Matching visual induction effects on screens of different size

Trevor Canham

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

Javier Vazquez-Corral

Computer Vision Center and the Computer Sciences Department at Universitat Autònoma de Barcelona, Cerdanyola del Vallès, Spain 🏠✉

Elise Mathieu

Grenoble INP – Phelma, France ✉

Marcelo Bertalmío

Instituto de óptica, Spanish National Research Council (CSIC), Spain ✉

This work is dedicated to the memory of Jihyun Yeonan-Kim.

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 phenomenon presents a practical challenge for the preservation of the artistic intentions of filmmakers, because 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 finding, we propose a method to preprocess 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.

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 toward 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 (Helson, 1963) aimed to quantify the perceptual phenomena first formally described by von Bezold (1874) and Gelb (1930), 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 increased (Figure 2). A similar result for the chromatic case was reported by Fach and Sharpe (1986), 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, whereas for lower spatial frequencies it takes the form of contrast. Although not all confirm these early observations, there exists a large body of later work (e.g., Brenner, Ruiz, Herraiz, Cornelissen, & Smeets, 2003; Brown & MacLeod, 1997; Harrar & Vienot, 2005; Monnier & Shevell, 2003; Shevell & Wei, 1998; Shevell & Monnier, 2005; Wesner & Shevell, 1992) corroborating the importance of the spatial distribution and variability of inducing surrounds. Regarding visual induction models, we single out the work of Otazu, Parraga, and Vanrell (2010), which is based on wavelet decompositions, and the very recent work of
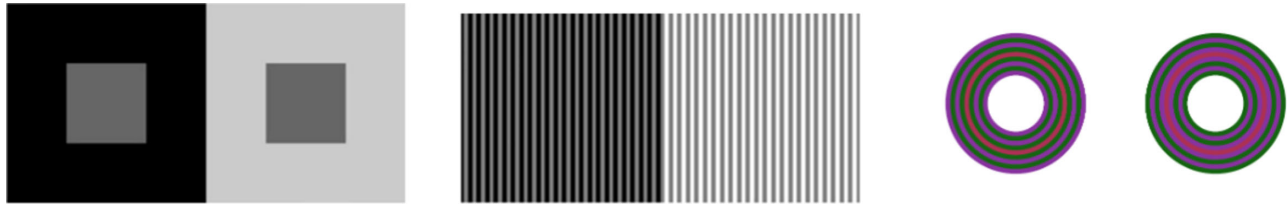
Figure 1. Induction examples. Left: lightness contrast; the center gray squares have the same luminance value but the one surrounded by white is perceived darker and the one surrounded by black is perceived lighter. Middle: lightness assimilation; all gray bars have the same luminance value but the gray bars surrounded by black are perceived as being darker than the ones surrounded by white, which are seen as being lighter. Right: chromatic induction; the central and inducing rings on both sides have the same RGB tristimulus values, but all rings are perceived differently due to their rearrangement.
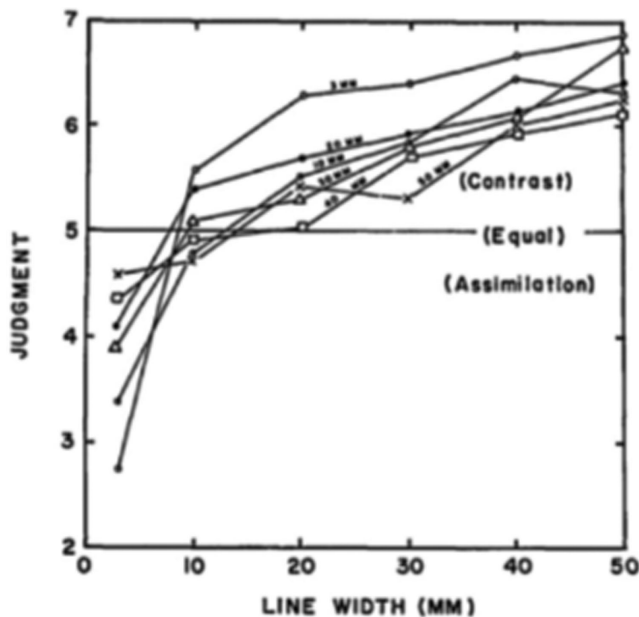

Figure 2. Induction type depends on spatial frequency. Low spatial frequencies induce contrast, while high spatial frequences induce assimilation. Reprinted with permission from Helson (1963). © The Optical Society.

Song, Faugeras, and Veltz (2019), that uses 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; 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 toward 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 that had an original theatrical release may be remastered for separate releases to home video, broadcast television, streaming, and so on. 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 viewer.

With these notions 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 becomes invertible. Finally, based on this invertible formulation we propose a method to preprocess 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 and a validation experiment with natural images; all these data are made available as Supplementary Material.

Neural field model for induction effects

The efficient representation principle, introduced by Attneave (1954) and Barlow (1961), 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 (Smirnakis, Berry, Warland, Bialek, & Meister, 1997). In fact, efficient representation requires that the statistics of the image input are matched by the coding strategy, and although a global part of this coding strategy must have evolved on long timescales (development, evolution), to be truly efficient the coding must also adapt to the local spatiotemporal 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 (Kohn, 2007).

Atick and Redlich (1992) make 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 photo-receptor level (Olshausen & Field, 2000), because 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 lateral geniculate nucleus 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. Bertalmío, Caselles, Provenzi, and Rizzi (2007) propose a variational method to improve the color appearance of images, that performs LHE. They introduce the following energy functional, whose minimization yields the method's result:

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

where I is an image channel 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 channel and α, β and γ are positive weights.

The first term in the functional of Equation 1 measures the dispersion around the mid-range response of $\frac{1}{2}$, as in the *gray world* hypothesis for color constancy which states that in a sufficiently varied scene the average color will be perceived as gray (an observation made by Judd (1940, 1979) and formalized by Buchsbaum (1980) and therefore the illuminant color can be estimated from the color average of the scene; this implies that the minimization of $E(I)$ will make the image mean tend to $\frac{1}{2}$, so that the first term is small, and corresponding to the case where the illuminant is white.

The second term in the functional measures the contrast as the sum of the absolute value of the pixel differences (weighted, through w, by the distance between said pixels); because of the negative sign in front of this term, minimizing $E(I)$ will increase the contrast.

Finally, because the third term measures the difference with the original image I_0, the minimization of $E(I)$ will yield a result that can't be too far away from I_0.

The gradient descent equation for this functional is

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

Starting from $I = I_0$, Equation 2 is iterated until a steady state is reached (corresponding with a minimum of E), that 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 (1997), whose minimization produces a (global) histogram equalization of the original image. Therefore, we can argue that the evolution Equation 2 performs *local* histogram equalization.

This method was applied channel-wise on color images in RGB (Bertalmío et al., 2007) and in a color opponent color space like CIELAB (Zamir, Vazquez-Corral & Bertalmío, 2017), and the results showed that the LHE method has several good properties:

(1) It has a very good local contrast enhancement performance, producing results without visual artifacts of any kind (only when the width of the locality kernel w is very small do haloes start to appear).

(2) It "flattens" the histogram, approaching histogram equalization, as expected due to the relationship of Equation 1 with the one in the histogram equalization model of Sapiro and Caselles (1997).

(3) It reproduces visual perception phenomena such as simultaneous contrast and the Mach Band effect; this is consistent with the functional of Equation 1 modeling perceived contrast in a localized manner, with close neighbors exerting a higher influence than far-away points.

(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 Bertalmío (2014b)).

Additionally, the LHE model of Bertalmío et al. (2007) is closely related to the neural field model of Wilson and Cowan, as pointed out in Bertalmío et al. (2007) and further discussed in Bertalmío and Cowan (2009). In particular, the evolution Equation 2 is very similar to the Wilson-Cowan equations (see Bressloff, Cowan, Golubitsky, Thomas, & Wiener 2002; Wilson & Cowan, 1972, 1973), which have a long and thriving history of modelling cortical low-level dynamics Cowan, Neuman, and Drongelen (2016). It has been proven recently (Bertalmío et al., 2020) 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 Bertalmío et al. (2007); 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 (Kingdom, 2011) or of neural activity (Murray, Vanrell, Otazu, & Parraga 2013), and also to validate existing ones. It is commonly accepted that visual illusions arise owing to neurobiological constraints (Purves, Wojtach, & Howe, 2008) 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 Bertalmío et al. (2007) and also when it considers orientation (Bertalmio, Calatroni, Franceschi, Franceschiello, & Prandi, 2020) 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.

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 Bertalmío et al. (2007) will always produce contrast, not assimilation, because, as we mentioned elsewhere in this article, assimilation is linked to high spatial frequencies (Shevell, 2003). To overcome the intrinsic limitations of Bertalmío et al. (2007) with respect to induction, we should introduce spatial frequency in the energy functional. In Bertalmío (2014a) 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 Bertalmío (2014a) 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 Bertalmío (2019).

Unfortunately, the modifications introduced to the LHE model in Bertalmío (2014a) 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 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 elsewhere in this discussion. 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 (Shapley & Enroth-Cugell, 1984). Furthermore, from Equation 3 and if $n = 1$, we can obtain Weber's law.
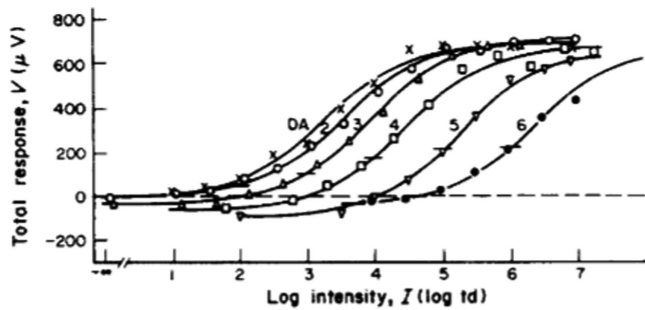
Figure 3. Photoreceptor response curves for primate cones, for different background levels. Reprinted from Valeton and van Norren (1983), Copyright 1983, with permission from Elsevier.

For this and other factors, it appears that the perceptual effects of light adaptation can be mostly accounted for by retinal processing (Meister & Berry, 1999).

Neural response nonlinearities and signal equalization

Neural adaptation performs a (constrained) signal equalization by matching the system response to the stimulus mean and variance (Dunn & Rieke, 2006), 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 are 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 Kane and Bertalmío (2016) 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 (Ozuysal & Baccus, 2012), and that the nonlinear responses of retinal ON and OFF cells are different (Kremkow et al., 2014; Turner & Rieke, 2016; Turner, Schwartz, & Rieke, 2018, see Figure 5). These data on neural activity is consistent with psychophysical data (Kane & Bertalmío, 2019;

Whittle, 1992) 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 of retinal ganglion cells (RGCs), with the excitatory center owing to the feed-forward cells (photoreceptors and bipolar cells) and the inhibitory surround owing 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 (Rucci & Victor, 2015) 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 receptive field size. But in Yeonan-Kim and Bertalmío (2016a) Yeonan-Kim and Bertalmío showed that, in fact, assimilation can *start* already in the retina. They took classic retinal models, those of Wilson (1997) and van Hateren (2005), 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 retinal fields that are much more extended than previously assumed, and RGC responses show a component that goes beyond the classical receptive field.

Based on this discussion, 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
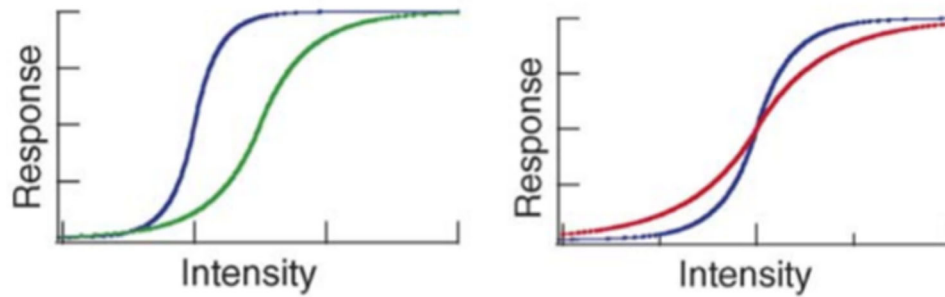
Figure 4. Neural adaptation to mean and variance. Left: neural response to higher (in green) and lower (in blue) mean luminance. Right: neural response to higher (in red) and lower (in blue) luminance variance. Adapted from Dunn and Rieke (2006), Copyright 2006, with permission from Elsevier.
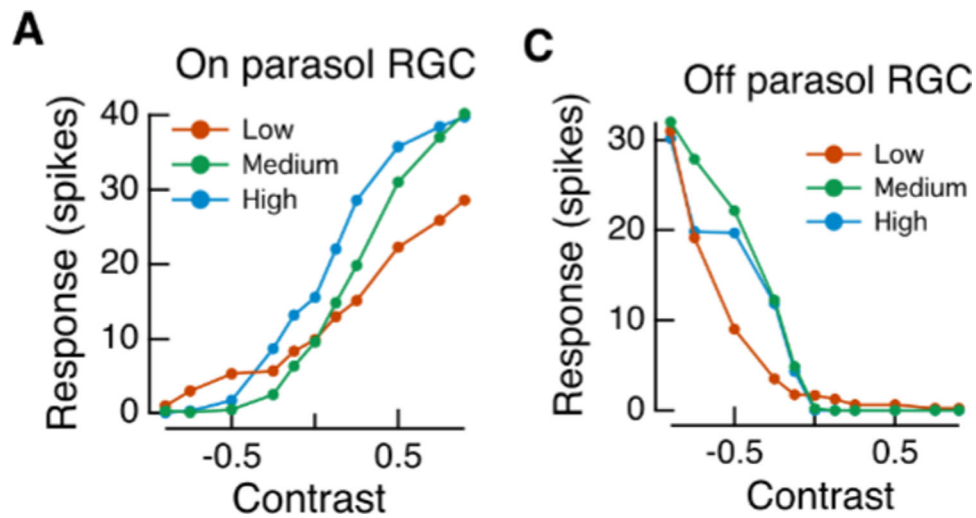

Figure 5. ON and OFF cells have different nonlinear response functions. Reprinted from Turner and Rieke (2016), Copyright 2016, with permission from Elsevier.

J_0, and the semi-saturation 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 Gaussian functions 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 anti-symmetric, hence allowing positive and negative responses to be of different magnitude. Let us note that Equation 5 is the gradient descent equation for an energy functional where the contrast term has this form:

$$\int_{\Omega^2} K_c(x, y)\phi(J(x) - J(y))dxdy \quad (6)$$

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.

Methods: LHEI model validation

To validate LHEI, we will use the chromatic induction data of Monnier (2008). In that work, observers were shown a test ring of some given chromaticity, surrounded by 16 concentric rings (one-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 seem to be lime and purple, selected because they differently stimulate the S cones only. Next to
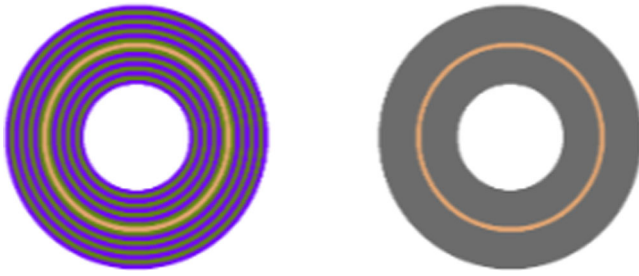
Figure 6. Chromatic experiment stimuli. Left: test ring surrounded by concentric inducing rings of two alternating chromaticities. Right: comparison ring over uniform background. Note that the comparison and test rings are presented at the same chromaticity, and in the actual experiment, these patterns are placed over a black surround.
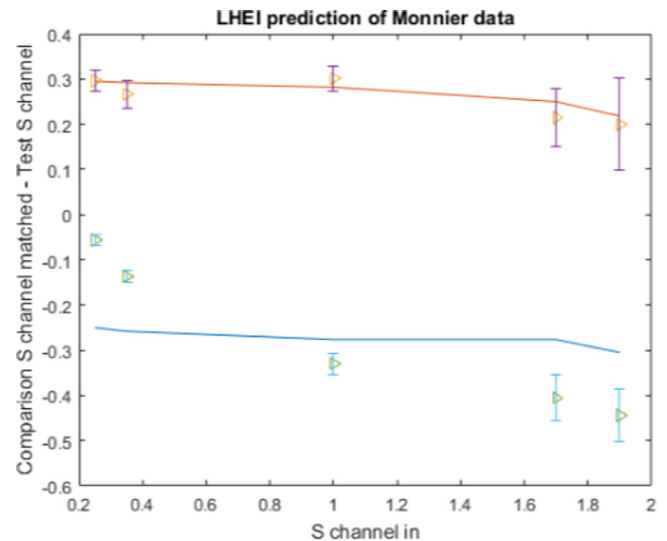

Figure 7. Results from applying the LHEI model to observer data from Monnier (2008). Triangles represent mean observer responses and lines represent LHEI model predictions, both in terms of S-chromaticity difference between test and comparison rings. The 95% confidence intervals are included for each of the observer data points. Values above zero: results when inducing rings next to the test ring have a lime hue. Values below zero: results when inducing rings next to the test ring have a purple hue.

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 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, owing 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).

We have optimized the parameters of the LHEI model 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. For each of the initial conditions, we run our method using both the original rings, and the comparison ring adjusted by observers as input. Then, our minimization looks at the difference between the test ring in these two 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.

Results

The resulting psychophysical data, averaged over the observers, is shown in Figure 7 as orange triangles

with purple error bars for the purple/lime patterns and green triangles with blue error bars for the lime/purple patterns. The error bars represent 95% confidence error intervals about the mean, averaged across observers and trial repetitions. The fits of the model are shown in solid lines, 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. (2019) using a neural field model based on the Wilson-Cowan formulation. For the purple/lime pattern case, our model makes predictions that are within the range of experimental error for all test ring S channel values; however, it does not properly fit the steeper slope of the lime/purple case.

Invertible model for induction effects

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 is 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 cannot 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 (Bertalmío et al., 2020), 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 a different size than the screen 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, and so on, 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 generally have different parameter values. The reason for this is that, as we mentioned in the section on neural processes adapt to the scene statistics, and by 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 preprocessed image $C(I)$ is shown on screen B its appearance, including induction effects, will be $M_B(C(I)) = M_B \cdot M_B^{-1} \cdot M_A(I) = M_A(I)$, that is, 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., Otazu et al., 2010; Song et al., 2019) 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 Kim, Batard, and Bertalmío (2016), the authors went back to the retinal models that were updated and analyzed in Yeonan-Kim and Bertalmío (2016b), 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 Bertalmío et al. (2007), but where the absolute value function in the second term of Equation 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, whereas the functional in Bertalmío et al. (2007) is non-convex and as a consequence its minimum has to be found by the iteration of the gradient descent equation. If we follow this approach to modify the contrast term of the energy functional associated to the LHEI method (Equation 6), we obtain:

$$\int_{\Omega^2} K_c(x, y)(J(x) - J(y))^2 dxdy, \quad (7)$$

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

With this modification, the gradient descent equation previously shown in Equation 5 now becomes:

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

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, $I : \Omega \to [0, +\infty)$, so $J : \Omega \to [0, 1)$.

(2) The second stage produces the output O as the convolution of J with the kernel S of Equation 9:

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

Let's call this model M.

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 Equation 10. Therefore, the inverse of M can be expressed as $M^{-1}(O) = NR^{-1}(S^{-1} * O)$.

Induction compensation method

Based on model M, defined in Equations 11 and 12, 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 the Section on the Invertible model of conduction effects, can perform decorrelation and contrast enhancement. For images with very high contrast, convolution with S_A could produce over enhancement, 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 the function NR_B^{-1} to them 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 12.

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 the section on the Invertible model of conduction effects to represent kernel S, we define what we call the compensation kernel S_C as $S_C = S_B^{-1} * S_A$:

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

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

The formulation was presented for single-channel images. For color images, we will apply the induction compensation C channel-wise. To this end, given an input image in the display-referred RGB space, we will first transform it to a cone-space representation (CAT02 LMS space) (Rich, 2006) by applying the electro-optical transfer function of the input space, converting to CIEXYZ 2-degree tristimulus values given the chromaticity coordinates of the primaries and white point, and finally applying the 3×3 linear transformation matrix from XYZ to CAT02. In this space, we will apply the first Naka-Rushton equation NR_A to the individual L, M, and S channels. In this step the Naka-Rushton exponents will be equal for all the channels. After this is done, we will further convert the color representation to an opponent one, with channels that we call Y, op_1 and op_2, computed as:

$$Y = L + M + S \quad (16)$$

$$op_1 = L - M \quad (17)$$

$$op_2 = 2S - (L + M). \quad (18)$$

Then, our method will convolve each of the channels with the compensation kernel S_C. Let us note that there will be two different compensation kernels: one for the chromatic channels and another for the achromatic one. Once this is done, our method will apply the inverted opponent channel transformation, clipping to the range [0,1] and the inverse of the second Naka-Rushton equation NR_B (see Equation 14). Again here, the Naka-Rushton exponent is kept equal for the three channels. Finally, to convert the processed image back to a state that is ready for display, the inverse 3×3 linear transformation (CAT02 to XYZ) from the forward process is applied followed by the primary matrix (XYZ to RGB) and the inverse electro-optical transfer function of the destination space.

To validate our method, we consider a scenario where *A* corresponds with a cinema screen and *B* with a mobile display. We perform psychophysical experiments for both achromatic and chromatic induction patterns 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 with the mobile viewing scenario) so that it matches the appearance of that region on the large-scale image (corresponding with 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 Canham, Murdoch, and Long (2018) is used resulting in a scaling factor of 0.39 between the two viewing scenarios. Using the data from these experiments the parameters of two separate kernels—one for the achromatic channel, another one for the chromatic channels—and the two Naka-Rushton exponents n_A, n_B will be found by minimizing the error between the observer data and the method results.

Methods: Achromatic induction

Following the preceding work Bertalmío, Batard, and Kim (2016), the achromatic experiment was intended to be a direct expansion of the experiments of Helson (1963) 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 comparison bar width. In this case however, observers reported the necessary correction factor directly by adjusting the luminance of the comparison bars in the mobile scaling to match those in the cinema scaling (test). The additional variable of starting comparison 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 1.
Laboratory setup: Experiments were conducted under dark surround viewing conditions on a calibrated
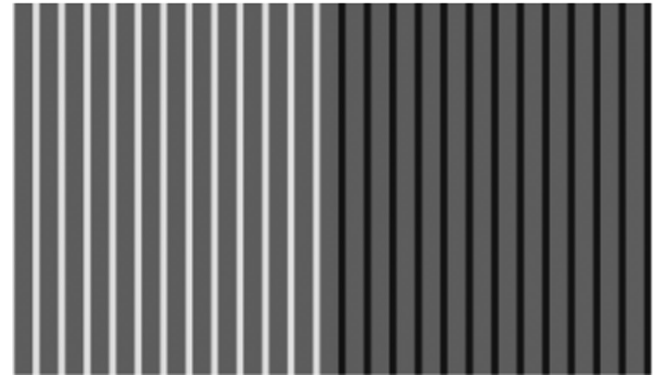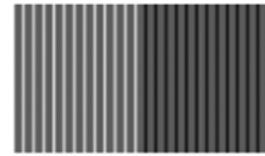

Figure 8. Achromatic experiment stimulus. Note that all gray bars are presented at the same value, and in the actual experiment, these patterns are placed over a black surround.

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 a 2.4 gamma decoding nonlinearity. These settings were 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 (Brainard, 1997; Pelli, 1997) to display stimuli. Observers adjusted the comparison bar luminance via keyboard input to the experimental test bed.

Figure 8 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 background, as opposed to the white surround they are presented over in the figure.) Preliminary experiments showed that the background color was a relevant factor, so we elected to display patterns over a black background to match the dark surround viewing condition. 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/m², while the black fields were presented at a value of 0.6 cd/m².
Procedure: Observers were ushered into the laboratory and the experimental instructions were read aloud. The instructions covered the purpose of the experiment, the

Factor	Type	Levels
Comparison width	Variable	$0.19°, 0.38°, 0.54°, 0.76°, 0.96°$
Initial comparison luminance	Variable	4.0, 8.1, 22 cd/m²
Inducing bar width	Constant	$0.19°$

Table 1. Achromatic experimental factors. Visual angles correspond with the cinema size patterns.

observer task and the control scheme. Before starting the trials, observers were given a test trial with the experimenter in the room to familiarize themselves with the controls. Then, observers completed the experiment task for the 15 patterns with different target bar widths and presentation values as they were presented on the screen in a random order.

Observers: Ten observers (two female and eight male) aged between 23 and 39 years participated in the experiment. All observers had normal or corrected acuity (20/20). Three observers are authors, and the remaining seven were naive to the purpose of the study.

Optimization: The compensation model *C* 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. The error was calculated as the sum squared difference in luminance values between the observer responses and a value sampled from the center of the comparison bars after the method was applied. By optimizing in this way, we show that the method can 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 800×800.

Results

Figure 9 shows the average observer responses in the experiment and the prediction provided by the compensation model C. We can see how the observer results are consistent with those of Helson (1963) in two key points: first, when the visual angle (equivalently the comparison bar line width) decreases, the appearance tends to assimilation, and hence the compensation requires enhancing the contrast; second, as the visual angle increases, the amount of necessary compensation should decrease. Our model responses are consistently inside the range of experimental error for the observer data.

Methods: Chromatic induction

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 compensation model *C* to the channels separately. In this case, we chose CIELAB space because it has some degree of perceptual uniformity. In initial experiments we found that this expansion to
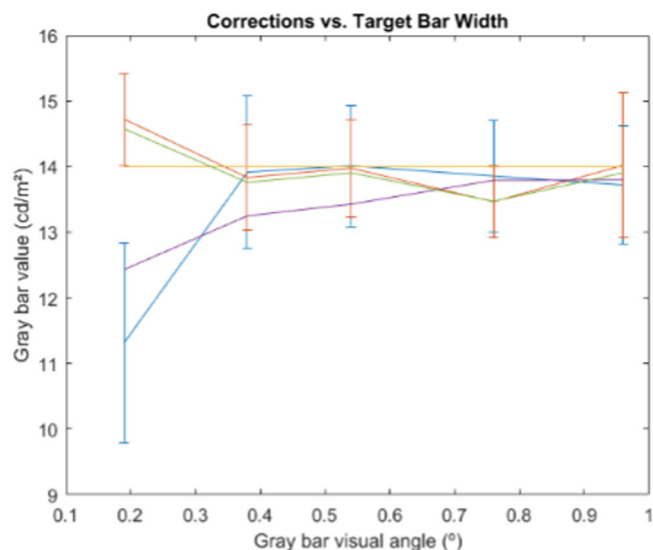

Figure 9. Achromatic experiment results. Yellow: original gray value stimulus. Orange: average observer-selected value for gray bars over black background, showing 95% confidence intervals. Blue: average observer-selected value for gray bars over white, showing 95% confidence intervals. Green: prediction of induction compensation model *C* for bars over black. Purple: model predictions for bars over white.

three-channel adjustment caused a great increase in the difficulty of the experimental task. Thus, we simplified the procedure by decreasing the number of variables. In this case, we test four color sets in the mobile and cinema sizes, and we test for three different comparison 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 2.

For the chromatic experiments we took inspiration from Monnier and Shevell (2004) and used concentric circular induction patterns as shown in Figure 10. Observers must adjust the CIELAB 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*). Otherwise, an equivalent procedure to the achromatic experiment was conducted in this case.

Laboratory setup: The laboratory conditions regarding the display, surround, experimental test bed were all the same as the achromatic experiment. One change, however, was that observers input their responses via a tangent element color correction panel, which allowed for multichannel adjustments to stimuli with

Factor	Type	Levels
Pattern scaling	Variable	cinema and mobile (39%) scaling
Initial comparison level	Variable	original, +10 a*b*, -10 a*b*
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°

Table 2. 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.
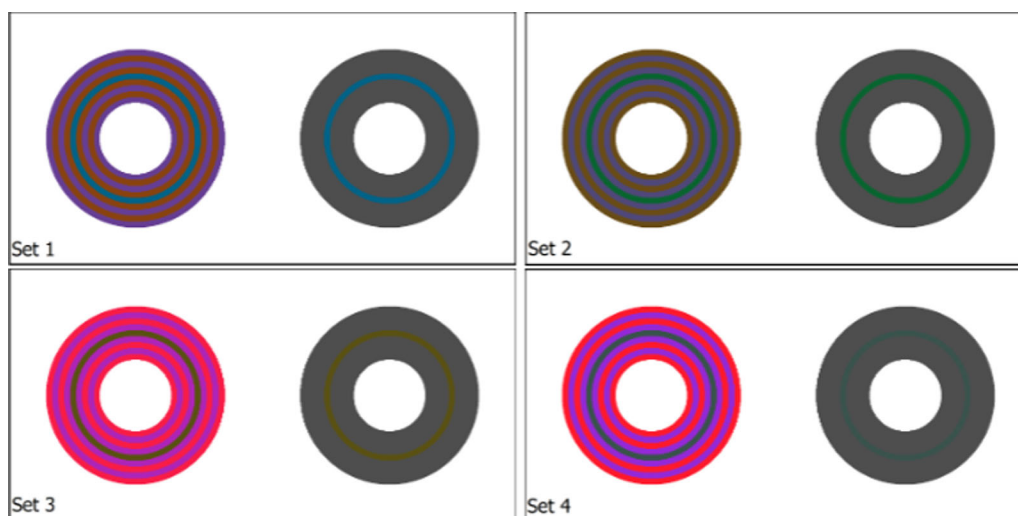

Figure 10. Chromatic experiment stimuli corresponding with sets one through four. On the right side of each set is the comparison ring surrounded by an achromatic field, which observers were asked to match to the test ring on the left, surrounded by the induction pattern. To illustrate the strength of the visual illusion, the comparison and test rings are presented with the same RGB value here. Note that in the actual experiment, these patterns are placed over a black background.

separate knobs allowing for a more natural and reactive experimental interface in comparison with keyboard input.

Stimuli: The stimuli for the chromatic experiments included four concentric circular induction patterns of different color arrangements similar to those used in Monnier and Shevell (2004), as shown in Figure 10. 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.

To find patterns that 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 range of colors that a display can reproduce, and Rec. 709 is the default gamut specification most commonly observed by display and television manufacturers). These patterns were then shown side by side at the different scaling factors tested in the experiment. Then, patterns for which a hue shift could be identified between the different scaling factors were singled out. Finally, 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 reproduced successfully 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, because the target colors were too close
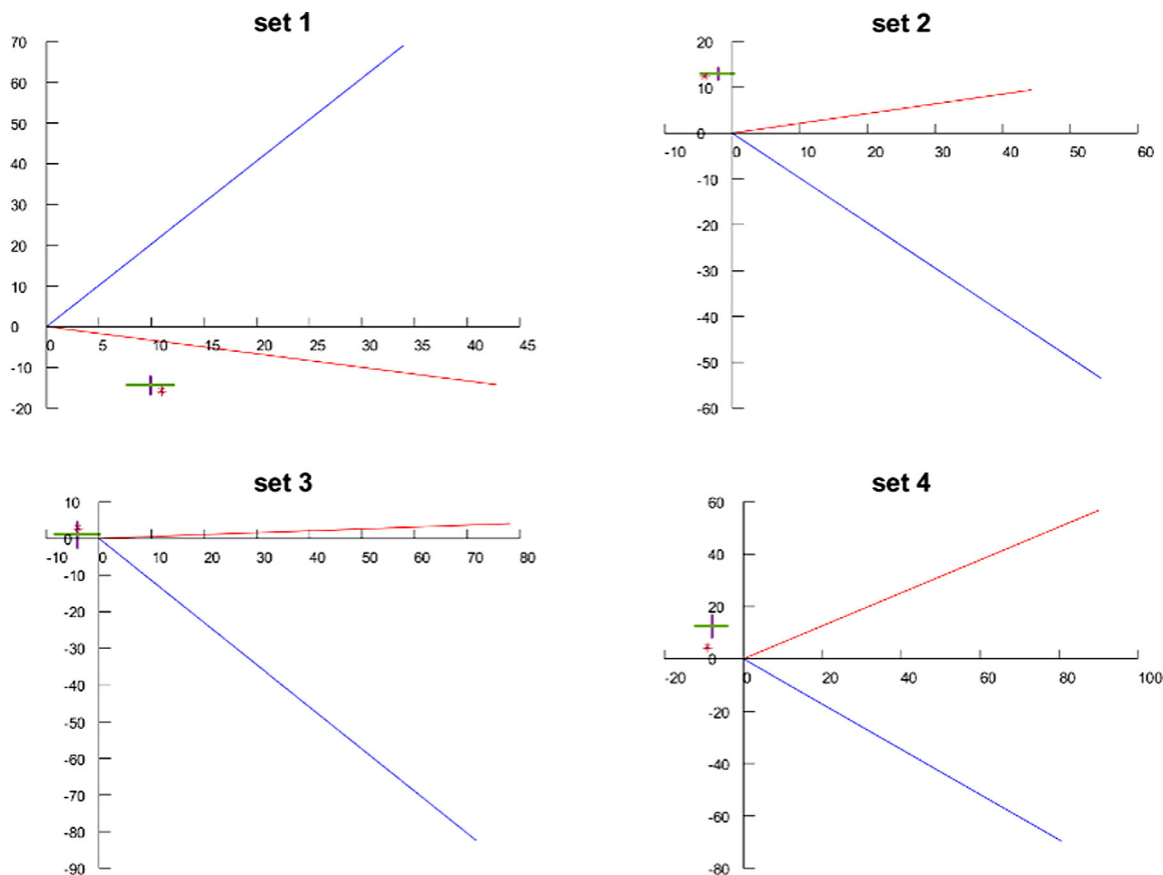
Figure 11. Chromatic experiment results for the four tested color sets, with the a*b* value of the test ring centered at the origin. The vectors represent the magnitude and direction of the color difference between the test ring and the inducing rings. The blue vectors represent the first inducers for each test case, or the color of the ring immediately adjacent to the test ring, and the red vectors represent the second inducer. The purple and green crosses represent the observer suggested compensation with 95% confidence error bars, and the red stars represent the response predicted by our correction method. The bottom right plot (set four) was used to test the model fit while the remaining three were used to train the model parameters.

to the gamut boundary for observers to make reliable observations.

Observers: For the second color set, four observers participated in the experiment (one female, three male) and for the remaining three color sets, three observers participated (three male). 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).

Optimization: We performed these experiments for four different concentric ring patterns, then optimized our model C so that it fits the data for three of these images and finally validated our results on the remaining image. To accomplish this, we apply the kernel S_C only to the two chromatic components of our opponent channel space. The forward model is otherwise applied as described earlier in the section, however during the inverse process we stop after applying the 3×3 transformation from CAT02 to XYZ, and convert this representation of the corrected image directly to

CIELAB, taking the monitor white point of D65 at 100 cd/m^2 for the reference illuminant. The optimization is performed in order to minimize the ΔE error on the test ring, and as we are using three different sets for training the minimization considers the maximum value of the ΔE error on the three test rings. Since our method is working in a color opponent space different from CIELAB, when convolving kernel S_C with our chromatic channels, shifts in the 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.

The values obtained for the case where set 4 is used for testing (corresponding to results in Figure 11 and column 5 in Table 3) 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 800×800.

	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 3. Error between the average observer response and our model's prediction. We have performed the training for all combinations of three sets, testing on the remaining set (columns two to five). Column six represents the original error, and column 7 represents our improvement, w.r.t. the original error in the test set.

Results

In Figure 11, 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. In order to illustrate their directional influence, 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 Helson (1963); Fach and Sharpe (1986); Monnier (2008). 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 compensation model C for fitting this data, in Figure 11 we added our results for the optimized values presented. Our resulting corrections predicted by C for each color set are depicted with a red star. We can see for the train cases that the predictions are within the range of experimental error, and for all cases compensate input in the proper direction.

To further study our model, Table 3 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 three of the sets and testing in the remaining one. This gives us four different cases. In the table, columns two to five represent each of the cases, with the model error for the testing set shown in blue; in particular, column five corresponds with the case illustrated in Figure 11. Column six 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 seven 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 rescaling the original image.

Methods: Induction in natural images

Because the compensation method C was designed with a direct imaging application in mind, it is important to analyze its effect in the context of natural images. In comparison with the synthetic stimuli used to optimize and validate the model, the context of real images introduces a significant increase in the spatial complexity of stimuli. For example, the image content could provide references for cognitive grouping feedback and other higher order processes which could be impactful to the induction effect (Murgia, 2016). Thus, before the method can be proposed for practical use, it is vitally important to first probe its behavior for a variety of test content and viewing contexts to ensure that it is tuned such that it improves the preservation of creative intent with changes in presentation size as a whole.

In addition to the image content, our experiments revealed a number of additional viewing scenario factors which were influential to the induction effect. First, in initial iterations of the achromatic experiments, we found that the background and surround conditions can completely change the nature of the induction effect (changing the direction of required compensation). We also observed that induction effects are not only

dependent on the relative scale adjustment between stimuli, but also on the absolute scaling. Due to this, a different correction would be required for stimuli with absolute scaling of 2 and 1, than for stimuli with absolute scaling of 1 and 0.5. Finally, our observers reported during the experiments that there were visible shifts in the appearance of the synthetic induction patterns with the amount of time spent viewing them.

Procedure: With all of these factors in mind, an online validation experiment was designed and conducted to evaluate the performance of the method on natural images. The experiment was designed and distributed using the PychoJS library and the psychophysics-centered hosting platform Pavlovia (Pierce et al., 2019). In the experiment, observers were first given instructions to extinguish any direct light sources as to make their viewing environment as dark as possible. Then, observers conducted the virtual chin rest test of Li, Joo, Yeatman, and Reinecke (2020) to determine their pixel-per-degree viewing angle so that stimuli could be adjusted to the correct presentation size. They were asked to maintain their seating position from this point in the experiment onward. Then, observers were given instructions that explained that they would be presented with original full-sized reference images and corresponding down-scaled pairs (some of which are altered with respect to their color and contrast and others which are unaltered) at timed intervals (5 seconds on, 2 seconds off), and would be asked to evaluate the match using one of the following options:

(1) The colors of this image have been altered
(2) The colors of this image may have been altered
(3) The colors of this image have not been altered

After this, the observers conducted the body of the experiment, iterating through the test images in random order.

Stimuli: The down-scaled versions include the original, unaltered image and the image corrected with our compensation method C. The method correction was applied following the process detailed at the beginning of the section, applying the kernel optimized in the achromatic experiment to the Y channel and the kernel optimized in the chromatic experiment to the opponent channels op_1 and op_2. A series of images were selected which reflect a cinema or television shooting and grading style. The images cover a range of scene types and include important memory colors such as skin tones, product labels, natural colors, and so on. To avoid observer fatigue, considering the two repetitions of each image and the minimum presentation time of approximately 10 seconds, the number of test images was limited to 33 to allow for observers to be able to

complete the experiment with ample observation time in 20 minutes.

Observers: A total of 16 observers participated in the experiment (10 male, six female) aged between 24 and 58 years. None of the observers are authors and all were naive to the purpose of the experiment. One-half of the observers work in an imaging related field and can thus be considered expert observers, and the other one-half were non-experts.

Results

Figure 12 illustrates the qualitative results of our induction compensation method C on some natural images. We can see how in our results the colors are subtly but noticeably more vivid, for example, the orange cone in the first row, the green teapot in the third, the kid's blue jacket and boots and the grass in the fourth row, the yellow fish in the bottom row. This increased vividness corresponds with a contrast enhancement in the chroma, which should 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 Zamir et al. (2017); Zamir, Vazquez-Corral, and Bertalmío (2021) and Bertalmío (2019).

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

In analyzing the quantitative experimental results, it is important to acknowledge that unlike the model optimization experiments, this experiment did not take place in a controlled laboratory setting and there could be significant variation in final image appearance between observers owing to display performance and calibration, viewing environment limitations, and adherence to the experimental cadence. Although we took steps to limit this variation in the online setting, this allowed for a lesser degree of control in comparison to the previous experiments. The results of Figure 13 show that, within this presentation context, the induction effect is subtle in natural images as in the majority of trials observers did not see a difference in color appearance between scaling settings. In contrast, the results of our method C were seen as having had their color appearance shifted in the majority of trials. Although this experiment was not a direct 2AFC comparison between the control and the results of our proposed method, these were the only two types of images presented to observers and thus the results can be interpreted comparatively, showing that the correction provided by our method was of greater magnitude than the shift caused by the induction effect in this scenario.

Figure 12. Left: original Images. Right: results of our method with parameters optimized for training sets one through three. Notice in our results how colors are slightly more vivid, and the absence of visual artifacts. Images sourced from Deng et al. (2009).
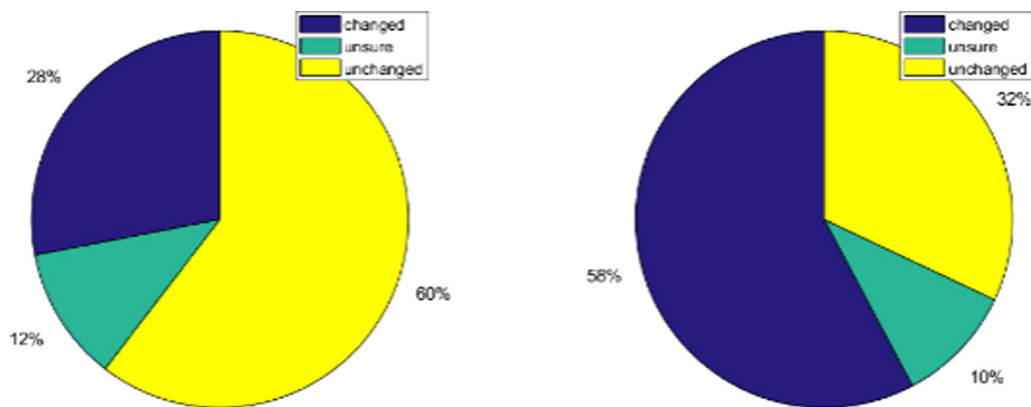
Figure 13. Validation experiment results: percentages of each response for the control (left) and for our proposed method (right).

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 Helson (1963) and Fach and Sharpe (1986), 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 C to the channels separately, with different optimization values compared with the achromatic case. We found that the use of the simple bar patterns of Fach and Sharpe (1986) 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 Monnier and Shevell (2004) and used concentric circular induction patterns, as shown in Figure 10, 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.

A procedure for the selection of color sets that exhibited a strong induction effect is explained in Methods: Chromatic induction. Although 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. Another interesting anomaly that can be observed from this experiment is that despite the random nature in which they were generated and selected, the final four patterns seem to be 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 toward assimilation with an increase in test pattern spatial frequency) can be broken. Although three of the color sets showed this behavior, 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 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 these data, we also found patterns that 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 (Monnier, 2008; Murgia, 2016). Although 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 (Singer & D'Zmura, 1994) on multiple visual pathways.

Although the model *C* 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 in fitting to these four test cases. These results were garnered by model fitting using three chromatic sets and testing on a fourth one. Looking at Table 3, one can see that the model *C* performs slightly better in predicting examples which are within its test set in comparison to when they are excluded, implying that its behavior is somewhat biased toward its training set.

Although the achromatic training set included a number of cases for which the necessary induction correction was negligible, our chromatic kernel is trained exclusively on examples which required a large correction between presentation sizes. Our process of selecting these examples showed that this effect is quite rare, even among synthetic patterns which are specifically designed to produce strong induction effects. The results of the validation experiment showed that the effect is even more subtle in the case of natural images when presented in a "wild" context with variation in display and viewing environment conditions, with a significant majority of observers reporting no color shift in the control examples. In comparison, the correction provided by our method was detectable by observers in the majority of cases. Thus, it is likely that our method in its current optimized state is producing an exaggerated correction for what the practical application requires. For this reason, the compensation method presented here is to be interpreted as a proof of concept and a contribution to the research in the field as opposed to a procedure which is ready to be used in practice.

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 *C*. 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 those 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 *C* makes corrections of similar accuracy when it is required to reach out into larger color volumes.

Conclusion

In this work, we have shown that a neural field model performing local histogram equalization is able to predict chromatic induction effects. This is a variational model, an embodiment of the efficient representation principle, and by regularizing its associated energy functional the model is still able to represent induction and now becomes invertible. This fact allows us to use the new invertible model as the basis for an induction compensation method, which we call C, to preprocess 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, both achromatic and chromatic. Our results show that the established assumption in the literature that induction tends toward assimilation as the spatial frequency increases is sometimes contradicted by the experimental data, and therefore can't be taken as a general principle.

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

(1) Our induction compensation technique *C* is based on a color appearance model that follows the classic formulation of a cascade of linear-nonlinear (L+NL) modules (Martinez-Garcia, Cyriac, Batard, Bertalmio, & Malo, 2018) 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 (Wark, Fairhall, & Rieke, 2009) 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 (Meister & Berry, 1999). So, for example, depending on the input the receptive field of a single neuron can have different sizes or preferred orientations (Coen-Cagli, Dayan, & Schwartz, 2012), or even change polarity

(ON/OFF) (Jansen et al., 2018). For our purposes of induction compensation, we should study how to make the convolution kernel *S* depend on the input, or instead to use a filter bank as is the traditional approach with L+NL models for visual perception (Wandell, 1995; Graham, 2011).

(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 (Bertalmío, Caselles, & Provenzi, 2009; Zamir et al., 2021; Zamir, Vazquez-Corral, & Bertalmío, 2014). 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 *C* 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. Additionally, it would be an interesting expansion of the work to include more test sets in the chromatic case, which do not produce strong illusions, as was done in the achromatic case, as this would likely better represent the behavior of the visual phenomenon in response to natural images.

Keywords: color perception, visual induction, efficient representation principle, neural field models, local histogram equalization, variational models, Wilson-Cowan equations

Acknowledgments

Commercial relationships: none.
Corresponding author: Trevor Canham.
Email: trevor.canham@upf.edu.
Address: Department of Information and Communication Technologies, Universitat Pompeu Fabra, Barcelona, Spain.

References

Atick, J. J., & Redlich, A. N. (1992). What does the retina know about natural scenes? *Neural Computation, 4*(2), 196–210.

Attneave, F. (1954). Some informational aspects of visual perception. *Psychological Review, 61*(3), 183.

Barlow, H. B. (1961). Possible principles underlying the transformation of sensory messages. *Sensory Communication, 1*, 217–234.

Bertalmío, M. (2014a). From image processing to computational neuroscience: a neural model based on histogram equalization. *Frontiers in Computational Neuroscience, 8*, 71.

Bertalmío, M. (2014b). *Image processing for cinema*. Boca Raton, FL: CRC Press, Taylor & Francis.

Bertalmío, M. (2019). *Vision models for high dynamic range and wide colour gamut imaging: Techniques and applications*. New York: Academic Press.

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. *Journal of Neurophysiology*.

Bertalmio, M., Calatroni, L., Franceschi, V., Franceschiello, B., & Prandi, D. (2020). Cortical-inspired wilson–cowan-type equations for orientation-dependent contrast perception modelling. *Journal of Mathematical Imaging and Vision*, 1–19.

Bertalmío, M., Caselles, V., & Provenzi, E. (2009). Issues about retinex theory and contrast enhancement. *International Journal of Computer Vision, 83*(1), 101–119.

Bertalmío, M., Caselles, V., Provenzi, E., & Rizzi, A. (2007). Perceptual color correction through variational techniques. *IEEE Transactions on Image Processing, 16*, 1058–1072.

Bertalmío, M., & Cowan, J. (2009). Implementing the Retinex algorithm with Wilson-Cowan equations. *Journal of Physiology, Paris*.

Bertalmío, M., Batard, T., & Kim, J. (2016). Correcting for induction phenomena on displays of different size. In *Vision Sciences Society Annual Meeting*.

von Bezold, W.. (1874). *Die farbenlehre*. Braunschweig: Westerman.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10*, 433–436.

Brenner, E., Ruiz, J., Herraiz, E. M., Cornelissen, F., & Smeets, J. (2003). Chromatic induction and the layout of colours within a complex scene. *Vision Research, 43*, 1413–1421.

Bressloff, P., Cowan, J., Golubitsky, M., Thomas, P., & Wiener, M. (2002). What geometric visual hallucinations tell us about the visual cortex. *Neural Computation, 14*(3), 473–491.

Brown, R., & MacLeod, D. (1997). Color appearance depends on the variance of surround colors. *Current Biology, 7*, 844–849.

Buchsbaum, G. (1980). A spatial processor model for object colour perception. *Journal of the Franklin Institute, 310*, 337–350.

Canham, T., Murdoch, M., & Long, D. (2018). Influence of ambient chromaticity on portable display color appearance. In *Smpte 2018 Annual Technical Conference and Exhibition*.

Coen-Cagli, R., Dayan, P., & Schwartz, O. (2012). Cortical surround interactions and perceptual salience via natural scene statistics. *PLoS Computational Biology, 8*(3), e1002405.

Cowan, J. D., Neuman, J., & Drongelen, W. van. (2016). Wilson–cowan equations for neocortical dynamics. *Journal of Mathematical Neuroscience, 6*(1), 1.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. In *Cvpr09*.

Dunn, F. A., & Rieke, F. (2006). The impact of photoreceptor noise on retinal gain controls. *Current Opinion in Neurobiology, 16*(4), 363–370.

Fach, C., & Sharpe, L. T. (1986). Assimilative hue shifts in color gratings depend on bar width. *Perception & Psychophysics, 40*(6), 412–418.

Gelb, A. (1930). Handbuch Norm. *Pathol. Physiol.*, *12*, 594.

Graham, N. V. (2011). Beyond multiple pattern analyzers modeled as linear filters (as classical v1 simple cells): Useful additions of the last 25 years. *Vision Research, 51*(13), 1397–1430.

Harrar, M., & Vienot, F. (2005). Regulation of chromatic induction by neighboring images. *Journal of the Optical Society of America, 22*, 2197–2206.

Hateren, H. van. (2005). A cellular and molecular model of response kinetics and adaptation in primate cones and horizontal cells. *Journal of Vision, 5*(4), 5–5.

Helson, H. (1963). Studies of anomalous contrast and assimilation. *Journal of the Optical Society of America, 53*(1).

Jansen, M., Jin, J., Li, X., Lashgari, R., Kremkow, J., & Bereshpolova, Y. et al. (2018). Cortical balance between on and off visual responses is modulated by the spatial properties of the visual stimulus. *Cerebral Cortex, 29*(1), 336–355.

Judd, D. (1940). Hue, saturation, and lightness of surface colors with chromatic illumination. *Journal of the Optical Society of America, 30*(1), 2–32.

Judd, D. (1979). *Contributions to color science*. Vol. 545, p. 516–522. National Bureau of Standards (USA).

Kane, D., & Bertalmío, M. (2016). System gamma as a function of image-and monitor-dynamic range. *Journal of Vision, 16*(6), 4–4.

Kane, D., & Bertalmío, M. (2019). A reevaluation of whittle (1986, 1992) reveals the link between detection thresholds, discrimination thresholds, and brightness perception. *Journal of Vision, 19*(1), 16–16.

Kim, J., Batard, T., & Bertalmío, M. (2016). Retinal processing optimizes contrast coding. *Journal of Vision, 16*(12), 1151–1151.

Kingdom, F. A. (2011). Lightness, brightness and transparency: A quarter century of new ideas, captivating demonstrations and unrelenting controversy. *Vision Research, 51*(7), 652–673.

Kohn, A. (2007). Visual adaptation: physiology, mechanisms, and functional benefits. *Journal of Neurophysiology, 97*(5), 3155–3164.

Kremkow, J., Jin, J., Komban, S. J., Wang, Y., Lashgari, R., & Li, X. et al. (2014). Neuronal nonlinearity explains greater visual spatial resolution for darks than lights. *Proceedings of the National Academy of Sciences, 201310442*.

Li, Q., Joo, S., Yeatman, J., & Reinecke, K. (2020). Controlling for participants' viewing distance in large-scale, psychophysical online experiments using a virtual chinrest. *Nature Scientific Reports*.

Martinez-Garcia, M., Cyriac, P., Batard, T., Bertalmio, M., & Malo, J. (2018). Derivatives and inverse of cascaded linear+ nonlinear neural models. *PLoS One, 13*(10), e0201326.

Meister, M., & Berry, M. J. (1999). The neural code of the retina. *Neuron, 22*(3), 435–450.

Monnier, P. (2008). Standard definitions of chromatic induction fail to describe induction with s-cone

patterned backgrounds. *Vision Research, 48*, 2708–2714.

Monnier, P., & Shevell, S. (2003). Large shifts in color appearance from patterned chromatic backgrounds. *Nature Neuroscience, 6*, 801–802.

Monnier, P., & Shevell, S. (2004). Chromatic induction from s-cone patterns. *Vision Research, 44*, 849–856.

Murgia, M. (2016, May). Perceptual belongingness determines the direction of lightness induction depending on grouping stability and intentionality. *Vision Research, 126*, 69–79.

Murray, N., Vanrell, M., Otazu, X., & Parraga, C. A. (2013). Low-level spatiochromatic grouping for saliency estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 35*(11), 2810–2816.

Olshausen, B., & Field, D. (2000). Vision and the coding of natural images. *American Scientist, 88*(3), 238–245.

Otazu, X., Parraga, C., & Vanrell, M. (2010). Toward a unified chromatic induction model. *Journal of Vision, 10*(5), 1–24.

Ozuysal, Y., & Baccus, S. A. (2012). Linking the computational structure of variance adaptation to biophysical mechanisms. *Neuron, 73*(5), 1002–1015.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision, 10*, 437–442.

Pierce, J., Gray, J., Simpson, S., MacAskill, M., Höchenberger, R., & Sogo, H., et al. (2019). Psychopy2: Experiments in behavior made easy. *Behavior Research Methods, 51*, 195–203.

Purves, D., Wojtach, W. T., & Howe, C. (2008). Visual illusions: an empirical explanation. *Scholarpedia, 3*(6), 3706.

Rich, D. C. (2006). Publication cie 159: A colour appearance model for colour management systems: Ciecam02. *Color Research & Application, 31*(2), 158–158.

Rucci, M., & Victor, J. D. (2015). The unsteady eye: an information-processing stage, not a bug. *Trends in neurosciences, 38*(4), 195–206.

Sapiro, G., & Caselles, V. (1997). Histogram modification via differential equations. *Journal of Differential Equations, 135*, 238–266.

Shapley, R., & Enroth-Cugell, C. (1984). Visual adaptation and retinal gain controls. *Progress in Retinal Research, 3*, 263–346.

Shevell, P., & Monnier, S.K.and. (2005). Color shifts from s-cone patterned backgrounds: Contrast sensitivity and spatial frequency selectivity. *Vision Research, 45*, 1147–1154.

Shevell, S. (2003). The science of color (pp. 149–190). Elsevier Science Ltd.

Shevell, S., & Wei, J. (1998). Chromatic induction: Border contrast or adaptation to surrounding light? *Vision Research, 38*, 1561–1566.

Singer, B., & D'Zmura, M. (1994). Color contrast induction. *Vision Research, 34*(23), 3111–3126.

Smirnakis, S. M., Berry, M. J., Warland, D. K., Bialek, W., & Meister, M. (1997). Adaptation of retinal processing to image contrast and spatial scale. *Nature, 386*(6620), 69.

Song, A., Faugeras, O., & Veltz, R. (2019). A neural field model for color perception unifying assimilation and contrast. *PLoS Computational Biology, 15*(6), e1007050.

Turner, M. H., & Rieke, F. (2016). Synaptic rectification controls nonlinear spatial integration of natural visual inputs. *Neuron, 90*(6), 1257–1271.

Turner, M. H., Schwartz, G. W., & Rieke, F. (2018). Receptive field center-surround interactions mediate context-dependent spatial contrast encoding in the retina. *bioRxiv*, 252148.

Valeton, J., & van Norren, D.. (1983). Light adaptation of primate cones: an analysis based on extracellular data. *Vision Research, 23*(12), 1539–1547.

Wandell, B. A. (1995). *Foundations of vision* (Vol. *8*). Sinauer Associates, Sunderland, MA.

Wark, B., Fairhall, A., & Rieke, F. (2009). Timescales of inference in visual adaptation. *Neuron, 61*(5), 750–761.

Wesner, M., & Shevell, S. (1992). Color perception within a chromatic context: Changes in red/green equilibria caused by noncontiguous light. *Vision Research, 32*, 1623–1634.

Whittle, P. (1992). Brightness, discriminability and the "crispening effect". *Vision Research, 32*(8), 1493–1507.

Wilson, H., & Cowan, J. (1972). Excitatory and inhibitory interactions in localized populations of model neurons. *Biophysical Journal, 12*, 1–24.

Wilson, H., & Cowan, J. (1973). A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue. *Biological Cybernetics, 13*(2), 55–80.

Wilson, H. R. (1997). A neural model of foveal light adaptation and afterimage formation. *Visual Neuroscience, 14*(3), 403–423.

Yeonan-Kim, J., & Bertalmío, M. (2016a). Retinal lateral inhibition provides the biological basis of long-range spatial induction. *PLoS One, 11*(12), e0168963.

Yeonan-Kim, J., & Bertalmío, M. (2016b). Retinal lateral inhibition provides the biological basis of long-range spatial induction. *PLoS One, 11*(12), e0168963.

Zamir, S. W., Vazquez-Corral, J., & Bertalmío, M. (2021). Vision models for wide colour gamut imaging in cinema. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 43*(5), 1777–1790, doi:10.1109/TPAMI.2019.2938499.

Zamir, S. W., Vazquez-Corral, J., & Bertalmío, M. (2014). Gamut mapping in cinematography through perceptually-based contrast modification. *IEEE Journal of Selected Topics in Signal Processing, 8*(3), 490–503.

Zamir, S. W., Vazquez-Corral, J., & Bertalmío, M. (2017). Gamut extension for cinema. *IEEE Transactions on Image Processing, 26*(4), 1595–1606.