeed, the odd-symmetrical filters act as first-order derivative filters which determine the direction of propagated activity in a filling-in process. Since the primary visual cortex (V1) and secondary visual cortex (V2) contain simple cells with odd "derivative" symmetrical shapes, the result suggests feasible V1 or V2 in computing non-gradient and gradient representations. Assuming in our hypothesis that the non-gradients representation "correlates" with surface representations; this result is in agreement with the conclusion of previous neurophysiological recordings that suggest neurons in V1 and V2 as the first stage to compute brightness/lightness surface perception (Huang and Paradiso 2008; Hung 2007; Vladusich et al. 2006; Roe et al. 2005; Friedman et al. 2003; Kinoshita and Komatsu 2001; MacEvoy et al. 1998; MacEvoy and Paradiso 2001; Rossi and Paradiso 1996; Rossi et al. 1996).

Dynamic Filtering

The dynamic filtering (DF) mechanism proposed in Chapter 3 successfully estimated different brightness/lightness phenomena (e.g., contrast effects, assimilation, induced brightness) for many visual illusions containing homogeneous illumination (except for Mach Bands), and also it predicted, phenomenologically, the tendency of some psy-

chophysical studies (see Chapter 3). Further, as we noted in Chapter 4, DF is compatible with the segregation mechanism. DF still predicts brightness/lightness phenomena predicted in the full reconstruction (summing both layers), being dominated by either or one of the two layers. It expands the scope of the DF mechanism for visual inputs that also contain gradients luminance patterns (e.g., illumination inhomogeneous).

Computationally, DF implements a spatial neural coding based on contextual modulation that reduces spatial redundancy (correlations) at boundary representations of an energy map that mimics the activity of complex cells in V1. The redundancy reduction (or efficient representation) links with other prominent coding strategies for the early visual system as the efficient coding (H. B. Barlow 1961; Atick & Redlich, 1992; Levy & Baxter, 1996; Simoncelli & Olshausen, 2001; Vincent et al., 2005), predictive coding (Srinivasan et al., 1982), whitening (Atick & Redlich, 1992) or response equalization (Keil, 2008a). have been chosen independently for each illusion and each model, to obtain the best possible replication of the visual illusion.

Recently, other models based on efficient representations of neural dynamics in V1 have shown promise in explaining distinct brightness phenomena (Bertalmío et al., 2020), but the parameters of these models were chosen independently for each simulation. Here, we believe the success in predicting a relatively large number of visual illusions without changing any parameter supports DF as a computational principle (see Chapter 3). However, the nature of the dynamic filter represents a difficulty. In order to learn the kernel for dynamic filtering, it is necessary somehow to compute local spatial autocorrelations. Besides, because DF is learned by data-driven, i.e., DF is reliant on the self-structure of the visual input but not on accumulated visual experience, such autocorrelations have been achieved in early visual areas as V1. Various models suggest that the columns are the locus for the computation of local spatial autocorrelation functions where strong intracolumnar excitation of the complex cells (in V1) could provide a substrate for detecting the local auto-correlation functions (Zucker 2006). However, physiological evidence of spatial autocorrelations in V1 is still lacking at the moment.

Limitations

In our simulations, we detected some counterexamples in addition to those described in our previous studies (Chapter 2 and Chapter 3) that challenge the core mechanism

and 'concepts' of the current model. In particular, these counterexamples emerge in predicting some visual illusions that contain certain articulated patterns. For instance, in the simultaneous contrast illusion, some studies have recorded a tendency to increase the perceived contrast if the background contains articulation patterns in a certain luminance range (Bressan & Actis-Grosso, 2006). Notice in Figure 5.1, the contrast in the articulated display is more "intense" than the original. As we noted in our second study (Chapter 3), dynamic filtering is inconsistent with this fact— since the more articulated patterns in visual input, the more spatial correlation with respect to the target test (i.e., Scenario 2 described in Chapter 3). Therefore, DF's would involve, on the target test, assimilation on the background instead of contrast.

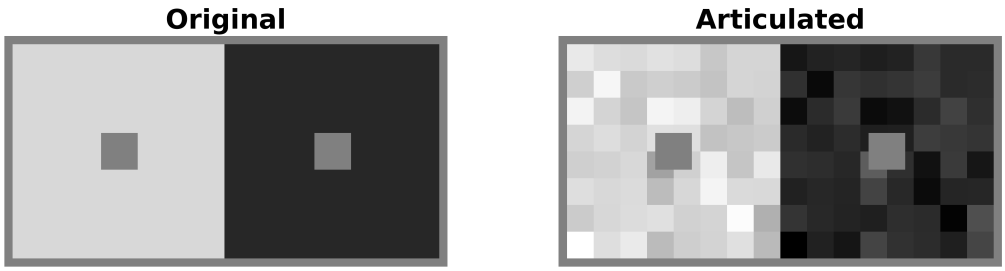


Figure 5.1: Example 1 that challenge the model. Original Simultaneous Contrast and Simultaneous Contrast with articulated patterns. The articulated patterns intensify the contrast phenomena (Bressan & Actis-Grosso, 2006). The articulated patterns have the same mean luminance as the background of the original displays.

We tried a "possible" solution. We readjusted the threshold parameter on the variability response of the segregation mechanism (equation 3 in Appendix A) in such a way that the articulated patterns were detected as gradient cues while the target tests were detected as non-gradients. In this way, on the one hand, the DF of the target test on the non-gradient layer was unaffected by articulated patterns, whereas, on the other hand, the articulated pattern served to increase balancing between gradient and the non-gradient layer. By this adjustment, the model estimated the tendency of articulated simultaneous contrast, but it was very slight (see Appendix B).

Another counterexample that challenges our model concerns articulated patterns that form luminance pseudo gradients. Physically, these luminance pseudo gradients are not gradients, but phenomenologically they seem to act as if they were. Figure 5.2 shows

an example of the phenomenon described.

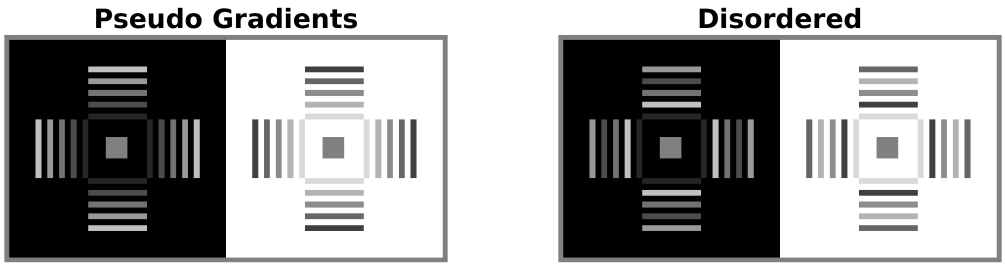


Figure 5.2: Example 2 that challenges the model. Simultaneous Contrast with a luminance pseudo gradient configuration and disordered configuration. In both cases, the luminance level of bars correspond to same values range.

Notice that the perception of luminance pseudo gradients. Similar to gradient configuration (in Chapter 4), a self-luminosity sensation emerges with luminance pseudo gradients, inducing, in turn, a stronger contrast effect than 'disordered' bars. The only way that the model can predict this phenomenon is by detecting pseudo gradients as gradients cues. A "first" solution would be to smooth visual input by a low-pass filter. However, the problem is that the luminance pseudo gradients contain gaps between bars, involving high variability responses – even with low-pass filtering – and therefore, they would be captured by the non-gradient layer instead of the gradient layer. We tried other strategies based on the current model's existing ideas, but they failed to find a satisfactory solution. Then, how detect luminance pseudo gradients as gradients cues for our model? Since the luminance pseudo gradients have a spatial structure (i.e., are not allocated randomly), a possible solution would be to introduce a selective integration mechanism that distinguishes contours: detecting boundaries that contain abrupt luminance changes – with spatial structure – susceptible to be gradient cues. Computationally this approach would be feasible because other contour systems that distinguish boundaries have been previously implemented successfully in brightness/lightness models (W. D. Ross & Pessoa, 2000), even considering biological constraints (Domijan, 2015; Grossberg & Todorovic, 1988). Perhaps, another possible mechanism would be to introduce a "texture system." After having segregated the bars (= texture), the residual gradient should remain. Diffusive reconstruction of these residual gradients should reconstruct the perceived gradient representation (see Keil, 2018 section 4).

Full reconstruction: Brightness and Lightness computations

What or how computations achieve and distinguish brightness and lightness from luminance is still unclear. Initially, we suggested linear computation between lightness and brightness by distinguishing (or balancing) gradient and non-gradient representations: the perceived illumination as gradient representations, lightness with non-gradient, and brightness with full reconstruction (summing both representations). However, we observed that this linearity only worked for certain visual configurations. We refer term linearity to indicate a direct distinction between perceived lightness, brightness, and apparent illumination through layers representations.

On the one hand, in visual images containing smooth inhomogeneous illumination (i.e., gradients) over the same surface background, our model easily distinguishes linearly lightness vs. brightness: In the case of positive and negative ramps configurations, the appearance of the surface in non-gradient representation balances in the "opposite direction" of the gradient luminance background (see Chapter 2 and Chapter 4). It corroborates with previous studies, which concluded luminance gradient as a perceptual feature that interacts in some level with lightness computations (Keil, 2008; Correani et al., 2006), and the perceived lightness goes with an equal but opposite error in perceived illumination induced by gradient configuration. Although this result would be in concordance with discounting illumination hypothesis, remember that the segregation mechanism erroneously detects an illumination edge as a non-gradient feature (see Chapter 2). Our model would operate as partial discounting illumination rather than total discounting illumination. Partial discounting illumination was suggested in previous studies, in which the luminance border produced by an illumination edge was perceived as partly a surface-brightness edge and partly a lightness edge (Logvinenko 2011).

On the other hand, in visual images containing a homogeneous (constant) illumination, we noted linearly in concordance with one would expect of the judgments of brightness and lightness inhomogeneous illumination (Blakeslee et al., 2008). However, due to the architecture of the extended model, this linearity depended on certain configurations of the background. As mentioned in Chapter 4, the response to luminance information of the extended model led to partial segregation introducing mixed artifacts between gradient and non-gradient layers.

In particular, for a background containing only a single luminance level or containing

high articulation patterns (e.g., checker), the mixed artifacts were located locally, and we observed almost complete linearity. However, the mixed artifacts extended through the layers for a background containing large regions with different luminance (e.g., low articulations), failing linearity (see Appendix C). We think the latter is a limitation in the reconstruction stage rather than the segregation mechanisms. Linearity could be solved in different ways: modifying luminance response or other complementary mechanisms without contradicting the presented results or conclusions.

In short, our results suggest linearity between lightness and brightness computations for visual images containing homogeneous or smooth inhomogeneous illumination. In the case of illumination edges, brightness and lightness computations seem more intricate (Logvinenko 2011) and possibly depend on other factors such as higher-level mechanisms or accumulated experience.

Future Directions

This thesis aims to help understand the underlying principles of low-level perception (as brightness and lightness) results from neural-based computations.

In Chapter 2 and Chapter 3, a segregation-based mechanism and a dynamic filtering mechanism were proposed. Although these mechanisms are prominent underlying computational principles in the visual system (for explaining brightness and lightness), neurophysiology studies will need to examine whether these mechanisms exist or are plausible as a neural coding strategy. For the segregation-based mechanism, we motivate future neurophysiology studies to record the variability response in neuronal populations. As our simulations suggest, the segregation-based mechanism could be implemented by a population coding, and it perhaps ubicated indistinctly in different visual areas as the retina, LGN, V1, or higher areas as V4. For the dynamic filtering mechanism, we suggest studying modulation responses of the extra-classical receptive field on simple and complex cells in V1 for stimuli that contain (in spatial context) high redundancy versus low redundancy in luminance patterns. Ideally, the modulation would work similarly to surround inhibition but as a function of spatial correlations. It could be tested with visual illusions containing contrast and assimilation. It would provide practical information on whether or not dynamic filtering exists and how brightness or lightness changes can be figured in cortical representations.

A consistent theme throughout the thesis work was the emphasis on studying how to predict a great number of visual illusions involving different phenomena under the presented mechanisms in Chapter 2, Chapter 3, and Chapter 4. Computationally, the main goal was to reduce the parameter space in each simulation (visual illusion) as low as possible. We build on the idea that if model parameters are adjusted for each illusion, then a model can hardly claim to propose a plausible mechanism for explaining underlying computations in perception. In this line, we want to motivate future computational models in predicting basic perceptions, a reduction in the parameter space in the simulations; we believe that it will be essential to better understand the scope of the hypothetical mechanisms, especially in the absence of neurophysiological studies. We also want to emphasize the importance of pointing out the inconsistencies of hypothetical mechanisms for future studies, which are sometimes omitted in previous studies. We believe, like visual illusions, understanding how models make accurate estimations is often best clarified by understanding why it sometimes does not. For instance, luminance pseudo gradient, articulated patterns, and edge illumination challenge our proposed mechanisms, and they could inspire future models to understand brightness and lightness perception better.

Altogether, this proposed computational model will provide important information concerning brightness and lightness computations. The work presented in this thesis provides a foundation upon which future simulations or computational models can be built. Understanding, unveiling principles, and coding strategies of basic perceptions through a computational model will lead to a better knowledge of neural coding for more complex processes.

Chapter 6

Conclusions

1. This thesis contributes new insight into brightness/lightness computations with a computational model for achromatic visual inputs.

2. The model successfully predicted many brightness and lightness phenomena for visual images with homogeneous or smooth inhomogeneous illumination, based on two hypothetical neural mechanisms: one based on variability in response that segregates gradient vs. non-gradient features the other on redundancy-reduction of edge representations.

3. In the simulations, some of these predictions were in concordance with psychophysical experiments, indicating these mechanisms could be prominent underlying computational principles in the visual system. However, there is a lack of neurophysiology studies correlating neural activity to such mechanisms, and this model must be considered only a proof of concept.

4. Since the model is compatible with low-level processing; we recommend in the future exploring these hypothetical mechanisms (if they exist) in early visual areas as V1/V2 and V4.

5. The neural mechanisms proposed in this thesis: redundancy-reduction hypothesis and variability-based segregation, should be compatible with all levels of information processing. This means that both might apply as well to other kinds of perceptions – even beyond brightness and lightness – as higher-order patterns and shapes that form the primitives for object recognition.

References

Adelson, E. H. (1995). Checkerboard optical illusion. Retrieved June 27, 2014 from MIT.

Agostini, Tiziano & Galmonte, Alessandra. (2002). A new effect of luminance gradient on achromatic simultaneous contrast. Psychonomic bulletin & review. 9. 264-9. 10.3758/BF03196281.

Anderson, B. & Winawer, J. (2005). Image segmentation and lightness perception. Nature. 434. 79-83. 10.1038/nature03271.

Anderson, B. L., Whitbread, M., & de Silva, C. (2014). Lightness, brightness, and anchoring. Journal of Vision, 14(9), 7–7. https://doi.org/10.1167/14.9.7

Anderson, B. L. & Winawer, J. (2005). Image segmentation and lightness perception. Nature 434:79-83

Arend, Lawrence & Spehar, Branka. (1993). Lightness, brightness, and brightness contrast: 1. Illuminance variation. Perception & psychophysics. 54. 446-56. 10.3758/BF03211767.

Atick, J. J., & Redlich, A. N. (1992). What Does the Retina Know about Natural Scenes? Neural Computation. https://doi.org/10.1162/neco.1992.4.2.196

Averbeck, Bruno, Latham, Peter & Pouget, Alexandre. (2006). Neural correlations, population coding and computation. Nature reviews. Neuroscience. 7. 358-66. 10.1038/nrn1888.

Barlow, H. (1961) "Possible principles underlying the transformation of sensory messages" in Sensory Communication, MIT Press

Baumgartner, G. (1960). Indirekte Größenbestimmung der rezeptiven Felder der Retina beim Menschen mittels der Hermannschen Gittertäuschung. Pflügers Archiv Für Die Gesamte Physiologie Des Menschen Und Der Tiere. https://doi.org/10.1007/BF00680926

Bertalmío M, Calatroni L, Franceschi V, Franceschiello B, Gomez Villa A, Prandi D. (2020). Visual illusions via neural dynamics: Wilson-Cowan-type models and the efficient representation principle. J Neurophysiol. 2020 May 1;123(5):1606-1618. doi: 10.1152/jn.00488.2019. Epub 2020 Mar 11. PMID: 32159409.

Blakeslee B, McCourt ME. (1999). A multiscale spatial filtering account of the White effect, simultaneous brightness contrast and grating induction. Vision Research. Volume 39, Issue 26, Pages 4361–4377, ISSN 0042–6989, https://doi.org/10.1016/S0042-6989(99)00119-4 PMID: 10789430

Blakeslee B, McCourt ME. (2004). A unified theory of brightness contrast and assimilation incorporating oriented multiscale spatial filtering and contrast normalization. Vision Research. Volume 44, Issue 21, Pages 2483–2503, ISSN 0042-6989, https://doi.org/10.1016/j.visres.2004.05.015 PMID:

Blakeslee, Barbara & McCourt, Mark. (2015). A common theoretical framework for brightness and lightness perception. Perception. 44. 359-362

15358084

Blakeslee, Barbara, Reetz, Daniel & McCourt, Mark. (, 2008). Coming to terms with lightness and brightness: Effects of stimulus configuration and instructions on brightness and lightness judgments. Journal of Vision. 8. 3.1-14. 10.1167/8.11.3.

Bressan, Paola & Actis-Grosso, Rossana. (2006). Simultaneous lightness contrast on plain and articulated surrounds. Perception. 35. 445-52. 10.1068/p5247.

Cataliotti, J & Gilchrist, Alan. (1995). Local and global processes in surface lightness perception. Perception & psychophysics. 57. 125-35.

Cohen MA, Grossberg S. (1987). Neural dynamics of brightness perception: Features, boundaries, diffusion, and resonance. Advances in Psychology. Volume 43,Pages 212–270, ISSN 0166-4115, ISBN 9780444701183, https://doi.org/10.1016/S0166-4115(08)61762-8

Correani, Alessia, Scott-Samuel, Nick & Leonards, Ute. (2006). Luminosity—A perceptual "feature" of light-emitting objects?. Vision research. 46. 3915-25. 10.1016/j.visres.2006.05.001.

Domijan, D. (2015). A Neurocomputational account of the role of contour facilitation in brightness perception. Frontiers in Human Neuroscience,9,93

https://doi.org/10.3389/fnhum.2015.00093

Faul Franz & Ekroll Vebjørn. (2011). On the filter approach to perceptual transparency. Journal of vision, 11. https://doi.org/10.1167/11.7.7 PMID: 21659428

Friedman HS, Zhou H, von der Heydt R (2003) The coding of uniform colour figures in monkey visual cortex. J Physiol 548:593-613

Gerrits, H. J. M., & Vendrik, A. J. H. (1970). Simultaneous contrast, filling-in process and information processing in man's visual system. Experimental Brain Research.

Gilchrist, Alan & Kossyfidis, Christos & Bonato, Frederick & Agostini, Tiziano & Cataliotti, Joseph & Li, Xiaojun & Spehar, Branka & Annan, Vidal & Economou, Elias. (1999). An anchoring theory of lightness perception. Psychological review. 106. 795-834. 10.1037//0033-295X.106.4.795.

Grossberg, Stephen & Todorovic, Dejan. (1988). Neural Dynamics of 1-D and 2-D Brightness Perception: A Unified Model of Classical and Recent Phenomena. Perception & psychophysics. 43. 241-77. 10.3758/BF03207869.

Grossberg, A. (2017) The visual world as illusion: The ones we know and the ones we don't. In Oxford Compendium of Visual Illusions (A. Shapiro and D. Todorovic, Eds.) Oxford, United Kingdom: Oxford University press, Chapter 7, pp. 90-118

Hénaff Olivier, Boundy-Singer Zoe, Meding Kristof, Ziemba Corey & Goris, Robbe. (2020). Representation of visual uncertainty through neural gain variability. Nature Communications. 11. 10.1038/s41467-020-15533-0.

Huang, Xin & Paradiso, Michael. (2008). V1 Response Timing and Surface Filling-In. Journal of neurophysiology. 100. 539-47. 10.1152/jn.00997.2007.

Jones, J. P., & Palmer, L. A. (1987a). An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex. Journal of Neurophysiology. https://doi.org/10.1152/jn.1987.58.6.1233

Kinoshita M, Komatsu H (2001) Neural representation of the luminance and brightness of a uniform surface in the macaque primary visual cortex. J Neurophysiol 86:2559-2570

Komatsu, H. (2006) The neural mechanisms of perceptual filling-in. Nature Review Neuroscience 7(3):220-231

Keil, M. S. (2006). Smooth gradient representations as a unifying account of Chevreul's illusion, mach bands, and a variant of the Ehrenstein disk. Neural Computation, 18(4), 871–903. http://dx.doi.org/10.1162/089976606775774705

Keil, M. S., Cristóbal, G., & Neumann, H. (2006). Gradient representation and perception in the early visual system-A novel account of Mach band formation. Vision Research, 46(17), 2659–2674. http://dx.doi.org/10.1016/j.visres.2006.01.038.

Keil, Matthias. (2008). Gradient Representations and the Perception of Luminosity. Vision research. 47. 3360-72. 10.1016/j.visres.2007.09.018.

Keil, M. S. (2008a). Does face image statistics predict a preferred spatial frequency for human face processing? Proceedings of the Royal Society B: Biological Sciences. https://doi.org/10.1098/rspb.2008.0486

Keil, Matthias. (2018). From Neuronal Models to Neuronal Dynamics and Image Processing. https://arxiv.org/abs/1801.08585

Levy, W. B., & Baxter, R. A. (1996). Energy Efficient Neural Codes. Neural Computation. https://doi.org/10.1162/neco.1996.8.3.531

Logvinenko, Alexander & Tokunaga, Rumi. (2011). Lightness constancy and illumination discounting. Attention, perception & psychophysics. 73. 1886-902. 10.3758/s13414-011-0154-2.

MacEvoy SP, Kim W, Paradiso MA (1998) Integration of surface information in primary visual cortex. Nat Neurosci 1:616-620.

MacEvoy SP, Paradiso MA (2001) Lightness constancy in primary visual cortex. Proc Natl Acad Sci U S A 98:8827-8831.

Morrone, M. C., & Burr, D. C. (1988). Feature detection in human vision: a phase-dependent energy model. Proceedings of the Royal Society of London. Series B, Containing Papers of a Biological Character. Royal Society (Great Britain). https://doi.org/10.1098/rspb.1988.0073

Morrone, M. C., & Owens, R. A. (1987). Feature detection from local energy. Pattern Recognition Letters. https://doi.org/10.1016/0167-8655(87)90013-4

Morgenstern Y, Rukmini DV., Monson BB, Purves D. (2014). Properties of artificial neurons that report lightness based on accumulated experience with luminance. Front Comput Neurosci. 2014; 8:134. https://doi.org/10.3389/fncom.2014.00134 PMID: 25404912

Pizlo, Zygmunt. (2001). Perception viewed as an inverse problem. Vision research. 41. 3145-61. 10.1016/S0042-6989(01)00173-0.

Purves D, Williams SM, Nundy S, Lotto RB. (2004). Perceiving the Intensity of Light. Psychol Rev. ; 111(1):142–58. Available from: http://www.ncbi.nlm.nih.gov/pubmed/14756591

Roe AW, Lu HD, Hung CP (2005) Cortical processing of a brightness illusion. Proc Natl Acad Sci U S A 102:3869-3874.

Ross, W. D., & Pessoa, L. (2000). Lightness from contrast: A selective integration model. Perception & Psychophysics, 62(6), 1160–1181. https://doi.org/10.3758/BF03212120

Rossi, A. F., Rittenhouse, C. D., & Paradiso, M. A. (1996). The representation of brightness in primary visual cortex. Science. https://doi.org/10.1126/science.273.5278.1104

Rossi AF, Paradiso MA (1996) Temporal limits of brightness induction and mechanisms of brightness perception. Vision Res 36:1391-1398.

Rudd Michael E. (2017). Lightness computation by the human visual system. Journal of Electronic Imaging, 26(3), 031209. https://doi.org/10.1117/1.JEI.26.3.031209

Schmid, A. C. & Anderson, B. L. (2017) Perceptual dimensions underlying lightness perception in homogeneous center-surround displays. Journal of Vision 17(2):6, https://doi.org/10.1167/17.2.6

Simoncelli EP, Olshausen BA. Natural image statistics and neural representation. Annu Rev Neurosci. 2001;24:1193-216. doi: 10.1146/annurev.neuro.24.1.1193. PMID: 11520932.

Srinivasan, M. V., Laughlin, S. B., & Dubs, A. (1982). Predictive coding: A fresh view of inhibition in the retina. Proceedings of the Royal Society of London - Biological Sciences. https://doi.org/10.1098/rspb.1982.0085

Vincent, B. T., Baddeley, R. J., Troscianko, T., & Gilchrist, I. D. (2005). Is the early visual system optimised to be energy efficient? Network: Computation in Neural Systems. https://doi.org/10.1080/09548980500290047

Vladusich T, Lucassen MP, Cornelissen FW (2006) Do cortical neurons process luminance or contrast to encode surface properties? J Neurophysiol 95: 2638–2649.

Yang Z, Purves D. (2004). The statistical structure of natural light patterns determines perceived light intensity. Proc Natl Acad Sci; 101(23):8745–50. Available from: https://www.pnas.org/content/101/ 23/8745. https://doi.org/10.1073/pnas.0402192101 PMID: 15152077

Zucker, Steven. (2006). Which Computation Runs in Visual Cortical Columns?. 23 Problems in Systems Neuroscience. 10.1093/acprof:oso/9780195148220.003.0011.

Zavagno, D., & Daneyko, O. (2012). The effect of non-adjacent luminance gradients on simultaneous lightness contrast. In C. Leth-Steensen, & R. Schoenherr (Eds), Proceedings of the 28th Annual Meeting of the International Society for Psychophysics (pp. 48–53). Ottawa, Canada: International Society for Psychophysics.

Appendix

A. Details and modification of parameters of the extended model

Filters. We used a set of Gabor filters to the encoding input image. The parameters and configuration of the set of filters were equal to our second study (Chapter 3). The single modification was the unbalanced parameter for Contrast-Luminance filters which we adjusted to 0.02 – see Appendix S1.Text in Chapter 3.

Variability response. To simplify computations, we modified the original variability map response presented in (Chapter 2, equation 6). Here, for each activity response filter (i.e. "activity in response to the input image by filter g") the model measures:

$$map_g = Var_L(R_g)$$

where Var_L indicates the local variance in a small neighborhood (7×7 pixels moving window, with mirror boundary conditions). Subsequently, the final variability map is computed as the square root of the sum of each variability filter response:

$$V_{map} = \sqrt{\sum_g map_g}$$

Notice that, mathematically, this computation could be interpreted as the local standard deviation between all filter responses if we assume the filter responses as independent random variables.

Gating map of segregation mechanism. We formulated the gating map, thresholding the variability response map as:

$$G_{map}(x,y) = \begin{cases} 1 & if\ \tilde{V}_{map}(x,y) > 0.1 \\ 0 & otherwise \end{cases} \tag{A.1}$$

where (x,y) indicates the spatial pixel coordinates and is the normalization of measured as:

$$\tilde{V}_{map} = \frac{V_{map}}{max_{x,y}\{V_{map}(x,y)\}}$$

Dynamic filtering. The extended model applies dynamic filtering (in parallel) to the gradient and non-gradient response after segregation. We adjusted the parameters with identical values of our second study (Chapter 3).

Reconstruction of gradient and non-gradient layer. The parameters and equations of the reconstruction stage for both layers – gradient and non-gradient – were the

same as the published studies (Chapter 2 and Chapter 3), but we fixed the stop criterion to 10^{-6}.

B. Simulation for Simultaneous Contrast With articulated patterns

Figure A1 illustrates the results of our extended model for Simultaneous Contrast with the articulated pattern. The tendency in increasing the contrast intensity on articulated S.C with respect to classical S.C was achieved by readjusting the threshold of segregation mechanism of the equation A.1 (see Appendix A) to 0.25 instead of 0.1 In full reconstruction, the contrast intensity prediction was approximately 0.16 for articulated S.C and approximately 0.12 for classical S.C.

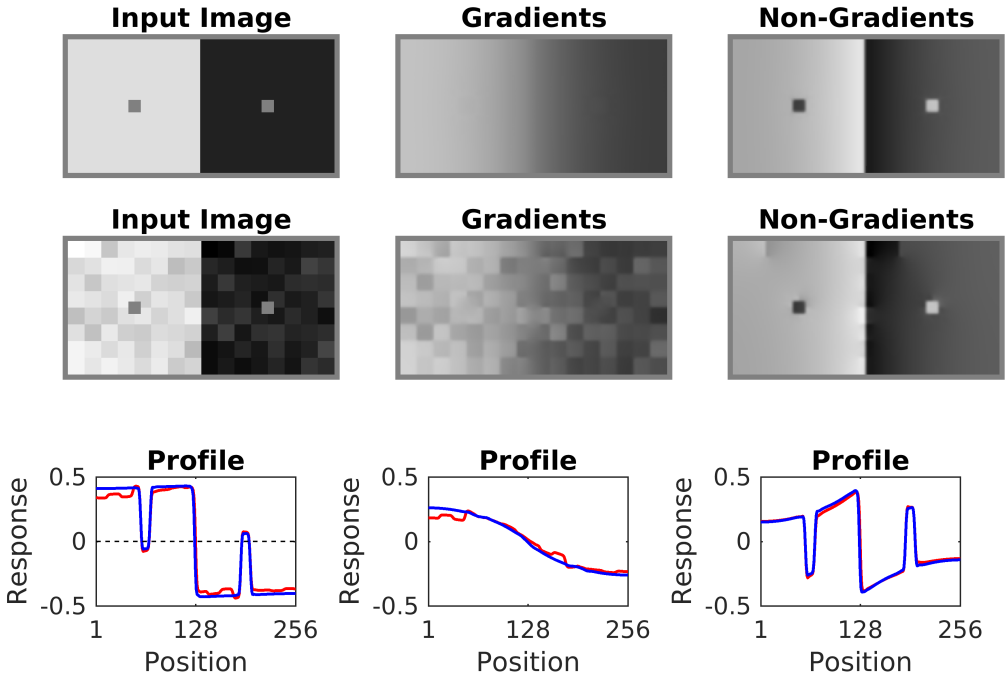


Figure A1: Simulations of Simultaneous Contrast with and without articulated patterns. In both cases, the luminance backgrounds were fixed equivalent average luminances. *Top.* The Classical Simultaneous Contrast and the corresponding gradient and non-gradient layer segregation. *Middle.* The Articulated Simultaneous Contrast and the corresponding gradient and non-gradient layer segregation. *Bottom.* The corresponding profiles of each estimation of in red (articulated), in blue (original). The (left to right) profiles show the full reconstruction, gradients, and non-gradients.

C. Simulation with different backgrounds.

Figure B1 illustrates the simulations for different visual inputs containing different background styles. The mixed artifacts were located locally for a single luminance level or high articulation patterns (e.g., checker). However, the mixed artifacts extended through the layers for large regions with different luminance (e.g., low articulations).

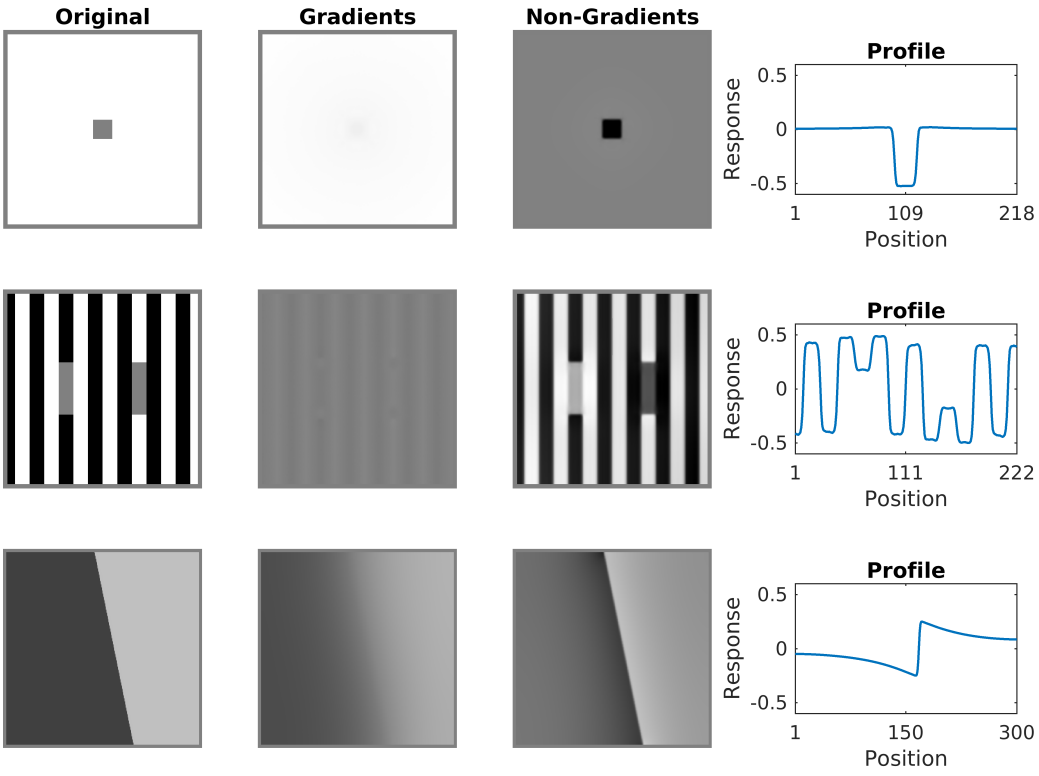


Figure B1: Simulations with different backgrounds. *Top.* A visual input with a background containing only a single level of luminance and its corresponding profile of non-gradient representation. *Middle.* A visual input with a background containing different luminance but with high articulation patterns and its corresponding profile of non-gradient representation. *Bottom.* A visual input with a background containing two large regions with different level luminance and its corresponding profile of non-gradient representation.