

RESEARCH

Matching visual induction effects on screens of different size by regularizing a neural field model of color appearance

Trevor D Canham^{1*}, Javier Vazquez-Corral¹, Elise Mathieu^{1,2} and Marcelo Bertalmío¹

*Correspondence:

trevor.canham@upf.edu

¹Department of Information and Communication Technologies, Universitat Pompeu Fabra, 08018 Barcelona, Spain

Full list of author information is available at the end of the article

Abstract

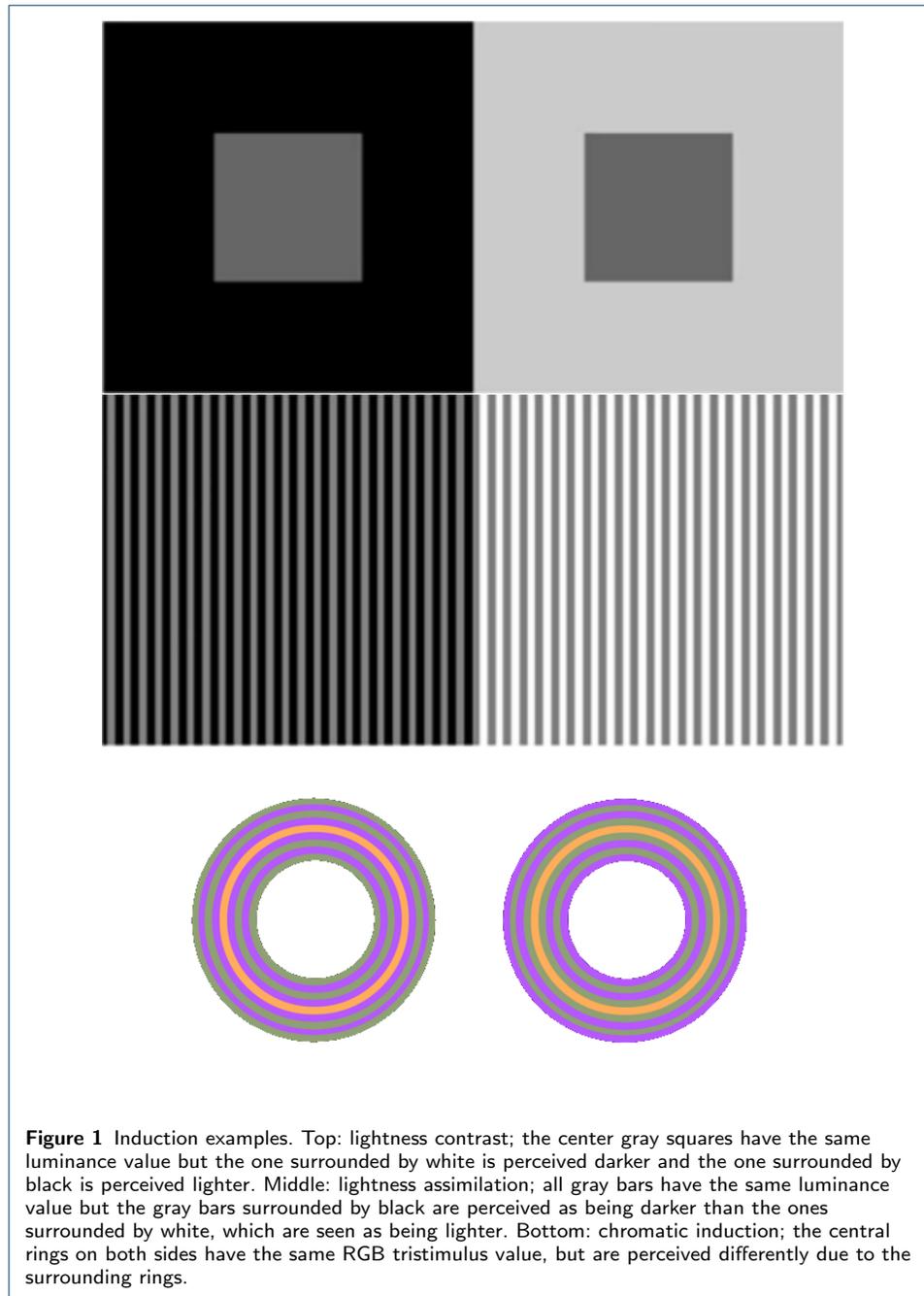
In the film industry, the same movie is expected to be watched on displays of vastly different sizes, from cinema screens to mobile phones. But visual induction, the perceptual phenomenon by which the appearance of a scene region is affected by its surroundings, will be different for the same image shown on two displays of different dimensions. This presents a practical challenge for the preservation of the artistic intentions of filmmakers, as it can lead to shifts in image appearance between viewing destinations. In this work we show that a neural field model based on the efficient representation principle is able to predict induction effects, and how by regularizing its associated energy functional the model is still able to represent induction but is now invertible. From this we propose a method to pre-process an image in a screen-size dependent way so that its perception, in terms of visual induction, may remain constant across displays of different size. The potential of the method is demonstrated through psychophysical experiments on synthetic images and qualitative examples on natural images.

Keywords: Color perception; Visual induction; Efficient representation principle; Neural field models; Local histogram equalization; Variational models; Wilson-Cowan equations

1 Introduction

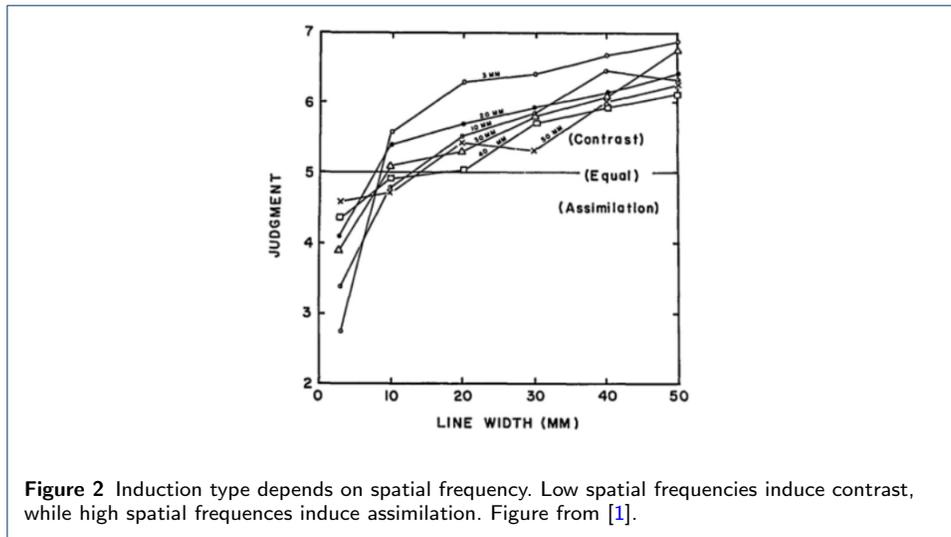
In visual perception, induction designates the effect by which the lightness and chroma of a stimulus are affected by its surroundings. Visual induction can take two forms: assimilation, when the perception of an object shifts towards that of its surround, or contrast, when the appearance of an image region moves away from that of its local neighborhood. See Figure 1 for some examples.

The groundbreaking experiments of Helson in 1963 [1] aimed to quantify the perceptual phenomena first formally described by von Bezold [2] and Gelb [3], using matching experiments with printed induction bar patterns and isolated Munsell patches. Specifically, observers had to judge the appearance of grey bars over white or black backgrounds. When the bars were very thin, the observers reported assimilation; as the bars increased in width, the assimilation effect became less pronounced, and after some point the observers started to report contrast, whose effect became increasingly more pronounced as the width of the bars turned larger. See Figure 2. A similar result for the chromatic case was reported by Fach and Sharpe [4], who modulated the spatial frequency of patterns as opposed to the target background proportionality variation of Helson. Their conclusion was that for higher



spatial frequencies visual induction takes the form of assimilation, while for lower spatial frequencies it takes the form of contrast. Though not all confirm these early observations, there exists a large body of later work (e.g. [5, 6, 7, 8, 9, 10, 11]) corroborating the importance of the spatial distribution and variability of inducing surrounds. Regarding visual induction models, we single out the work of Otazu et al. [12], which is based on wavelet decompositions, and the very recent work of Song et al. [13], that employs a neural field model.

In the film industry, the same movie is expected to be watched on displays of vastly different sizes, from cinema screens to mobile phones. But the typical viewing angle



depends on screen size, being larger for larger displays, and therefore the same image content will have a higher spatial frequency when seen on a small screen than when seen on a larger one. As a consequence, the visual induction effects on both screens may not be of the same magnitude or even type: in the smaller display, induction effects of the contrast kind will have less magnitude and tend towards assimilation.

It is common practice in motion picture distribution to manually modify the original mastered picture when distributing to different display scenarios. In this process, a skilled artist works to ensure that the visual storytelling intentions of the piece in its original format are preserved. For instance, a piece which had an original theatrical release may be remastered for separate releases to home video, broadcast television, streaming, etc. However, new standards and developments in the display industry of the past decade (high dynamic range, wide color gamut, 4K and 8K displays, mobile devices) have increased the variability in potential content destinations to the point where manual processing is no longer a feasible solution for handling distribution masters. For this reason, it is an increasingly relevant effort for the motion picture industry to develop solutions to adjust content automatically considering the specific viewing scenario parameters of the user.

With the abovementioned in mind, in this work we make three main contributions. First, we show that a neural field model based on the efficient representation principle is able to predict induction effects, and this model is validated using existing psychophysical data. Second, we prove that by regularizing its associated energy functional the model is still able to predict induction but now it becomes invertible. Finally, based on this invertible formulation we propose a method to pre-process an image in a screen-size dependent way so that its perception, in terms of visual induction, may remain constant across displays of different size. The potential of the method is demonstrated through novel psychophysical experiments on synthetic images; all this data is made available as supplementary material.

2 The efficient representation principle, local histogram equalization and the Wilson-Cowan neural field equations

The efficient representation principle, introduced by Attneave [14] and Barlow [15], is a general strategy observed across mammalian, amphibian and insect species, where visual processing considers the statistics of the visual stimulus and adapts to its changes [16]. In fact, efficient representation requires that the statistics of the image input are matched by the coding strategy, and while a global part of this coding strategy must have evolved on long timescales (development, evolution), in order to be truly efficient the coding must also adapt to the local spatio-temporal changes of natural images occurring at timescales of hours (e.g. from daybreak to dawn), seconds (e.g. when we move from one environment into another), or fractions of a second (e.g. when our eyes move around). By constantly adapting to the statistical distribution of the stimulus, the visual system can encode signals that are less redundant and this in turn produces metabolic savings by having weaker responsiveness after adaptation, since action potentials are metabolically expensive [17].

Atick [18] makes the point that there are two different types of redundancy or inefficiency in an information system like the visual system:

- 1 **If some neural response levels are used more frequently than others.** For this type of redundancy, the optimal code is the one that performs *histogram equalization*. There is evidence that the retina is carrying out this type of operation at photoreceptor level [19], as their response curves match the cumulative histogram of the luminance distribution of the environment.
- 2 **If neural responses at different locations are not independent from one another.** For this type of redundancy the optimal code is the one that performs decorrelation. There is evidence in the retina, the LGN and the visual cortex that receptive fields act as optimal “whitening filters”, *locally* decorrelating the signal.

From the above, a local histogram equalization (LHE) process would simultaneously reduce both types of redundancy. In [20] Bertalmío et al. propose a variational method to improve the color appearance of images, that performs LHE. They introduce the following energy functional:

$$E(I) = \frac{\alpha}{2} \int_{\Omega} (I(x) - \frac{1}{2})^2 dx - \gamma \int_{\Omega^2} w(x, y) |I(x) - I(y)| dx dy + \frac{\beta}{2} \int_{\Omega} (I(x) - I_0(x))^2 dx, \quad (1)$$

where I is an image in the range $[0, 1]$, Ω is the image domain, x, y are pixels, w is a distance function such that its value decreases as the distance between x and y increases, I_0 is the original image and α, β and γ are positive weights. The gradient descent equation for this functional is

$$I_t(x) = -\alpha(I(x) - \frac{1}{2}) + \gamma \int_{\Omega} w(x, y) \text{sgn}(I(x) - I(y)) dy - \beta(I(x) - I_0(x)) \quad (2)$$

Starting from $I = I_0$, equation 2 is iterated until a steady state is reached, which will be the result of this algorithm.

The energy in Equation 1 introduces the influence of spatial neighbors through the distance function w . Without it (and with $\beta = 0$) the energy becomes the one proposed by Sapiro and Caselles in [21], whose minimization produces a (global) histogram equalization of the original image. Therefore, we can argue that the evolution equation 2 performs *local* histogram equalization.

The LHE method of [20] has several good properties:

- 1 It has a very good local contrast enhancement performance, producing results without halos, spurious colors or any other kind of visual artifact.
- 2 It “flattens” the histogram, approaching histogram equalization.
- 3 It reproduces visual perception phenomena such as simultaneous contrast and the Mach Band effect.
- 4 It yields very good color constancy results, being able to remove strong color casts and to deal with non-uniform illumination (a challenging scenario for most color constancy algorithms, as discussed in [22]).

Additionally, the LHE model of [20] is closely related to the neural field model of Wilson and Cowan, as pointed out in [20] and further discussed in [23]. In particular, the evolution equation 2 is very similar to the Wilson-Cowan equations (see [24, 25, 26]), which have a long and thriving history of modelling cortical low-level dynamics [27]. It has been proven recently [28] that the Wilson-Cowan equations are not variational, in the sense that they can’t be minimizing an energy functional, and that the simplest modification that makes them variational yields the LHE method of [20]; furthermore, the LHE model provides a better reproduction of visual illusions than the Wilson-Cowan model. The study of visual illusions has always been key in the vision science community, as the mismatches between reality and perception provide insights that can be very useful to develop new models of visual perception [29] or of neural activity [30], and also to validate existing ones. It is commonly accepted that visual illusions arise due to neurobiological constraints [31] that limit the ability of the visual system, and are therefore related to efficient representation. In short, the LHE method (in its original formulation of [20] and also when it considers orientation [32]) is the generalization of the Wilson-Cowan equations that makes them compliant with the efficient representation principle, and at the same time this allows for an improved reproduction of visual perception phenomena.

3 Modifying the LHE model so that it predicts induction

Looking at equation 2, we can see that the spatial arrangement of the image data is only taken into account by the weighting function w . But in practice w is very wide, and therefore we can expect that the local contrast enhancement procedure of [20] will always produce contrast, not assimilation, since as we mentioned previously assimilation is linked to high spatial frequencies [33]. In order to overcome the intrinsic limitations of [20] with respect to induction, we should introduce spatial frequency in the energy functional. In [34] this is done by making the parameter γ in equation 2 change both spatially and with each iteration, according to the local standard deviation: if the neighborhood over which it is computed is sufficiently small, the standard deviation can provide a simple estimate of spatial frequency. But also, the standard deviation is commonly used in the vision literature as an estimate of local contrast. The model in [34] can predict lightness assimilation and

further improves efficiency by reducing redundancy: flattening the histogram and whitening the power spectrum. Other attempts to modify the LHE formulation so that it better deals with induction are discussed in [35].

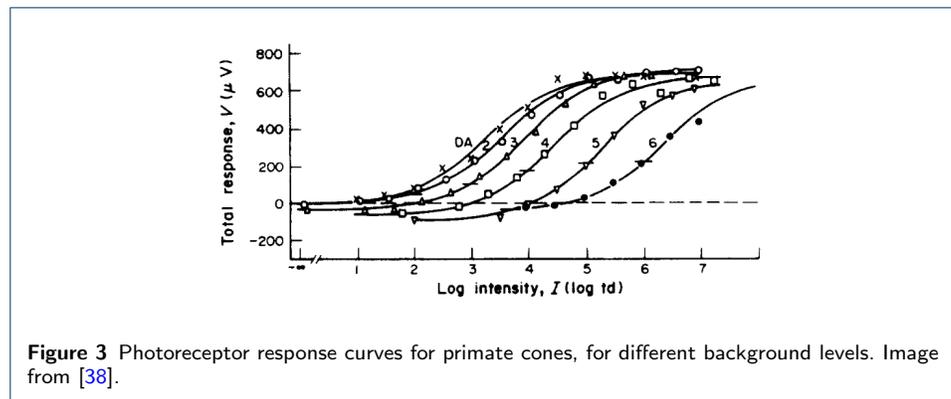
Unfortunately, the modifications introduced to the LHE model in [34] do not fit well with the basic postulates of Wilson and Cowan's theory. This is why in this section we propose to adapt the LHE model in a different manner in order to predict induction, with changes that are motivated by neurophysiology data and that now keep the model consistent with the Wilson-Cowan formulation. Specifically, we want to take into account the following biological phenomena.

Photoreceptor response

Photoreceptor response curves can be approximated very well with the Naka-Rushton equation:

$$R(I) = R_{max} \frac{I^n}{I^n + I_s^n}, \quad (3)$$

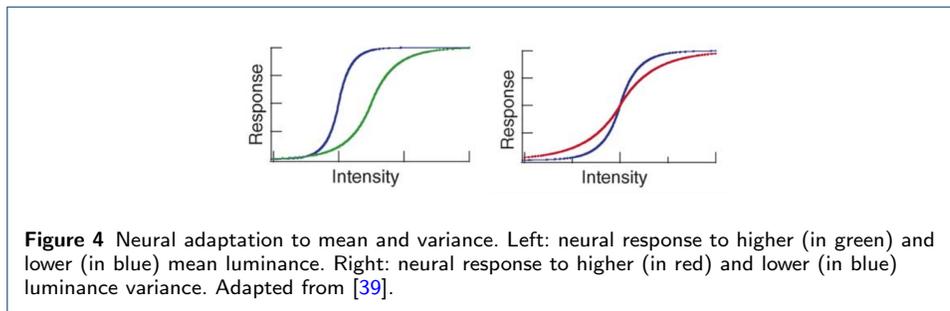
where R is the response, R_{max} is the maximum or saturation response, I is the intensity, n is an exponent of around 0.75, and I_s is the so-called semi-saturation value, the intensity at which the response is one-half of its maximum value and that roughly corresponds to the average intensity level. Notice that the Naka-Rushton equation is a monotonically increasing function and is therefore invertible; this point will become important later on. If we increase I_s and plot R in linear-log coordinates, as in Figure 3, then the curve moves to the right, the same curve-shifting phenomena observed when the background level increases. Therefore, light adaptation can be seen as changing the semi-saturation constant in the Naka-Rushton equation [36]. Furthermore, from Eq. 3 and if $n = 1$, we can obtain Weber's law. For this and other factors, it appears that the perceptual effects of light adaptation can be mostly accounted for by retinal processing [37].



Neural response nonlinearities and signal equalization

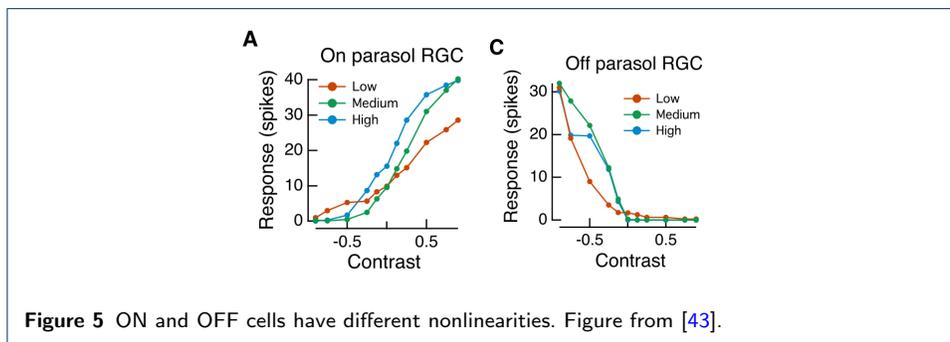
Neural adaptation performs a (constrained) signal equalization by matching the system response to the stimulus mean and variance [39], thus ensuring visual fidelity under a very wide range of lighting conditions. Figure 4 (left) shows that when the

mean light level is high, the nonlinear curve that models retinal response to light intensity is a sigmoid function with less steep slope than when the mean light level is low. Figure 4 (right) shows that at a given ambient level, the slope of the sigmoid is lower when the contrast is higher. In both cases, the data is consistent with the nonlinearity of the neural response to light performing histogram equalization, since the nonlinearity behaves as the cumulative histogram (which is the classical tool used in image processing to equalize a histogram) does: darker images and images with lower contrast typically have less variance and therefore their cumulative histograms are steeper. The psychophysical experiments in [40] corroborate that the visual system performs histogram equalization by showing how observers prefer display nonlinearities that allow the displayed image to be perceived as having a brightness distribution as close to uniform (i.e. with an equalized histogram) as possible.



Asymmetry of neural response nonlinearity

Recent works from neurophysiology prove that OFF cells (those that respond to stimuli with values below the average stimulus level) change their gain more than ON cells during adaptation [41], and that the nonlinear responses of retinal ON and OFF cells are different [42, 43, 44], see Figure 5. This data on neural activity is consistent with psychophysical data [45, 46] that demonstrates that our sensitivity to brightness is enhanced at values near the average or background level.



Retinal lateral inhibition can explain assimilation

Lateral inhibition creates the typical center-surround structure of the receptive field (RF) of retinal ganglion cells (RGCs), with the excitatory center due to the feed-forward cells (photoreceptors and bipolar cells) and the inhibitory surround due

to the inhibitory feedback from interneurons (horizontal and amacrine cells). This center-surround organization is a very important instance of efficient representation, performing signal decorrelation and allowing to represent with less resources large uniform regions because they generate little or no activity. It should also be pointed out that a more recent work [47] contends that decorrelation is already performed by the rapid eye movements that happen during fixations, and therefore that the signal arrives already decorrelated at the retina: the subsequent spatial filtering performed at the retina and downstream must have other purposes, like enhancing contrast.

Classical studies assumed that assimilation had to take place at a later stage than the retina, most probably at the cortex, because it needs a much longer range of interaction between image regions than what lateral inhibition could provide with the classical RF size. But in [48] Yeonan-Kim and Bertalmio showed that, in fact, assimilation can *start* already in the retina. They took classic retinal models, those of Wilson [49] and van Hateren [50], and adapted them so that parasol RGCs have a surround that is now dual, with a narrow component of large amplitude and a wide component of smaller amplitude. This different form for the surround is based on more recent neurophysiological data showing that retinal interneurons have RFs that are much more extended than previously assumed, and RGC responses show a component that goes beyond the classical RF.

4 Proposed modification of the LHE model

Based on the above we propose the following two-stage model:

- 1 The image stimulus I , which is a scalar-valued linear image (i.e. an image channel proportional to light intensity) is passed through the photoreceptor nonlinearity, modeled as a Naka-Rushton equation, yielding J_0 :

$$J_0 = NR(I) = \frac{I^n}{I^n + I_s^n}, \quad (4)$$

where the exponent of the NR equation is chosen so as to maximize the equalization of the histogram of J_0 , and the semisaturation constant I_s is the median average of the image.

- 2 The following evolution equation is run until a steady state is reached:

$$J_t(x) = -\alpha(J(x) - K_m * J(x)) + \gamma \int_{\Omega} K_c(x, y) \sigma(J(x) - J(y)) dy - \beta(J(x) - J_0(x)) \quad (5)$$

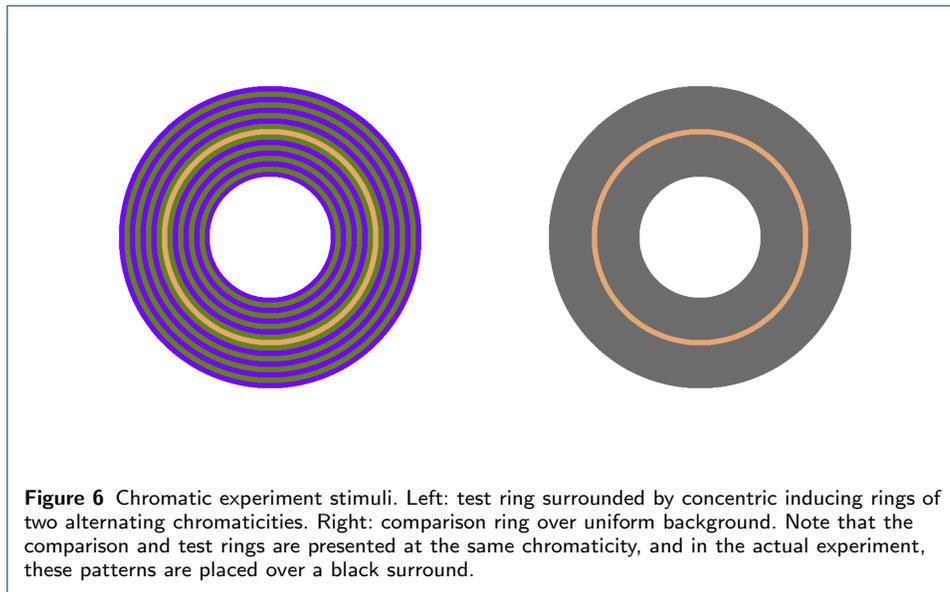
Here K_m, K_c denote kernels each expressed as a sum of two Gaussians and $*$ is the convolution operation, so now instead of a global mean $1/2$ as in Equation 2 we have a local mean $K_m * J(x)$ and local neighbors exert more influence but very far apart points can affect the response as well. Furthermore σ is a sigmoid function such that $\sigma(0) = 0$ but not necessarily antisymmetric, hence allowing positive and negative responses to be of different magnitude. Let us note that Eq. 5 is the gradient descent equation for an energy functional of

this form:

$$E(J) = \frac{\alpha}{2} \int_{\Omega} (J(x) - K_m * J(x))^2 dx - \gamma \int_{\Omega^2} K_c(x, y) \phi(J(x) - J(y)) dx dy + \frac{\beta}{2} \int_{\Omega} (J(x) - J_0(x))^2 dx, \quad (6)$$

where $\phi(\cdot)$ is a function whose derivative is the sigmoid $\sigma(\cdot)$.

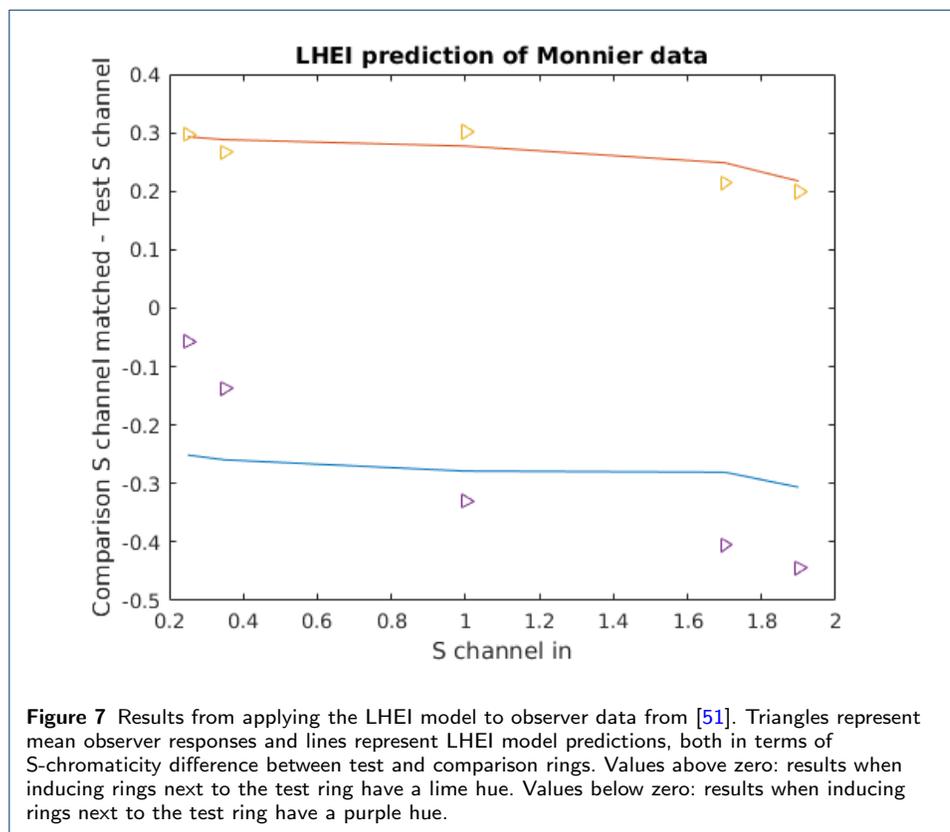
Let's call this model LHEI (*I* for "induction") for the sake of brevity. In order to validate it, we will use the chromatic induction data of [51]. In that work, observers were shown a test ring of some given chromaticity, surrounded by 16 concentric rings (half on each side of the test) that constitute the inducing pattern. This is the test image. The surrounding rings alternated between two chromaticities, which in isolation appear lime and purple, selected because they differentially stimulate the S cones only. Next to this image, the observer was shown a comparison ring, with the same dimensions as the test ring, but in this case simply presented over a uniform grey background (i.e. without inducing patterns). This is the comparison image. Observers adjusted the hue, saturation and brightness of the comparison ring in order to match the appearance of the test ring. See Figure 6 for an illustration of this experimental set-up.



The resulting chromaticity of the comparison ring is not the same as the chromaticity of the test, due to the induction effects produced by the lime and purple rings that surround the test ring: the difference in the S-chromaticity (associated to the S cones) between test and comparison rings is a color shift that quantifies the induction and can be plotted against the S-chromaticity of the test ring. Monnier performed this experiment with four observers, seven test-ring chromaticities, and the two possible alternating orders for the inducing rings (lime followed by purple, or the other way round).

The resulting psychophysical data, averaged over the observers, is shown in Figure 7 as orange triangles for the purple/lime patterns and purple triangles for the

lime/purple patterns. We have optimized the parameters of the LHEI model (see Section 8.1 for details) so that when we apply it to the test and comparison images, the resulting S-chromaticity difference between test and comparison ring is as close as possible to the one reported in the psychophysical experiments. Figure 7 shows in solid lines the fits of the LHEI model, in orange for the purple/lime pattern case and in blue for the lime/purple case. As we can see the fit is quite good, and qualitatively similar to the one obtained by Song et al. [13] for the same data using a neural field model based on the Wilson-Cowan formulation.



5 Induction compensation and the invertibility of models of color appearance

As stated in the introduction we want to derive a method that *matches* induction effects among screens of different size, not a method that *estimates* induction effects and their appearance. The difference is very relevant, and it's similar to the fact that colorimetry and color spaces allow us to determine quite accurately when two colors are perceived as different or the same, but they can't tell us the perceived appearance of said colors, as there are many external factors that play a role in this; we must remark though that this approach contrasts with that of works like [28], where the output of the algorithm was explicitly simulating the appearance.

Let's say that we have a color appearance model M that is invertible and capable of reproducing induction effects. We consider two viewing scenarios A and B in which the same image stimulus I is presented on a display, and both scenarios have

identical viewing conditions except that the screen in A has different size than in B. In this study we isolate for viewing angle and its effect on color perception; it is well known that other viewing parameters, like ambient illumination, screen luminance and dynamic range, display color gamut, etc. may have a significant impact on perception, but the usual practice in the literature, given the challenges in modeling vision, is to vary one of these elements while the others are kept fixed. The model M predicts for image I an appearance $M_A(I)$ in scenario A and an appearance $M_B(I)$ in scenario B. These appearances will be different because M_A and M_B are two instances of model M that will in general have different parameter values. The reason is that, as we mentioned in Section 2, neural processes adapt to the scene statistics, and for scene we mean the whole field of view, *a part of which* is the screen where the image stimulus is displayed: therefore, different viewing angles will result in different scenes, consequently yielding different adaptation processes. In fact, in linear-nonlinear (L+NL) models of vision (and the model M we will be proposing shortly will be of this kind), adaptation is actually defined as the change of the model parameters when the input changes, and the full-view scenes in A and B provide different inputs to the visual system because the viewing angle of the screen is different.

Then, our induction matching goal can be expressed as determining the parameters for the compensation method $C = M_B^{-1} \cdot M_A$, because when the pre-processed image $C(I)$ is shown on screen B its appearance, including induction effects, will be $M_B(C(I)) = M_B \cdot M_B^{-1} \cdot M_A(I) = M_A(I)$, i.e. the same as if the image was seen on screen A . In short, having an invertible appearance model M for induction allows us to have an explicit analytical expression for C , and the parameter values for C might be found so that they match psychophysical data. Furthermore, and very importantly, we don't need to optimize M so that it accurately predicts induction effects in image appearance, which is a very challenging open problem: we just need to optimize C so that the induction effects *match* in the two conditions.

This implies, however, that neither the LHEI model nor any of the color induction models in the literature (e.g. [12, 13]) can be used for our induction compensation goal, as they are not invertible. In what follows we show how to modify the LHEI model so as to make it invertible.

In [52] the authors went back to the retinal models that were updated and analyzed in [53], studied what were their most essential elements, and produced the simplest possible form of equations to model the retinal feedback system that are nonetheless capable of predicting a number of significant contrast perception phenomena like brightness induction (assimilation and contrast) and the band-pass form of the contrast sensitivity function. These equations form a system of partial differential equations that minimize an energy functional, closely related to the one of the LHE method of [20], but where the absolute value function in the second term of Eq. 1 is raised to the power of two. This has the effect of *regularizing* the functional, making it convex, and therefore its minimum can be computed with a single convolution, while the functional in [20] is non-convex and as a consequence its minimum has to be found by the iteration of the gradient descent equation. If we modify the energy functional associated to the LHEI method (Eq. 6) following this approach

we obtain:

$$E(J) = \frac{\alpha}{2} \int_{\Omega} (J(x) - K_m * J(x))^2 dx - \gamma \int_{\Omega^2} K_c(x, y) (J(x) - J(y))^2 dx dy + \frac{\beta}{2} \int_{\Omega} (J(x) - J_0(x))^2 dx, \quad (7)$$

where as usual Ω is the rectangular domain of the image that is displayed (i.e. not the whole field of view).

The gradient descent equation associated to this functional is:

$$J_t(x) = -\alpha(J(x) - K_m * J(x)) + \gamma \int_{\Omega} K_c(x, y) (J(x) - J(y)) dy - \beta(J(x) - J_0(x)) \quad (8)$$

Now the minimum can be computed directly by convolving the input image J_0 with a kernel S :

$$S = \mathcal{F}^{-1} \left(\frac{\beta}{\alpha + \beta - \gamma - \alpha \mathcal{F}(K_m) + \gamma \mathcal{F}(K_c)} \right), \quad (9)$$

where \mathcal{F} represents the Fourier transform. The kernel S clearly has an inverse kernel S^{-1} such that $S * S^{-1} = \delta$:

$$S^{-1} = \mathcal{F}^{-1} \left(\frac{\alpha + \beta - \gamma - \alpha \mathcal{F}(K_m) + \gamma \mathcal{F}(K_c)}{\beta} \right) \quad (10)$$

We propose the following modified version of the LHEI model, also consisting of two stages:

- 1 The first stage is identical to the first stage of the LHEI model:

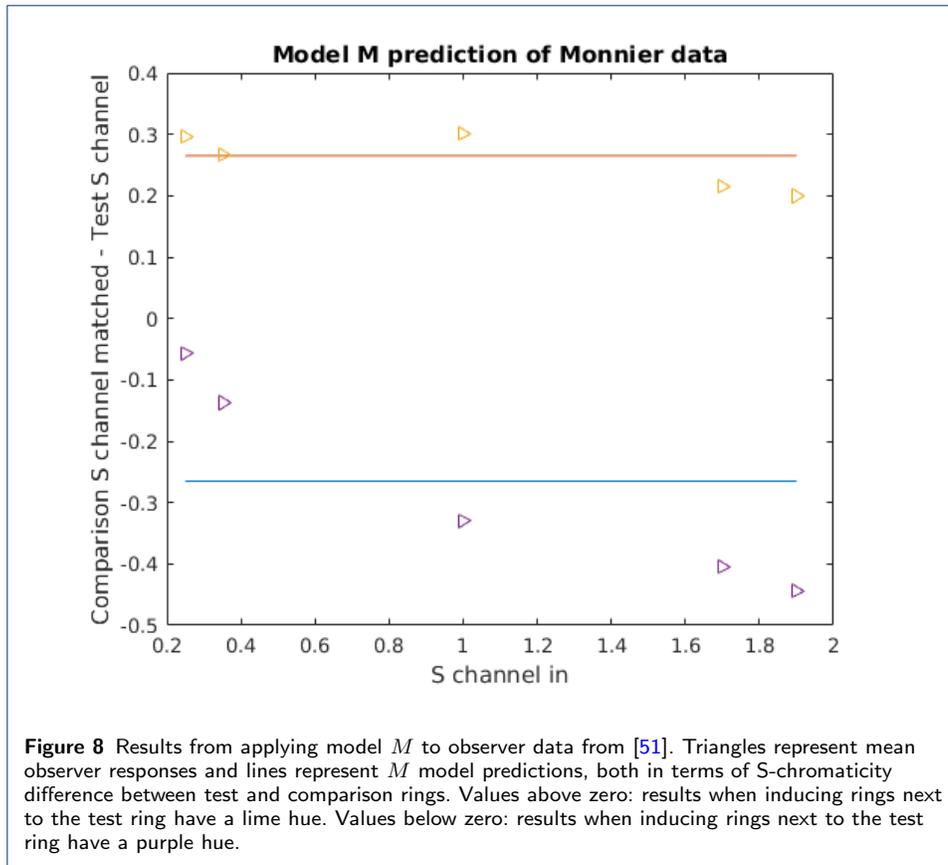
$$J = NR(I) = \frac{I^n}{I^n + I_s^n}, \quad (11)$$

where we recall that we consider I to be a scalar-valued image, $I : \Omega \rightarrow [0, +\infty)$, so $J : \Omega \rightarrow [0, 1)$.

- 2 The second stage produces the output O as the convolution of J with the kernel S of Eq. 9:

$$O = S * J \quad (12)$$

Let's call this model M . As we did for the LHEI model, we have optimized the parameters of model M so as to fit the chromatic induction data of Monnier (see Section 8.1 for details). The resulting fit is shown in Figure 8, where we see that model M also reproduces the induction effects, although with less accuracy than the LHEI model. This is consistent with the observation made in [51], that a simple linear model can't reproduce the nonlinear nature of the color shifts observed in the psychophysical experiments; and while the M model begins with a non-linear stage, its main effect is provided by the second, linear stage.



The output $O = M(I)$ can be expressed as $M(I) = S * NR(I)$. The inverse of the Naka-Rushton equation is

$$NR^{-1}(J) = I_s \cdot \left(\frac{J}{1-J}\right)^{\frac{1}{n}}, \quad (13)$$

and the inverse kernel S^{-1} was defined in Eq. 10. Therefore, the inverse of M can be expressed as $M^{-1}(O) = NR^{-1}(S^{-1} * O)$.

6 Proposed method for induction compensation

Based on model M , defined in Eqs. 11 and 12 above, we propose the following method for induction compensation for screens of different size.

If an image I is to be shown on screen B producing the same induction effects as if it were shown on screen A , in both cases under the same viewing conditions, then a compensation method C must be applied to the image I , yielding an image $C(I)$. When $C(I)$ is displayed on screen B the induction effects are the same as when I is displayed on screen A . The compensation method C is:

$$C(I) = M_B^{-1}(M_A(I)) = NR_B^{-1}(S_B^{-1} * S_A * NR_A(I)) \quad (14)$$

The linear filter S of model M has a center-surround form that, as mentioned in Section 3, can perform decorrelation and contrast enhancement. For images with

very high contrast, convolution with S_A could produce overenhancement, resulting in some undershoot or overshoot values falling outside the range $[0, 1)$, and in some cases these values might still remain out of range after convolution with S_B^{-1} , making it impossible to apply to them the function NR_B^{-1} because its domain is $[0, 1)$. To prevent these issues, in practice we clip all out-of-range values of $S_B^{-1} * S_A * NR_A(I)$ so that negative values are set to 0 and values greater than 1 are set to 1; nonetheless, it is not expected that this clipping procedure produces visible artifacts, as attested by the natural image examples in Figure 13.

In order to validate our method we consider a scenario where A corresponds to a cinema screen and B to a mobile display. We perform psychophysical experiments for both achromatic and chromatic images where observers look at a display with two scales of the same image, and they have to adjust the values of a given region of the small scale image (corresponding to the mobile viewing scenario) so that it matches the appearance of that region on the large scale image (corresponding to the cinema viewing condition). For cinema, three picture heights viewing distance (a common figure for mastering) is assumed resulting in a vertical viewing angle of 18.92° ; for the mobile condition, the same viewing angle as in [54] is used resulting in a scaling factor of 0.39 between the two viewing scenarios.

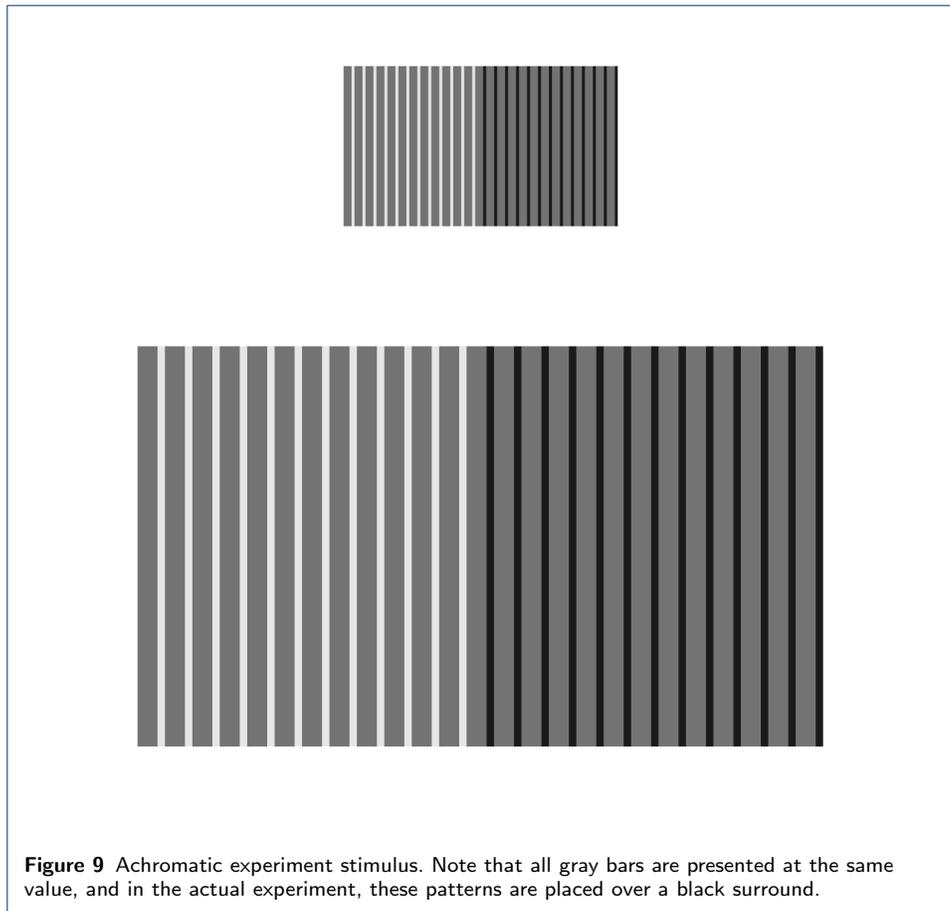
6.1 Achromatic case

These experiments follow the approach used by Helson [1], in which observers had to judge the appearance of grey bars over white or black backgrounds. In our case, though, the stimuli is emissive (images are seen on a display, instead of being printed on paper) and observers independently adjust the luminance of the gray bars on each side background (white/black) so that they match the appearance of the corresponding gray bars on the larger reference, the cinema-scale image; see Figure 9.

The experiment was conducted with ten observers and with five different widths for the gray bar stimuli. Figure 10 shows the average observer responses in the experiment and the prediction provided by our model (see Section 8 for details on the experimental set-up and the optimization). We can see how the observer results are consistent with those of [1] in two key points: firstly, when the visual angle (equivalently the line width) decreases, the appearance tends to assimilation, and hence the compensation requires enhancing the contrast; secondly, as the visual angle increases, the amount of necessary compensation should decrease. Our model responses are consistently inside the confidence intervals for the observer data.

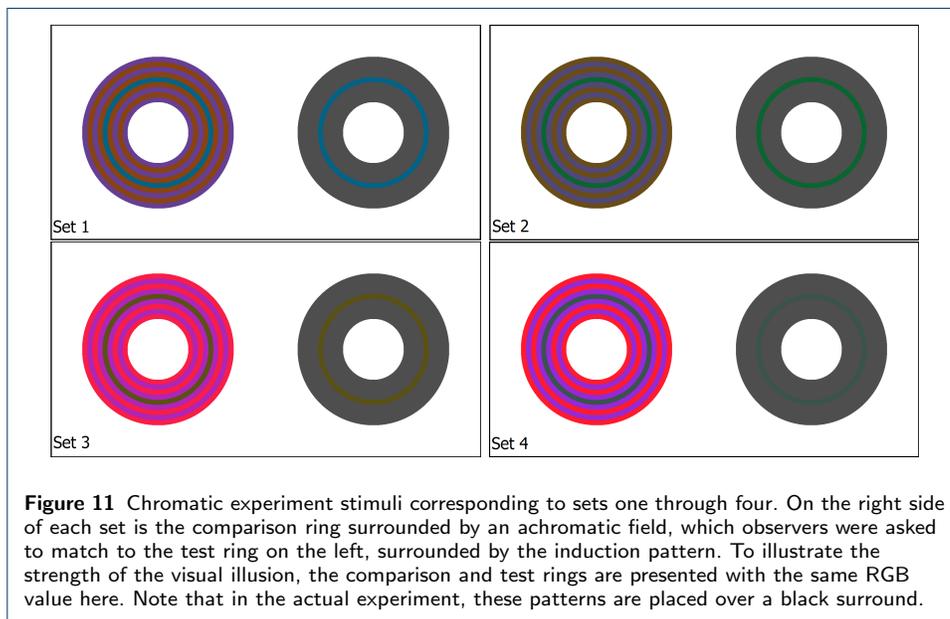
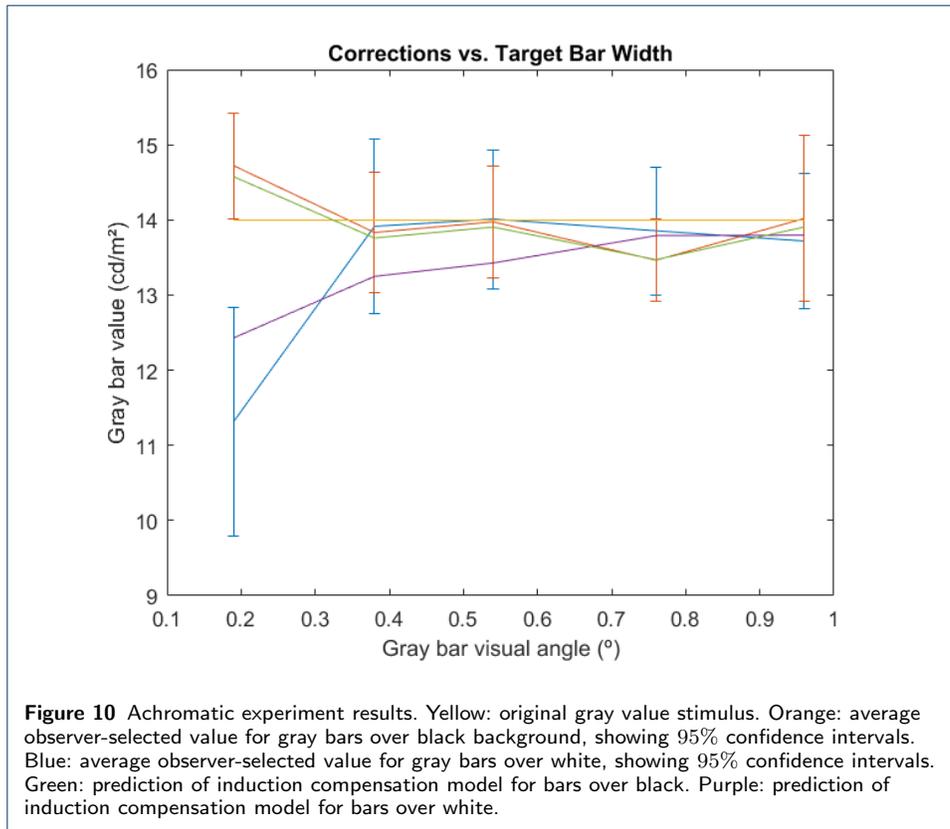
6.2 Chromatic case

For the chromatic experiments we took inspiration from [55] and used concentric circular induction patterns as shown in Figure 11. Observers must adjust the tristimulus values of the central ring of the comparison pattern (the achromatic circular ring on the right side of each set) so that it matches the appearance of the central ring of the test pattern (the concentric circular pattern on the left side of each set). This procedure was repeated for patterns at mobile and cinema scaling settings, and the observer reported correction was found by taking the difference between responses (*cinema* – *mobile*).



We performed these experiments for four different concentric ring patterns, then optimized our induction compensation model so that it fits the data for three of these images and finally validated our results on the remaining image; see Section 8 for details on the experimental procedure and optimization method. In Figure 12, the results of the chromatic experiment are plotted in the two dimensional a^*b^* plane. The results are limited to the chroma channels, as the corrections reported by observers in the L^* dimension were not statistically significant (95% confidence error ranges overlapped the origin for all tested cases.) For each of the patterns the origin of the coordinate system is placed at the starting a^*b^* value of the test ring. The plots depict the value of the inducing rings with a blue vector for the value of the inducing ring that is closer to the test ring, which we call the first inducer, and a red vector for the value of the other inducing rings, that we call the second inducer. The average observer response is depicted with purple and green 95% confidence error bars.

Looking at the observers' responses, the induction compensation results selected by observers tend to show contrast mainly in the direction opposite to the first inducer, which implies that the appearance of the mobile viewing condition shows assimilation in the direction of the first inducer (because assimilation is compensated by contrast). In this way, the results for sets two through four were consistent with the classic assumptions on induction, as well as the results of [1, 4, 51]. However,

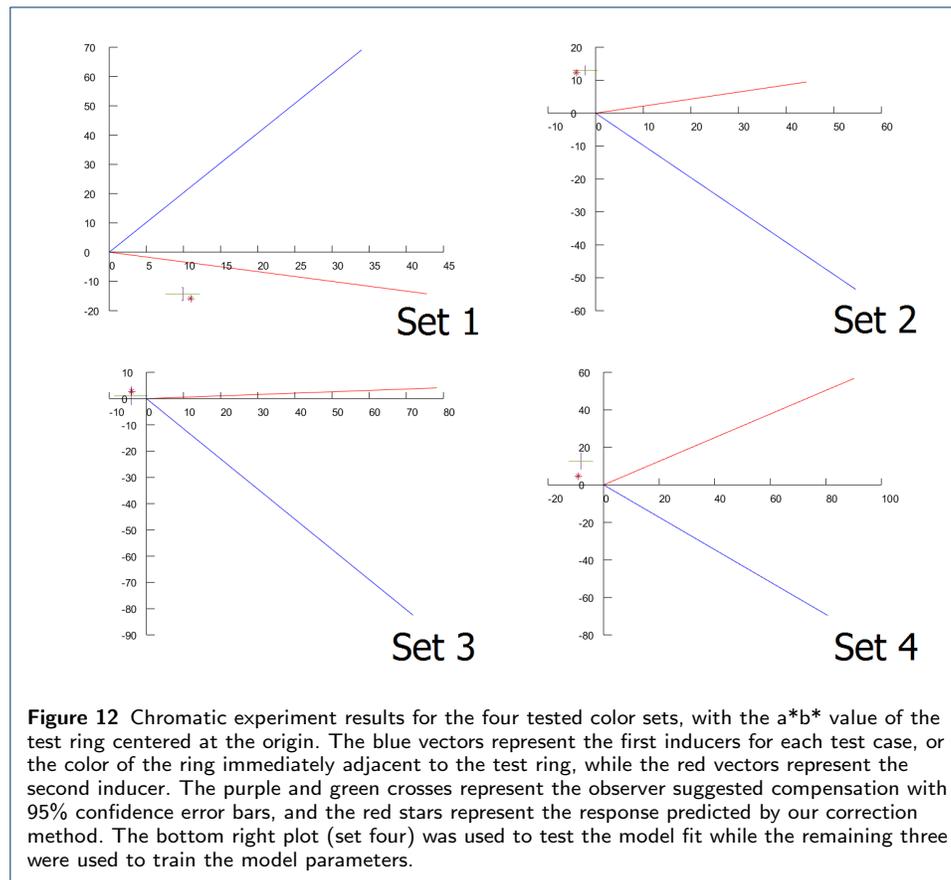


our first set shows that this cannot be taken as a general rule, as observers reported the necessary correction to be roughly in the assimilation direction of the second inducer.

Regarding the ability of our method for fitting this data, in Figure 12 we added our results when we train using sets 1-3 and we test in set 4. Our resulting corrections

predicted by the method for each color set are depicted with a red star. We can see for the train cases that the kernel prediction is within the range of experimental error, and for all cases compensates input in the proper direction.

To further study our model, Table 6.2 shows the error (measured as ΔE difference) between the average observer response and our model's prediction. As explained above, given that we have four different sets, we perform our experiments by training in 3 of the sets and testing in the remaining one. This gives us 4 different cases. In the table, columns 2 to 5 represent each of the cases, with the model error for the testing set shown in blue; in particular, column 5 corresponds to the case illustrated in Figure 12. Column 6 presents in red what we call the 'original' error, the ΔE difference between the original data and the result of the observer correction. Finally, column 7 shows the improvement that our method presents over the original error, which is in the range [45% – 60%], therefore highlighting the advantage of applying our compensation method instead of doing nothing and just re-scaling the original image.



6.3 Example results on natural images

Figure 13 illustrates the results of our induction compensation method on some natural images. The original images are shown on the left, and on the right are the results of applying our method (in this case, with the parameters optimized for sets 1 to 3, as in Figure 12, and corresponding to column 5 in Table 1.)

	Train on: 2-4 Test on: 1	Train on: 1,3,4 Test on: 2	Train on: 1,2,4 Test on: 3	Train on: 1-3 Test on: 4	Original error	Improvement (on test set)
Set 1	8,92	1,58	2,84	1,87	17,39	48,70%
Set 2	1,50	5,35	2,84	1,95	13,20	59,47%
Set 3	1,61	1,58	1,67	1,65	4,29	61,05%
Set 4	1,17	1,46	2,79	8,18	15,00	45,48%

Table 1 Error between the average observer response and our model’s prediction. We have performed the training for all combinations of 3 sets, testing on the remaining set (columns 2-5). Column 6 represents the original error, and column 7 represents our improvement, w.r.t. the original error, in the test set.

We can see how in our results the colors are subtly but noticeably more vivid, e.g. the orange cone in the 1st row, the green teapot in the 3rd, the kid’s blue jacket and boots and the grass in the 4th row, the yellow fish in the bottom row. This increased vividness corresponds to a contrast enhancement in the chroma, which will be cancelled out by the visual assimilation (and resulting chroma contrast reduction) produced when observing the image under a smaller field of view; the relationship between contrast enhancement and more vivid colors is discussed in detail in [56, 57, 35].

While these results do not show visual artifacts of any kind, these problems can’t be ruled out as they might appear if the method’s parameters are optimized differently and/or the method is tested on other images.

7 Discussion

The primary goal of these experiments was to determine the correction required to match the appearance of induction pattern targets at two different field of view scales. Based on the results of [1, 4], we took the simple hypothesis that a greater degree of contrast would always be observed in the larger field of view pattern. Thus, the correction from small pattern to large for a given channel should always be in the direction of contrast, and the results of the achromatic experiment confirmed our hypothesis.

For the chromatic experiments, making the assumption that the phenomenon of induction occurs after visual signals are separated into different visual pathways, we chose to make color matches in an opponent space with the intention of applying our corrective method to the channels separately. We found that the use of the simple bar patterns of [4] caused a multitude of problems in the chromatic case. Observers reported weak induction effects as well as strong afterimages when shifting their gaze between test patterns. In addition, the direct comparison of the cinema-sized pattern to the mobile pattern was confusing to observers, as the inducers in the smaller pattern appeared to be significantly less saturated. As a solution, we took inspiration from [55] and used concentric circular induction patterns, as shown in Figure 11, whose main features are that their circular shape results in less afterimages when compared to the bars, and their use of dual inducing colors leads to a stronger effect, allowing for more significant results to be gleaned from the experiment.

In order to find patterns which exhibited a strong inductive effect, an experiment was performed in which 100 patterns containing regions with randomly selected $L^*a^*b^*$ values within the Rec. 709 color gamut were generated (a color gamut is the set of colors that a display can reproduce, and Rec. 709 is the default gamut specification most commonly observed by display and television manufacturers).



Figure 13 Left: original Images. Right: results of our method with parameters optimized for training sets 1-3. Notice in our results how colors are slightly more vivid, and the absence of visual artifacts. Images sourced from [58]

These patterns were then shown side by side at the different scaling factors tested in the experiment. Then, patterns for a which a hue shift could be identified between the different scaling factors were singled out. Then, the experiment procedure was conducted for a single observer using all selected patterns from the previous step. From these results the final patterns were selected based on the criteria that the sensation of the target ring could be successfully reproduced in isolation, given the gamut of the monitor and that a statistically significant correction (given 95% confidence intervals) was called for by the observer between the two test pattern scaling factors.

After several iterations of the experiment, six total color sets were found. Administering the test to multiple observers revealed that two of the sets should be removed, as the target colors were too close to the gamut boundary for observers to make reliable observations. While this experiment was more or less informal in nature, its results demonstrate the rarity of strong induction effects given random color combinations, even if they are arranged in synthetic patterns which emphasize induction. Given that adjacent color combinations in the real world are less than random [59], it would be interesting to determine if patterns featuring adjacent color sets sampled from real images would lead to a higher or lower frequency of detectable illusions. Another interesting anomaly which can be observed from this experiment is that despite the random nature in which they were generated and selected, the final four patterns appear quite similar to each other, all containing an inducer of violet hue.

The results of the chromatic experiments presented here clearly show that the original hypothesis (that contrast effects will shift towards assimilation with an increase in test pattern spatial frequency) can be broken. While three of the color sets showed this behaviour, we can see that the first color set breaks the trend, and is closer to requiring correction in the assimilation direction with respect to the second inducer. One pattern which may provide a clue to this differing behavior is related to the violet inducers which appear in each pattern. In all patterns for which induction effects behaved as expected, the violet inducer was directly adjacent to the test and was the primary induction influence. However, for the first color set, the violet field serves as the second inducer which is not directly adjacent to the test field, but still acts as the primary induction influence.

Outside of this data, we also found patterns which broke our simple hypothesis in preliminary experiment iterations. From these iterations we observed that the luminance level of the background/surround and the hues of inducing and target patches to be relevant factors. This type of conflicting and paradoxical finding seems to be common in the study of induction, with many works being based on the discovery of scenarios which contradict previous findings [51, 60]. While this phenomenon can be found in all research topics and is a sign of progress, its frequency in this area is an indication that induction as a whole is still very much an open problem, despite its earliest formal works dating back nearly a century and a half. This can be justified by the fact that the phenomenon is the result of complex interactions involving both physiological and cognitive processes [61] on multiple visual pathways. It is this fact which makes it such an interesting topic in vision science, as to understand it thoroughly would be a great contribution to understanding the behavior of the human visual system.

While the model was designed with the intention that it would always correct the test targets in the contrast direction, we have had some success in optimizing a kernel which works more generally. These results were garnered by model fitting using three chromatic sets and testing on a fourth one. We have found that a different choice of optimization and validation sets produce somewhat different results, due to the optimization procedure falling in local minima. Due to this fact, and that induction data was gathered in a limited range of synthetic scenarios, the compensation method presented here is to be interpreted as a proof of concept and a work in progress as opposed to a procedure which is ready to be used in practical applications.

A further interesting challenge is that the compensation value observers reported to adjust between mobile to cinema appearance could be outside of any given monitor gamut space, or outside of the gamut of physically realizable colors, depending on the position of the test target and the magnitude and direction of the induction shift between screen sizes. In these scenarios, induction effects will only be partially compensated for by the method. We encountered this issue with three of our four test sets, and opted to clip all observer and kernel reported corrections to the Rec. 709 gamut. By doing this our model's results can be readily reproduced by the most common displays, including our own, which we used for visual proofing during its development. We later performed a preliminary analysis with input encoded under the larger standard color gamuts Rec. 2020 and CIE 1931 XYZ, where the clipping of observer corrections is smaller for the former and almost negligible for the latter. The results showed that the method makes corrections of similar accuracy when it is required to reach out into larger color volumes.

We believe there are three main avenues to explore in order to improve our proposed approach:

- 1 Our induction compensation technique is based on a color appearance model that follows the classic formulation of a cascade of linear-nonlinear (L+NL) modules [62] and has a biological correlate, consisting of a nonlinear stage (the Naka-Rushton equation that models photoreceptor responses) followed by a linear stage (convolution with a kernel that models lateral inhibition in the retina). A L+NL model is valid for stimuli of a given distribution seen under given viewing conditions, in which case it may provide a good match to the firing rate. But visual adaptation, an essential feature of the neural systems of all species by which changes in the stimuli produce a change in the input-output relation of the system [63], alters the visual system response. Visual adaptation is clearly a key element of the efficient representation principle and it affects, among other things, the spatial receptive field and temporal integration properties of neurons, requiring changes in the linear and/or the nonlinear stages of a L+NL model in order to explain neural responses [37]. So for example depending on the input the receptive field of a single neuron can have different sizes or preferred orientations [64], or even change polarity (ON/OFF) [65]. For our purposes of induction compensation, we should study how to make the convolution kernel S depend on the input, possibly designing a filter bank as is the traditional approach with L+NL models for visual perception [66, 67].

- 2 Another option is to study how to make the LHEI model invertible while keeping it as a nonlinear neural field model (i.e. without regularizing its associated functional), looking into a gradient *ascent* equation or alternatively considering changing the sign of the parameter γ in the model, as it has been shown that with one sign for γ the model increases the contrast while with the opposite sign the model reduces the contrast [68, 57, 69]. We believe this option has more potential because the resulting compensation model would not be of L+NL form.
- 3 Finally, a third avenue to explore, compatible with the previous two, would be to design and carry out psychophysical experiments for induction compensation where observers are asked to adjust values over the whole image and not just on a particular region like the gray bars or the test ring. Using this data, the correction method could be optimized such that it produces a balanced correction for all of the spatially adjacent regions in the patterns simultaneously, accounting for their interdependent effects.

8 Methods

8.1 Monnier data

For the minimization of Monnier’s data, we follow the same idea as in the Song *et al.* paper [13]. This means that, for each of the initial conditions, we run our method using both the original image (composed by the original rings), and the comparison image adjusted by observers (a single ring on an achromatic background) as input. Then, our minimization looks at the difference between the test ring in these two images. In other words, we look for the parameters of our method that produce the same internal representation to both of the images. The error between the two images is computed as the L_2 difference between the value of the central rings. Finally, the error for each of the initial conditions is summed up to obtain the total error to minimize. All our values in the minimization are related to the size of the input images, which is 1500 by 1500.

8.1.1 LHEI

The parameter values obtained for this model are $\alpha = 3.4809$, $\beta = 0.0257$, $\gamma = 0.6259$. The kernels are $K_c = k_1 \cdot (G_{507} + G_{507})$, $K_m = k_2 \cdot (G_{300} + 2G_{50})$, where we denote as G_N a Gaussian with standard deviation of N pixels, and the scalar values k_1, k_2 are normalization constants so that the integral of K_c and K_m is 1 in both cases. The sigmoid σ is approximated to $\sigma(z) = z^p$ when $z > 0$, and to $\sigma(z) = -|z|^q$ otherwise, where $p = 0.503$, $q = 0.544$.

8.1.2 Model M

For this minimization we write the kernel S as $S = \mathcal{F}^{-1}(1/(C_2 + C_1\mathcal{F}(K)))$, where K is a weighted sum of the four Gaussians given by K_c and K_m . The values obtained for this model are: $K = 0.99G_{125} + 0.81G_{25} + 3.38G_5 + 0.82G_{27}$, $C_1 = 3.21$, and $C_2 = 1.31$. As a note, we ensure that the mean value after the convolution equals the mean value before the convolution.

8.2 Psychophysical experiment

8.2.1 Apparatus and stimuli

Experiments were conducted under dark surround viewing conditions on a calibrated Sony PVM-A250 reference monitor, representing an easily controllable cinema-like viewing environment. The monitor was calibrated to Rec. 709 primaries with a D65 white point and was verified routinely before experimental sessions using a Klein K10-A colorimeter. The experimental cadence was controlled by a MATLAB test bed using the Psychophysics toolbox [70, 71] to display stimuli. Observers input their responses via a Tangent Element color correction panel which allowed for multi-channel adjustments to stimuli with separate knobs allowing for a more natural and reactive experimental interface in comparison to keyboard input.

The test patterns were selected based on the criteria that they evoked a large degree of induction, such that a significant corrective factor could be found in spite of the potentially large observer response variability, and that the target color had significant room for adjustment considering the limited gamut of the experimental display. The scaling factor between mobile and cinema conditions was determined as follows. For cinema, three picture heights viewing distance (a common figure for mastering) is assumed resulting in a vertical viewing angle of 18.92° . For the mobile condition, the same viewing angle as in [54] is used resulting in a scaling factor of 0.39 between the two viewing scenarios.

The observer task in each experiment was that of asymmetric matching, with the goal of determining the necessary corrective factor between induction effects viewed at different field of view scaling factors. To accomplish this, observers were presented in each trial with a pair of patterns, which we call test and comparison. The test pattern serves as a reference, and contains a field (the test field) which is to be matched to an analogous region in the comparison pattern (the comparison field). At the start of each experiment, observers were seated at a three picture heights viewing distance from the screen in order to achieve the experimental viewing angles determined above and the experimental instructions were read to them aloud. These instructions described the experimental task, cadence, and control scheme. Observers were then allowed to start the experiments, starting with a practice trial. Each experiment took observers an average of 20 minutes to complete.

8.2.2 Achromatic experiment

Following the preceding work [72], the initial experiment was intended to be a direct expansion of the experiments of [1] for the case of emissive stimuli. To this effect, we used the same type of induction pattern with a fixed inducer bar width and varied the target bar width. In this case however, observers reported the necessary correction factor directly by adjusting the luminance of the target bars in the mobile scaling (comparison) to match those in the cinema scaling (test). The additional variable of starting comparison pattern target bar luminance was also varied between experimental presentations such that observers could approach their response from different directions. The complete matrix of experimental factors is shown in Table 2.

Table 2 Achromatic experimental factors. Visual angles correspond to the cinema size patterns

Factor	Type	Levels
Target width	variable	0.19°, 0.38°, 0.54°, 0.76°, 0.96°
Initial comparison luminance	variable	4.0, 8.1, 22 cd/m^2
Inducing bar width	constant	0.19°

Stimuli Figure 9 shows the presented stimuli for the achromatic experiment. As can be seen in the figure, two patterns were presented on screen (except over a black surround, as opposed to the white surround they are presented with in the figure). Each pattern consists of two sides, representing positive (white/gray) and negative contrast (black/gray) respectively. Observers adjusted these two sides separately, but both are included simultaneously such that lightness references remain constant. The white fields were presented just below the maximum monitor white at a value of $90 cd/m^2$, while the black fields were presented at a value of $0.6 cd/m^2$.

Observers Ten observers (2F,8M) aged between 23-39 took the experiment. All observers had normal or corrected acuity (20/20). Three observers are authors, while the remaining seven were naive to the purpose of the study.

8.2.3 Chromatic experiment

Based on evidence that the phenomenon of induction occurs after visual signals are separated into different visual pathways, we expand on the achromatic experiments by making color matches in an opponent channel space, with the intention of applying our corrective method to the channels separately. In this case, we chose CIE $L^*a^*b^*$ space due to it having some degree of perceptual uniformity. In initial experiments we found that this expansion to three channel adjustment caused a great increase in the difficulty of the experimental task. Thus, we simplified the procedure by reducing the number of variables. In this case we test four color sets in the mobile and cinema sizes, and we test for three different test ring starting colors. To avoid observer fatigue, experiments were conducted two color sets at a time. The complete matrix of experimental factors is shown in Table 3.

Table 3 Chromatic experimental factors. Test pattern element sizes are relative to the cinema condition, but their size in proportion to each other is preserved for the mobile case.

Factor	Type	Levels
Pattern scaling	variable	cinema and mobile (39%) scaling
Initial test level	variable	original, +10 a^*b^* , -10 a^*b^* cd/m^2
Test pattern colors	variable	sets 1-4
Test pattern diameter	constant	11.0°
Test to comparison distance (center to center)	constant	15.6°
Pattern center diameter	constant	4.39°

Stimuli Using the patches shown in Figure 11, observers were given the task of adjusting the comparison ring, seen on the left surrounded by an achromatic field to the test ring, which can be seen on the right surrounded by the induction pattern. Observers repeated experiment trials with the patterns shown at mobile and cinema scaling factors, and the corrective results were derived by taking the difference between their responses in these separate conditions.

The two relevant features of these patterns are that their circular shape results in less after-images when compared to the bars, and their use of dual inducing

colors leads to a stronger induction effect, allowing for more significant results to be gleaned from the experiment. The L^* value of all rings in these patterns is kept consistent such that the focus of the observers' task could be on correction for chromatic induction. This said, observers were still permitted to adjust the L^* channel value as equiluminance between pattern regions was not confirmed.

Observers For the second color set, four observers participated in the experiment (1F, 3M) and for the remaining three color sets, three observers participated (3M). In both cases observer age ranged between 23 and 36, and two observers are authors while the remainder were naive to the purpose of the study. All observers had normal or corrected acuity (20/20).

8.3 Model fitting and validation

Following Equation 14, our goal is to fit the two exponents of the Naka-Rushton equations and the two convolutions S_B^{-1} and (S_A) . Following the approach used in Section 8.1.2 to represent kernel S , we rewrite the minimization in terms of what we call the compensation kernel S_C , $S_C = S_B^{-1} * S_A$:

$$S_C = S_B^{-1} * S_A = \mathcal{F}^{-1} \left(\frac{D_2 + D_1 \mathcal{F}(K_F)}{C_2 + C_1 \mathcal{F}(K_F)} \right) \quad (15)$$

where K_F is the weighted sum of four Gaussians, and C_1 , C_2 , D_1 , D_2 are real numbers.

We use the cone-space representation of CAT02 LMS space [73] and apply the first Naka-Rushton equation to individual L, M, and S channels. These values are then transformed to an opponent color representation, with channels that we call Y, op1 and op2, and are calculated as follows: $Y = L+M+S$, $op1 = L-M$, and $op2 = 2S-(L+M)$. The method output is computed by convolution with the kernel S_C , followed by the inverted opponent channel transformation, clipping to the range $[0, 1)$ and finally by the inverse of the second Naka-Rushton equation (see Equation 14). As discussed below, the parameters for the S_C kernel and the two Naka-Rushton exponents n_A, n_B are found by minimizing the error between the observer data and the method results.

Two practical issues to comment regarding our implementation are the following. First, we impose that the mean value in each channel after the convolution with S_C should equal the mean of the channel before the convolution. In this way, we avoid the introduction of colour shifts in our results. Second, our method is working in a color opponent space different from $L^*a^*b^*$. For this reason, when convolving kernel S_C with our chromatic channels, shifts in the perceived brightness (and L^* channel value) may occur. To better comply with the observer responses, which reported no L^* correction to be necessary, we decided to replace the L^* channel of our result by the L^* channel of the original image.

8.3.1 Luminance Correction

The luminance correction method was optimized to fit the achromatic experiment data as generally as possible, meaning that the mean observer response from each target bar width experiment were considered simultaneously in the error function.

By optimizing in this way, the method could be made to work generally for different spatial configurations.

The values obtained for this experiment following the above procedure are:

$n_A = 0.7861$, $n_B = 0.7063$, $K_F = -1.14G_{156} + 1.86G_{29} + 0.13G_3 - 1.76G_{40}$, $C_1 = 3.94$, $C_2 = 2.54$, $D_1 = 2.46$, $D_2 = 2.72$. These values have been obtained in relation to images of size 800 by 800.

8.3.2 Chrominance Correction

We apply the kernel S_C only to the two chromatic components of our opponent channel space. The optimization is performed in order to minimize the *Delta E error on the test ring, and as we are using three different sets for training the minimization considers the maximum value of the Delta E error on the three test rings.*

The values obtained for the case where set 4 is used for testing (corresponding to results in Figure 12 and column 5 in Table 6.2) are: $n_A = 0.5187$, $n_B = 0.4439$, $K_F = -1.53G_{103} - 0.67G_{43} + 0.67G_4 + 0.34G_{26}$, $C_1 = 2.81$, $C_2 = 1.30$, $D_1 = 2.27$, $D_2 = 1.60$. Let us note that these values have been obtained in relation to images of size 800 by 800.

Acknowledgements

The authors would like to thank those who served as experimental observers, who this work would not have been possible without. This work has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement number 761544 (project HDR4EU) and under grant agreement number 780470 (project SAUCE), and by the Spanish government and FEDER Fund, grant ref. PGC2018-099651-B-I00 (MCIU/AEI/FEDER, UE).

Abbreviations

LGN - lateral geniculate nucleus
 LUT - look up table
 PDE - partial differential equation
 LHE - local histogram equalization
 RF - receptive fields
 RGC - retinal ganglion cells
 DoG - difference of gaussians
 L+NL - linear-nonlinear

Declarations

8.4 Ethics approval and consent to participate
All observers in the experiment gave consent to their participation

8.5 Consent for publication
Not applicable

8.6 Availability of data and materials
The test patterns used in the chromatic and achromatic experiments and the raw responses of observers are included in the submission.

8.7 Competing interests
The authors declare that they have no competing interests.

8.8 Funding
This work has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement number 761544 (project HDR4EU) and under grant agreement number 780470 (project SAUCE), and by the Spanish government and FEDER Fund, grant ref. PGC2018-099651-B-I00 (MCIU/AEI/FEDER, UE).

8.9 Author's contributions
TC planned and conducted psychophysical experiments and participated in writing the paper. JVC led the method optimization effort and participated in writing the paper. EM assisted in the method optimization effort. MB created the paper's concept, supported the research in an advisory role and participated in writing the paper. All authors read and approved the final manuscript.

Author details

¹Department of Information and Communication Technologies, Universitat Pompeu Fabra, 08018 Barcelona, Spain.

²Grenoble INP - Phelma, Grenoble, France.

References

1. Helson H. *Studies of Anomalous Contrast and Assimilation*. *Journal of the Optical Society of America*. 1963;53(1).
2. von Bezold W. *Die Farbenlehre*. Braunschweig: Westerman; 1874.
3. Gelb A. *Handbuch norm. pathol. physiol.* 1930;12:594.
4. Fach C, Sharpe LT. Assimilative hue shifts in color gratings depend on bar width. *Perception 'I&' Psychophysics*. 1986;40(6):412–418.
5. Brenner E, Ruiz JS, Herraiz EM, Cornelissen FW, Smeets JB. Chromatic induction and the layout of colours within a complex scene. *Vision Research*. 2003;43:1413–1421.
6. Brown RO, MacLeod DIA. Color appearance depends on the variance of surround colors. *Current Biology*. 1997;7:844–849.
7. Harrar M, Vienot F. Regulation of chromatic induction by neighboring images. *Journal of the Optical Society of America*. 2005;22:2197–2206.
8. Monnier P, Shevell SK. Large shifts in color appearance from patterned chromatic backgrounds. *Nature Neuroscience*. 2003;6:801–802.
9. Shevell SK, Wei JP. Chromatic induction: Border contrast or adaptation to surrounding light? *Vision Research*. 1998;38:1561–1566.
10. Shevell SK, Monnier P. Color shifts from S-cone patterned backgrounds: Contrast sensitivity and spatial frequency selectivity. *Vision Research*. 2005;45:1147–1154.
11. Wesner MF, Shevell SK. Color perception within a chromatic context: Changes in red/green equilibria caused by noncontiguous light. *Vision Research*. 1992;32:1623–1634.
12. Otazu X, Parraga CA, Vanrell M. Toward a unified chromatic induction model. *Journal of Vision*. 2010;10(12)(5):1–24.
13. Song A, Faugeras O, Veltz R. A neural field model for color perception unifying assimilation and contrast. *PLOS Computational Biology*. 2019 Jun;15(6):e1007050.
14. Attneave F. Some informational aspects of visual perception. *Psychological review*. 1954;61(3):183.
15. Barlow HB. Possible principles underlying the transformation of sensory messages. *Sensory communication*. 1961;1:217–234.
16. Smirnakis SM, Berry MJ, Warland DK, Bialek W, Meister M. Adaptation of retinal processing to image contrast and spatial scale. *Nature*. 1997;386(6620):69.
17. Kohn A. Visual adaptation: physiology, mechanisms, and functional benefits. *Journal of neurophysiology*. 2007;97(5):3155–3164.
18. Atick JJ, Redlich AN. What does the retina know about natural scenes? *Neural computation*. 1992;4(2):196–210.
19. Olshausen BA, Field DJ. Vision and the coding of natural images. *American Scientist*. 2000;88(3):238–245.
20. Bertalmio M, Caselles V, Provenzi E, Rizzi A. Perceptual Color Correction Through Variational Techniques. *IEEE Transactions on Image Processing*. 2007;16:1058–1072.
21. Sapiro G, Caselles V. Histogram modification via differential equations. *Journal of Differential Equations*. 1997;135:238–266.
22. Bertalmio M. *Image Processing for Cinema*. CRC Press, Taylor & Francis; 2014.
23. Bertalmio M, Cowan J. Implementing the Retinex algorithm with Wilson-Cowan Equations. *Journal of Physiology, Paris*. 2009;.
24. Bressloff PC, Cowan JD, Golubitsky M, Thomas PJ, Wiener MC. What Geometric Visual Hallucinations Tell Us about the Visual Cortex. *Neural Computation*. 2002;14(3):473–491.
25. Wilson HR, Cowan JD. Excitatory and inhibitory interactions in localized populations of model neurons. *Biophysical journal*. 1972;12:1–24.
26. Wilson HR, Cowan JD. A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue. *Biological Cybernetics*. 1973;13(2):55–80.
27. Cowan JD, Neuman J, van Drongelen W. Wilson–Cowan equations for neocortical dynamics. *The Journal of Mathematical Neuroscience*. 2016;6(1):1.
28. Bertalmio M, Calatroni L, Franceschi V, Franceschiello B, Gomez-Villa A, Prandi D. Visual illusions via neural dynamics: Wilson-Cowan-type models and the efficient representation principle. *Journal of Neurophysiology*. 2020;.
29. Kingdom FA. Lightness, brightness and transparency: A quarter century of new ideas, captivating demonstrations and unrelenting controversy. *Vision Research*. 2011;51(7):652–673.
30. Murray N, Vanrell M, Otazu X, Parraga CA. Low-Level Spatiochromatic Grouping for Saliency Estimation. *IEEE Trans Pattern Anal Mach Intell*. 2013;35(11):2810–2816.
31. Purves D, Wojtach WT, Howe C. Visual illusions: An Empirical Explanation. *Scholarpedia*. 2008;3(6):3706. Revision #89112.
32. Bertalmio M, Calatroni L, Franceschi V, Franceschiello B, Prandi D. Cortical-inspired Wilson-Cowan-type equations for orientation-dependent contrast perception modelling. *Journal of Mathematical Imaging and Vision*. 2020;.
33. Shevell SK. Color appearance. In: *The Science of Color*. Elsevier Science Ltd; 2003. p. 149–190.
34. Bertalmio M. From image processing to computational neuroscience: a neural model based on histogram equalization. *Frontiers in computational neuroscience*. 2014;8:71.
35. Bertalmio M. *Vision Models for High Dynamic Range and Wide Colour Gamut Imaging: Techniques and Applications*. Academic Press; 2019.

36. Shapley R, Enroth-Cugell C. Visual adaptation and retinal gain controls. *Progress in retinal research*. 1984;3:263–346.
37. Meister M, Berry MJ. The neural code of the retina. *Neuron*. 1999;22(3):435–450.
38. Valeton Jm, Norren DV. Light adaptation of primate cones: An analysis based on extracellular data. *Vision Research*. 1983;23(12):1539–1547.
39. Dunn FA, Rieke F. The impact of photoreceptor noise on retinal gain controls. *Current opinion in neurobiology*. 2006;16(4):363–370.
40. Kane D, Bertalmío M. System gamma as a function of image-and monitor-dynamic range. *Journal of vision*. 2016;16(6):4–4.
41. Ozuysal Y, Baccus SA. Linking the computational structure of variance adaptation to biophysical mechanisms. *Neuron*. 2012;73(5):1002–1015.
42. Kremkow J, Jin J, Komban SJ, Wang Y, Lashgari R, Li X, et al. Neuronal nonlinearity explains greater visual spatial resolution for darks than lights. *Proceedings of the National Academy of Sciences*. 2014;p. 201310442.
43. Turner MH, Rieke F. Synaptic rectification controls nonlinear spatial integration of natural visual inputs. *Neuron*. 2016;90(6):1257–1271.
44. Turner MH, Schwartz GW, Rieke F. Receptive field center-surround interactions mediate context-dependent spatial contrast encoding in the retina. *bioRxiv*. 2018;p. 252148.
45. Whittle P. Brightness, discriminability and the “crispning effect”. *Vision research*. 1992;32(8):1493–1507.
46. Kane D, Bertalmío M. A reevaluation of Whittle (1986, 1992) reveals the link between detection thresholds, discrimination thresholds, and brightness perception. *Journal of vision*. 2019;19(1):16–16.
47. Rucci M, Victor JD. The unsteady eye: an information-processing stage, not a bug. *Trends in neurosciences*. 2015;38(4):195–206.
48. Yeonan-Kim J, Bertalmío M. Retinal lateral inhibition provides the biological basis of long-range spatial induction. *PLoS one*. 2016;11(12):e0168963.
49. Wilson HR. A neural model of foveal light adaptation and afterimage formation. *Visual neuroscience*. 1997;14(3):403–423.
50. van Hateren H. A cellular and molecular model of response kinetics and adaptation in primate cones and horizontal cells. *Journal of vision*. 2005;5(4):5–5.
51. Monnier P. Standard definitions of chromatic induction fail to describe induction with S-cone patterned backgrounds. *Vision Research*. 2008;48:2708–2714.
52. Kim J, Batard T, Bertalmío M. Retinal processing optimizes contrast coding. *Journal of Vision*. 2016;16(12):1151–1151.
53. Yeonan-Kim J, Bertalmío M. Retinal lateral inhibition provides the biological basis of long-range spatial induction. *PLoS one*. 2016;11(12):e0168963.
54. Canham TD, Murdoch MJ, Long DL. Influence of Ambient Chromaticity on Portable Display Color Appearance. In: *SMPTE 2018 Annual Technical Conference 'I&' Exhibition*. Society of Motion Picture and Television Engineers; 2018. .
55. Monnier P, Shevell SK. Chromatic induction from S-cone patterns. *Vision Research*. 2004;44:849–856.
56. Zamir SW, Vazquez-Corral J, Bertalmío M. Gamut Extension for Cinema. *IEEE Transactions on Image Processing*. 2017;26(4):1595–1606.
57. Zamir SW, Vazquez-Corral J, Bertalmío M. Vision models for wide colour gamut imaging in cinema. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2020;.
58. Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L. ImageNet: A Large-Scale Hierarchical Image Database. In: *CVPR09*; 2009. .
59. Gonzalez RC, Woods RE. *Digital Image Processing (3rd Edition)*. USA: Prentice-Hall, Inc.; 2006.
60. Murgia M. Perceptual belongingness determines the direction of lightness induction depending on grouping stability and intentionality. *Vision Research*. 2016 May;126:69–79.
61. Singer B, D’Zmura M. Color contrast induction. *Vision Research*. 1994;34(23):3111–3126.
62. Martínez-García M, Cyriac P, Batard T, Bertalmío M, Malo J. Derivatives and inverse of cascaded linear+nonlinear neural models. *PLoS one*. 2018;13(10):e0201326.
63. Wark B, Fairhall A, Rieke F. Timescales of Inference in Visual Adaptation. *Neuron*. 2009;61(5):750–761.
64. Coen-Cagli R, Dayan P, Schwartz O. Cortical surround interactions and perceptual salience via natural scene statistics. *PLoS computational biology*. 2012;8(3):e1002405.
65. Jansen M, Jin J, Li X, Lashgari R, Kremkow J, Bereshpolova Y, et al. Cortical Balance Between ON and OFF Visual Responses Is Modulated by the Spatial Properties of the Visual Stimulus. *Cerebral Cortex*. 2018;29(1):336–355.
66. Wandell BA. *Foundations of vision*. vol. 8. Sinauer Associates Sunderland, MA; 1995.
67. Graham NV. Beyond multiple pattern analyzers modeled as linear filters (as classical V1 simple cells): Useful additions of the last 25 years. *Vision research*. 2011;51(13):1397–1430.
68. Bertalmío M, Caselles V, Provenzi E. Issues about retinex theory and contrast enhancement. *International Journal of Computer Vision*. 2009;83(1):101–119.
69. Zamir SW, Vazquez-Corral J, Bertalmío M. Gamut Mapping in Cinematography Through Perceptually-Based Contrast Modification. *IEEE Journal of Selected Topics in Signal Processing*. 2014;8(3):490–503.
70. Brainard DH. *The Psychophysics Toolbox*. *Spatial Vision*. 1997;10:433–436.
71. Pelli DG. The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision*. 1997;10:437–442.
72. Bertalmío M, Batard T, Kim J. Correcting for Induction Phenomena on Displays of Different Size. In: *Vision Sciences Society Annual Meeting*; 2016. .
73. Rich DC. Publication CIE 159: A colour appearance model for colour management systems: CIECAM02. *Color Research & Application*. 2006;31(2):158–158.