Visual illusions via neural dynamics: Wilson-Cowan-type models and the efficient representation principle

Marcelo Bertalmío
DTIC, Universitat Pompeu Fabra, Barcelona, Spain
marcelo.bertalmio@upf.edu

Luca Calatroni
Université Côte d'Azur, CNRS, INRIA,
Laboratoire d'Informatique, Signaux et Systèmes de Sophia Antipolis, France
calatroni@i3s.unice.fr

Valentina Franceschi
IMO, Université Paris-Sud, Orsay, France
valentina.franceschi@u-psud.fr

Benedetta Franceschiello
FAA, LINE, Radiology, CHUV, Lausanne, Switzerland
benedetta.franceschiello@fa2.ch

Alexander Gomez-Villa
DTIC, Universitat Pompeu Fabra, Barcelona, Spain
alexander.gomez@upf.edu

Dario Prandi
Université Paris-Saclay, CNRS, CentraleSupélec,
Laboratoire des signaux et systèmes, Gif-sur-Yvette, France
dario.prandi@l2s.centralesupelec.fr

Abstract

We reproduce supra-threshold perception phenomena, specifically visual illusions, by Wilson-Cowan-type models of neuronal dynamics. Our findings show that the ability to replicate the illusions considered is related to how well the neural activity equations comply with the efficient representation principle. Our first contribution consists in showing that the Wilson-Cowan (WC) equations can reproduce a number of brightness and orientation-dependent illusions. Then, we formally prove that there can't be an energy functional that the Wilson-Cowan dynamics are minimizing. This leads us to consider an alternative, variational modelling which has been previously employed for local histogram equalization (LHE) tasks. In order to adapt our model to the architecture of V1, we perform an extension that has an explicit dependence on local image orientation. Finally, we report several numerical experiments showing that LHE provides a better reproduction of visual illusions than the original WC formulation and that its cortical extension is capable to reproduce also complex orientation-dependent illusions.

New & Noteworthy: We show that the Wilson-Cowan equations can reproduce a number of brightness and orientation-dependent illusions. Then, we formally prove that there can't be an energy functional that the Wilson-Cowan equations are minimizing, making them sub-optimal with respect to the efficient representation principle. We thus propose a slight modification that is consistent with such principle and show that this provides a better reproduction of visual illusions than the original Wilson-Cowan formulation. We also consider the cortical extension of both models in order to deal with more complex orientation-dependent illusions.

1 Introduction

The goal of this work is to point out the intimate connections existing between three popular approaches in vision science: the Wilson-Cowan equations, the study of visual brightness illusions, and the efficient representation theory.

As other articles in this special issue make abundantly clear, Wilson-Cowan equations have a long and successful story of modelling cortical low-level dynamics [**?**]. Nonetheless, the study of psychophysics by Wilson-Cowan equations ([**?**, **?**, **?**, **?**, **?**, **?**, **?**, **?**]) is a topic that hasn't been addressed much in neuroscience, and we are not aware of publications in which Wilson-Cowan equations are used for predicting brightness illusions. In this work, we aim to fill this gap.

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 [**?**] or of neural activity [**?**, **?**], and also to validate the existing ones. It is commonly accepted that visual illusions arise due to neurobiological constraints [**?**] that modify the underpinned mechanisms of the visual system.

The efficient representation principle, introduced by Attneave [**?**] and Barlow [**?**], states that neural responses aim to overcome these neurobiological constraints and to optimize the limited biological resources by being tailored to the statistics of the images that the individual typically encounters, so that visual information can be encoded in the most efficient way. This principle is a general strategy observed across mammalian, amphibian and insect species [**?**] and is embodied by neural processing according to abundant experimental evidence [**?**, **?**, **?**].

Our work aims at pulling together the three approaches just mentioned, providing a more unified framework to understand vision mechanisms. First, we show that the Wilson-Cowan equations are able to qualitatively reproduce a number of visual illusions. Secondly, we formally prove that Wilson-Cowan equations (with constant input) are not variational, in the sense that they are not minimizing any energy functional. Next, we detail how a simple modification turning the Wilson-Cowan equations variational yields a local histogram equalisation method that is consistent with the efficient representation principle. We finally show how this new formulation provides a better reproduction of visual illusions than the Wilson-Cowan model.

We remark that our model has to be intended as a proof of concept, whose objective is the reproduction of perceptual phenomena at a macroscopic level with no quantitative assessment on analogous psychophysical data. There are in fact very important limitations for doing that, since such comparison would require both a perfect knowledge of how behavioural data were collected, and a tuning of the model parameters to match with the observed perception. Nonetheless, we believe that the numerical evidence of our experiments and our theoretical considerations can be used for future research studies comparing our computational results with the ones corresponding to experiments coming from psychophysics.

2 Materials and methods

2.1 Visual illusions

Computational models able to reproduce visual illusions represent very effective methods to test new hypotheses and generate new insights, both for neuroscience and applied disciplines such as image processing. Illusions can be classified according to the main feature detection mechanisms involved during the visual process [**?**]. In this contribution we considered two main groups of visual illusions to assess the efficacy of our model in reconstructing the perceptual process: *brightness illusions* and *orientation-dependent illusions*.

2.1.1 Brightness illusions

Brightness illusions are a class of phenomena where image regions with the same gray level are perceived as having different brightness, depending on the shapes, arrangement and gray level of

the surrounding elements. Fig. **??** shows the nine brightness illusions we have chosen to perform tests on in this paper. They are all very popular and at the same time they represent a diverse set, as we can see from the following descriptions.

White's illusion: the left gray rectangle appears darker than the right one, while both are identical [**?**] (Fig. **??**(a)).

Simultaneous brightness contrast: the left gray square appears lighter than the right one, while both are identical [**?**] (Fig. **??**(b)).

Checkerboard illusion: the mid-gray square in the fifth column appears darker than the one in the seventh column, while both are identical [**?**] (Fig. **??**(c)).

Chevreul illusion: a pattern of homogeneous bands of increasing intensity from left to right is presented. However, the bands in the image are perceived as inhomogeneous, i.e. darker and brighter lines appear at the borders between adjacent bands [**?**] (Fig. **??**(d)).

Chevreul cancellation: when the order of the bands is reversed, now decreasing in intensity from left to right, the effect is cancelled [**?**] (Fig. **??**(e)).

Dungeon illusion: two gray rectangles are perceived as darker or lighter depending on the gray intensities of both the background and the grid, see [**?**]. The left rectangle is perceived as darker than the one on the right (Fig. **??**(f)).

Grating induction: the background grating (which can be tuned to different orientations) induces the appearance of a counter-phase grating in the homogeneous gray horizontal bar [**?**] (Fig. **??**(g)).

Hong-Shevell illusion: the mid-gray half-ring on the left appears darker than the one on the right, while both are identical [**?**] (Fig. **??**(h)).

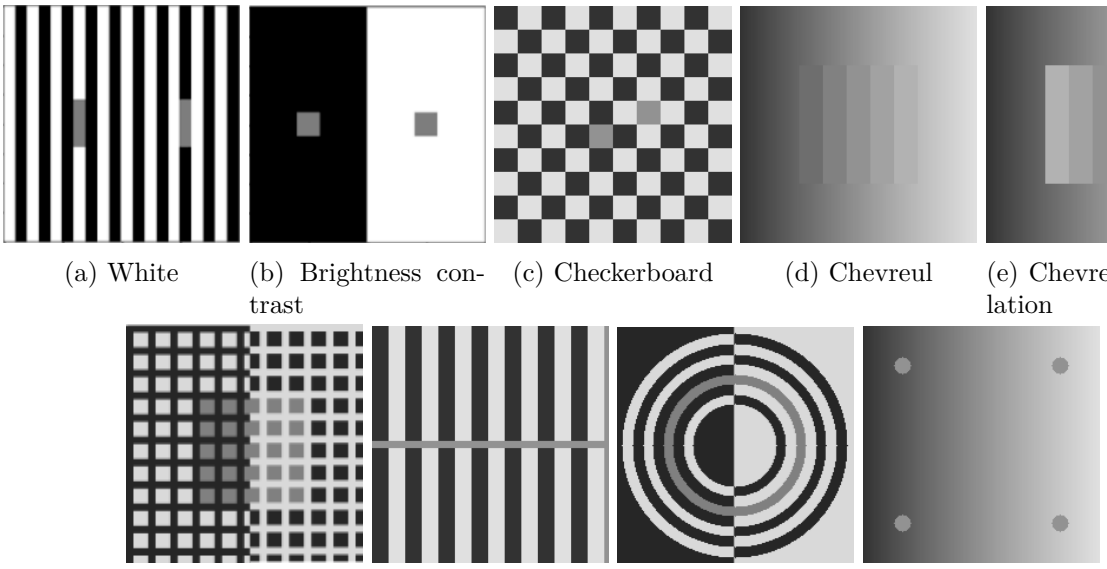
Luminance illusion: four identical dots over a background where intensity increases from left to right, and the dots on the left are perceived being lighter than the ones on the right [**?**] (Fig. **??**(i)).

2.1.2 Orientation-dependent illusions

We also consider orientation-dependent illusions, where the perceptual phenomenon (e.g. in terms of brightness or contrast) is affected by the orientation of the image elements.

Poggendorff illusion. The Poggendorff illusion, presented in the modified version considered in this work in Fig. **??**(a), is a very well known geometrical optical illusion in which the presence of a central surface induces a misalignment of the background lines. This illusion depends both on the orientation of the background lines and the width of the central surface [**?**], as the more the angle is close to $\pi/2$ the less is the bias, but in this example the perceived bias is also dependent on the brightness contrast between central surface and background lines.

Tilt illusion. The Tilt illusion is a phenomenon where the perceived orientation of a test line or grating is altered by the presence of surrounding lines or a grating with a different orientation. In our case we consider the effect that the orientation of a surround grating pattern has on the perceived contrast of a grating pattern in the center: the inner circles in Figs. **??**(b) and **??**(c) are identical but the latter is perceived as having more contrast than the former.



(a) White (b) Brightness contrast (c) Checkerboard (d) Chevreul (e) Chevreul cancellation

(f) Dungeon (g) Grating induction (h) Hong-Shevell (i) Luminance

Figure 1: From left to right, top to bottom: White's illusion, Brightness contrast, the Checkerboard illusion, the Chevreul illusion, Chevreul cancellation, the Dungeon illusion, the Grating induction, the Hong-Shevell illusion and the Luminance illusion.

(a) Poggendorff illusion. (b) Tilt illusion, same θ. (c) Tilt illusion, different θ.

Figure 2: From left to right: a modified version of the Poggendorff illusion based on Grating Induction, a modified Tilt illusion with concentric circles having the same orientation and a modified Tilt illusion with concentric circles having different orientations.

2.2 Wilson-Cowan-type models for contrast perception

In this section we introduce four different evolution equations derived from the Wilson-Cowan formulation, that will be studied in this paper. We recall that, denoting by $a(x,t)$ the state of a population of neurons with spatial coordinates $x \in \mathbb{R}^2$ at time $t > 0$, the Wilson-Cowan equations proposed in [**?**, **?**] can be written[1] as

$$\frac{\partial}{\partial t}a(x,t) = -\beta a(x,t) + \nu \int_{\mathbb{R}^2} \omega(x\|y)\sigma(a(y,t))\,dy + h(x), \tag{2.1}$$

where $\beta > 0$ and $\nu \in \mathbb{R}$ are fixed parameters, $\sigma : \mathbb{R} \to \mathbb{R}$ is a non-linear sigmoid saturation function, the kernel $\omega(x\|y)$ models interactions at two different spatial locations x and y (we will assume that the integral of ω is normalised to 1) and h is the input signal.

2.2.1 Wilson-Cowan equations do not fulfill any variational principle

Over the last thirty years, the use of variational methods in imaging has become increasingly popular as a regularisation strategy for solving general ill-posed imaging problems in the form

$$\text{find } u \quad \text{s.t.} \quad f = \mathcal{T}(u). \tag{2.2}$$

Here, f represents a given degraded image and $\mathcal{T}$ a (possibly non-linear) operator describing the degradation (e.g. noise, blur, under-sampling, etc.)

Due to the lack of fundamental properties such as existence, uniqueness and stability of the solution of the problem (**??**), the idea of regularisation consists of incorporating *a priori* information on the desired image $u_\star$ and on its closeness to the data f by means of suitable variational terms. This gives rise, in particular, to variational methods where one looks for an approximation $u_\star$ of the real solution u by solving

$$u_\star = \arg\min \mathcal{E}(u), \tag{2.3}$$

where $\mathcal{E}$ is the energy functional combining regularisation and data fit, depending also on the given image f. A popular way to solve the variational problem consists in finding $u_\star$ as the steady-state solution of the evolution equation given by the gradient descent of the energy functional

$$\frac{\partial}{\partial t}u = -\nabla\mathcal{E}(u), \qquad u|_{t=0} = f, \tag{2.4}$$

under appropriate conditions on the boundary of the image domain.

In the context of vision science, evolution equations have been originally used as a tool to describe the physical transmission, diffusion and interaction phenomena of stimuli in the visual cortex [**?**, **?**, **?**]. Variational methods are the main tool of ecological approaches, that pose the efficient coding problem [**?**] as an optimisation problem to be solved with evolution equations that minimise an energy functional [**?**] involving natural image statistics and biological constraints. The resulting solution is optimal because it has minimal redundancy.

However, we must remark that, while considering the gradient descent of an energy functional gives always an evolution equation, the reverse is not true: not every evolution equation is minimising an energy functional. In fact, this is the case for the Wilson-Cowan equations, which do not fulfil any variational principle, as we prove in Appendix **??**. As a consequence, they are sub-optimal in reducing the redundancy.

We remark that it is possible to define an energy that decreases along trajectories of (**??**), as done in [**?**]. This ensures in particular that even though the evolution is not variational, its steady states (i.e., solutions of (**??**) that are constant in time) can indeed be obtained as critical points of this energy.

[1] In [**?**] the sigmoid function is applied outside of the integral term and not only on the activity $a(y,t)$ as in (**??**). This corresponds to an "activity-based" model of neuron activation, while (**??**) corresponds to a "voltage-based" one. See [**?**], where the two models are shown to be equivalent.

2.2.2 A modification of the Wilson-Cowan equations complying with efficient representation

Remarkably, the efficient representation principle has correctly predicted a number of neural processing aspects and phenomena like the photoreceptor response performing histogram equalisation, the dominant features of the receptive fields of retinal ganglion cells (lateral inhibition, the switch from bandpass to lowpass filtering when the illumination decreases, and, remarkably, colour opponency, with photoreceptor signals being highly correlated but color opponent signals having quite low correlation), or the receptive fields of cortical cells having a Gabor function form [**?**, **?**, **?**]. Efficient representation is the only framework able to predict the functional properties of neurons from a simple principle, and given how simple the assumptions are it's really surprising that this approach works so well [**?**].

In [**?**] it is shown how a slight modification of the Wilson-Cowan formulation leads to a variational model, as we now present. Assuming that the activity signal a is in the range $[0, 1]$, we can re-write equation (**??**) in terms of a sigmoid $\hat{\sigma}$ shifted by $\frac{1}{2}$ (which we take as the average signal value) and inverted in sign, thus getting:

$$\frac{\partial}{\partial t}a(x,t) = -\beta a(x,t) - \nu \int_{\mathbb{R}^2} \omega(x\|y)\hat{\sigma}\left(a(y,t) - \frac{1}{2}\right) dy + h(x). \tag{2.5}$$

Note that this is just a re-writing of equation (**??**), so it is still not associated to any variational method. However, if we now assume $\hat{\sigma}$ to be odd and replace the $\frac{1}{2}$ term by $a(x,t)$, we obtain

$$\frac{\partial}{\partial t}a(x,t) = -\beta a(x,t) + \nu \int_{\mathbb{R}^2} \omega(x\|y)\hat{\sigma}(a(x,t) - a(y,t))\, dy + h(x), \tag{2.6}$$

and this equation is now a gradient descent equation, as it does fulfil a variational principle.

Furthermore, under the proper choice of parameters β, ν and input signal h, this evolution equation performs local histogram equalisation (LHE) [**?**]. This is key for our purposes, since, as Atick points out [**?**], one of the main types of redundancy or inefficiency in an information system like the visual system happens when some neural response levels are used more frequently than others, and for this type of redundancy the optimal code is the one that performs histogram equalisation.

It is therefore expected that the modification of the Wilson-Cowan equations in (**??**), which better complies with the efficient representation principle, should be more effective in reducing redundancy than the original Wilson-Cowan model of equation (**??**).

2.2.3 Accounting for orientation

Models (**??**) and (**??**) ignore orientation and as such they are not well-suited to explain a number of visual phenomena. For this reason, following [**?**], we extend them to a third dimension, representing local image orientation, as follows. We let $La : Q \times [0, \pi) \to \mathbb{R}$ be the cortical activation in V1 associated with the signal a, so that $La(x, \theta)$ encodes the response of the neuron with spatial preference x and orientation preference θ to a. Mathematically, such activation is obtained via a suitable convolution with the receptive profiles of V1 neurons, as explained in Appendix **??**, see also [**?**, **?**, **?**, **?**, **?**]. Then, denoting by $A(x, \theta, t)$ the cortical response at time t for any $t > 0$, the natural extension of equations (**??**) and (**??**) to the orientation dependent case is given by the two models:

$$\frac{\partial}{\partial t}A(x,\theta,t) = -\beta A(x,\theta,t) + \nu \int_0^\pi \int_Q \omega(x,\theta\|y,\phi)\sigma\Big(A(y,\phi,t)\Big)\, dy\, d\phi + Lh(x,\theta), \tag{2.7}$$

$$\frac{\partial}{\partial t}A(x,\theta,t) = -\beta A(x,\theta,t) + \nu \int_0^\pi \int_Q \omega(x,\theta\|y,\phi)\hat{\sigma}\big(A(x,\theta,t) - A(y,\phi,t)\big)\, dy\, d\phi + Lh(x,\theta), \tag{2.8}$$

where $Lh(x, \theta)$ denotes the cortical activation in V1 corresponding to the visual input h at spatial location x and orientation preference θ. We remark that these models describe the dynamic

behaviour of activations in the 3D space of positions and orientation. As explained in Appendix **??**, once a stationary solution is found, the two-dimensional perceived image can be found by simply applying the formula

$$a(x) = \frac{1}{\pi}\int_0^\pi A(x,\theta)\,d\theta. \tag{2.9}$$

2.2.4 Models under consideration

We summarise here the four models we are going to test in the following sections. The orientation-independent WC and LHE models are:

$$\frac{\partial}{\partial t}a(x,t) = -(1+\lambda)a(x,t) + \frac{1}{2M}\int_Q \omega(x,y)\sigma\left(a(y,t)\right)\,dy + \lambda f_0(x) + \mu(x) \tag{WC-2D}$$

$$\frac{\partial}{\partial t}a(x,t) = -(1+\lambda)a(x,t) + \frac{1}{2M}\int_Q \omega(x,y)\hat{\sigma}\left(a(x,t)-a(y,t)\right)\,dy + \lambda f_0(x) + \mu(x), \tag{LHE-2D}$$

which relate to (**??**) and (**??**) by simply choosing parameters as $\beta = 1+\lambda$ and $\nu = 1/2M$ where $M > 0$ is a normalisation constant, and input signal $h(x) = \lambda f_0(x) + \mu(x)$, where $\lambda > 0$, $f_0(x)$ is the local intensity at $x \in Q$ of given image f_0 and $\mu(x)$ denotes a local average of the initial stimulus f_0 around x (a choice motivated by the averaging behaviour of cells in the magnocellular pathway [**?**] and already considered in similar models e.g. [**?**, **?**]).

The orientation-dependent WC and LHE models can be similarly written as:

$$\begin{aligned}\frac{\partial}{\partial t}A(x,\theta,t) =& -(1+\lambda)A(x,\theta,t) + \frac{1}{2M}\int_0^\pi\int_Q \omega(x,\theta||y,\phi)\sigma\left(A(y,\phi,t)\right)\,dy\,d\phi \\ &+ \lambda L f_0(x,\theta) + L\mu(x,\theta),\end{aligned} \tag{WC-3D}$$

$$\begin{aligned}\frac{\partial}{\partial t}A(x,\theta,t) =& -(1+\lambda)A(x,\theta,t) + \frac{1}{2M}\int_0^\pi\int_Q \omega(x,\theta||y,\phi)\hat{\sigma}\left(A(x,\theta,t)-A(y,\phi,t)\right)\,dy\,d\phi \\ &+ \lambda L f_0(x,\theta) + L\mu(x,\theta),\end{aligned} \tag{LHE-3D}$$

which can analogously be related to (**??**) and (**??**) by choosing the very same parameters as above and by now taking as cortical activation in V1 corresponding to h the quantity $Lh(x,\theta) = \lambda L f_0(x,\theta) + L\mu(x,\theta)$.

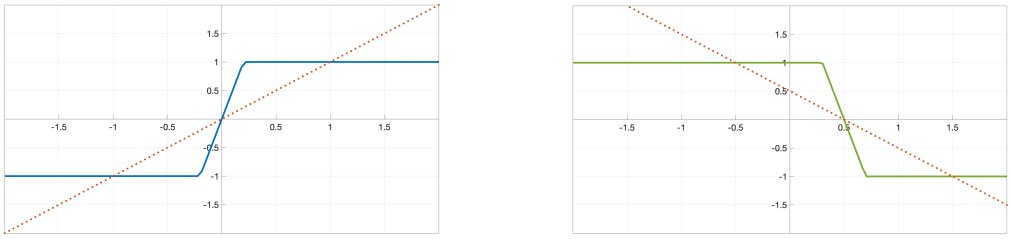
2.2.5 Numerical implementation

All four relevant equations (**??**), (**??**), (**??**), and (**??**) are numerically implemented via a forward Euler time-discretisation, as presented in [**?**]. For a given image a, the cortical activation La is recovered via standard wavelet transform methods, as presented in [**?**] (see also [**?**]). The codes, written in Julia [**?**], are available at the following link: `http://www.github.com/dprn/WCvsLHE`.

All the considered images are of size 200×200 pixels, and take values in the interval $[.15, .85]$ in order to avoid out-of-range issues. We always consider $K = 30$ discretised orientations, as done in [**?**] for instance. As presented in Appendix **??**, the receptive profiles associated to the discretised orientations selected are obtained via cake wavelets [**?**], for which the frequency band `bw` is set to `bw`$= 5$. The interaction kernel is taken to be a 2D or 3D Gaussian with standard deviation $\boldsymbol{\sigma}_\omega$, the local mean average μ is obtained via Gaussian filtering with standard deviation $\boldsymbol{\sigma}_\mu$. In our experiments we used the following two piece-wise linear functions as sigmoids:

$$\hat{\sigma}(\rho) := \min\{1, \max\{\alpha\rho, -1\}\}, \qquad \sigma(\rho) := -\hat{\sigma}\left(x - \frac{1}{2}\right), \tag{2.10}$$

with $\alpha = 5$, see Figure **??**. Note that $\hat{\sigma}$, which will be used for LHE models, is odd and centered in zero while σ, which will be used for WC models, is shifted in $1/2$ and shows a reversed behaviour. This in fact corresponds to a change of sign in the integral terms of LHE models w.r.t. the WC ones, as discussed in Section **??**.



(a) $\hat{\sigma}$ and the line $y = x$

(b) σ and the line $y = -x + \frac{1}{2}$

Figure 3: Sigmoid functions in the form (**??**), with $\alpha = 5$, as considered in our experiments.

Finally, the evolution stops when the L^2 relative distance between two successive iterations is smaller than a tolerance $\tau = 10^{-2}$, which identifies convergence of the iterates to a stationary state.

3 Results

In this section, we present the results obtained by applying the four models described above to the visual illusions described in Section **??**. Our objective is to understand the capability of these models to *replicate* the visual illusions under consideration. That is, we are interested in whether the output produced by the models qualitatively agrees with the human perception of the phenomena. We stress that our study is purely qualitative as it has to be intended as a proof of concept showing how Wilson-Cowan-type dynamics can be effectively used to replicate the perceptual effects due to the observation of visual illusions. We do not address here the match with empirical data since those depend on several experimental conditions for which a correspondence with the model parameters is not clear. A dedicated study on experiments motivated by psychophysics, addressing the validation of our models and, possibly, allowing for the creation of ground-truth references for a quantitative assessment is left for future research.

Due to the lack of a universal metric adapted to the task of assessing the replication of visual illusions, we will evaluate replication or lack thereof by presenting relevant line profiles, i.e., plots of brightness levels along a single row (line), of images produced by the four models in consideration (a common tool used by several brightness/lightness/color models before [**?**, **?**]). These lines are chosen as to cross a section of the image called *target*: A gray region in the image (or set of regions in the case of the Chevreul illusion), where the brightness illusion appears.

In all the results shown in this section, the original visual stimulus profile is represented as a blue dashed line. The line profiles of the output models are represented as solid red (**??**), green (**??**), magenta (**??**), and cyan (**??**) lines.

The parameters appearing in the models have been chosen independently for each illusion and each model, in order to obtain the best possible replication of the visual illusion. Here, by best-replication we mean that the extracted line-profiles correctly mimic the perceptual outcome from a qualitative point of view.The chosen parameters are presented in Table **??**.

	WC-2D				LHE-2D				WC-3D				LHE-3D			
Illusion	σ_μ	σ_ω	λ	M	σ_μ	σ_ω	λ	M	σ_μ	σ_ω	λ	M	σ_μ	σ_ω	λ	M
White	10	20	.7	1.4	10	50	.7	1	20	30	.7	1.4	2	50	.7	1
Brightness	2	10	.7	1.4	2	10	.7	1	2	10	.7	1.4	2	10	.7	1
Checkerboard	10	70	.7	1.4	10	70	.7	1	10	70	.7	1.4	10	70	.7	1
Chevreul	2	5	.7	1	2	10	.7	1	2	40	.5	1	5	7	.7	1
Chevreul canc.	2	2	.9	1	5	3	.9	1	2	20	.5	1.4	5	3	.9	1
Dungeon	6	10	.7	1.4	5	40	.7	1	2	50	.7	1.4	5	50	.7	1
Gratings	2	6	.7	1	2	6	.7	1	2	6	.7	1	2	6	.7	1
Hong-Shevell	5	20	.7	1	5	.5	.7	1	10	30	.7	1	10	30	.7	1
Luminance	2	6	.7	1	2	6	.7	1	2	6	.7	1	2	6	.7	1
Poggendorff	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	3	10	.5	1
Tilt	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	15	20	.7	1

Table 1: Parameters used in the tests.

3.1 Orientation-independent brightness illusions

Table **??** summarises the replication results obtained for the illusions described in Section **??**: if the model replicates the illusion we indicate in the table the used parameters, otherwise a cross (✗) denotes no replication, i.e. the failure of the model to reproduce computational results corresponding to the visual perception of the considered illusion.

White's illusion. The chosen line profile for the plots in Fig. **??** corresponds to the central horizontal line of the image, which crosses both gray patches. As both plots show, all four models correctly predict the left target to be darker than the right one.

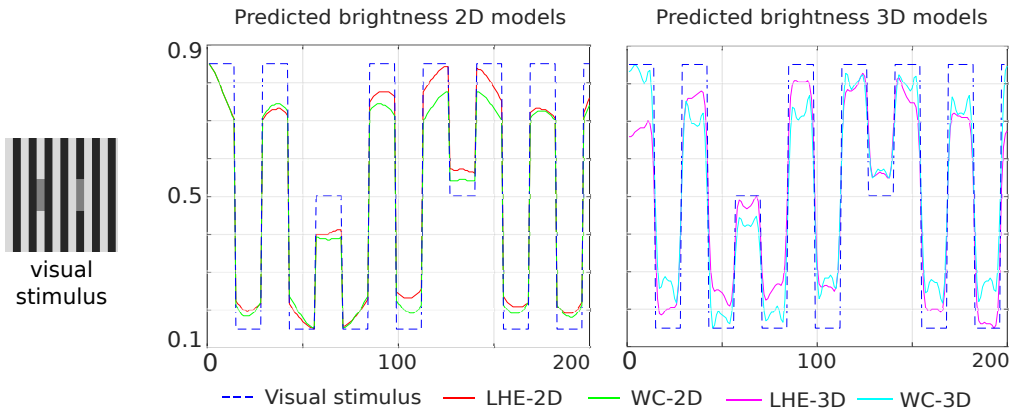


Figure 4: Predicted brightness in White illusion

Simultaneous brightness contrast. The plots in Fig. **??** show the line profiles of the central horizontal line of the image, which crosses the two gray squares. We see that our four models replicate this illusion (left square lighter than the right square). In both the 2D and the 3D case, we observe that LHE methods result in an enhanced contrast effect w.r.t. WC methods.

Checkerboard illusion. The chosen line profiles for this illusion are the two horizontal lines crossing, respectively, the left gray target and the right one. In Fig. **??**, we chose to plot the first half of the line profile corresponding to the left target and the second half of the one corresponding to the right target. The profiles of all the four models show replication of this illusion, by which the left target is perceived darker than the right one.

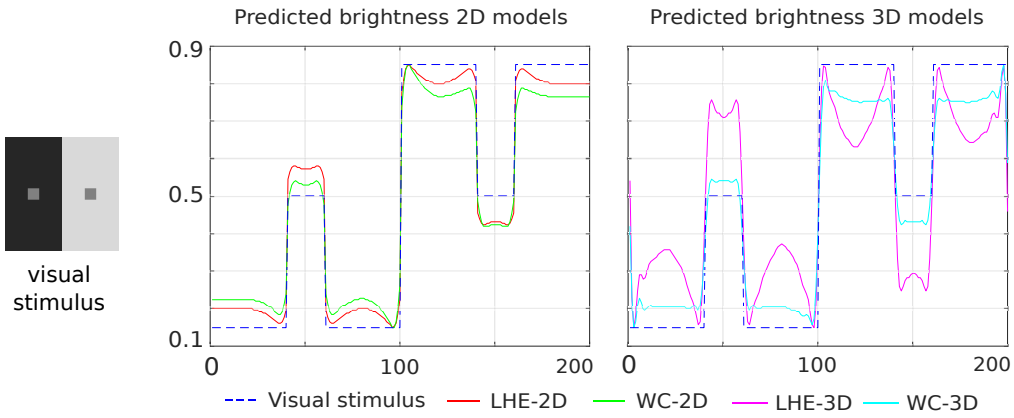


Figure 5: Predicted brightness in simultaneous brightness contrast

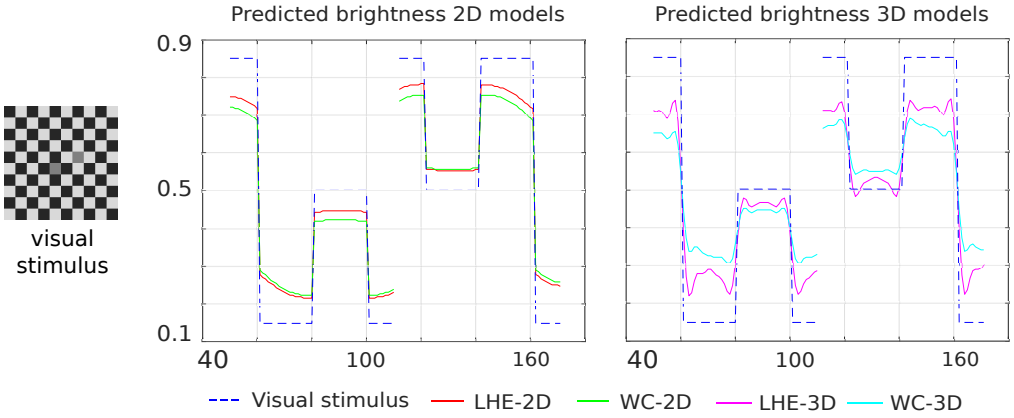


Figure 6: Predicted brightness in Checkerboard illusion

Chevreul illusion. Fig. **??** presents the line profiles for the central horizontal line.All four models correctly replicate the perceived changes within each band.

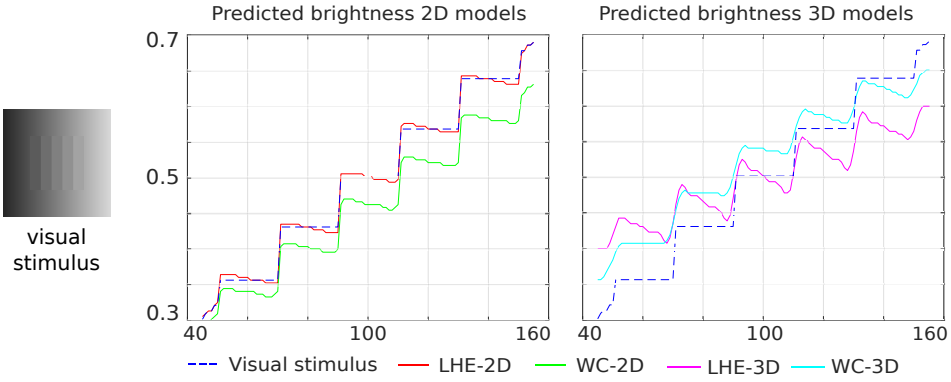


Figure 7: Predicted brightness in Chevreul illusion

.

Chevreul cancellation. The line profiles for the central horizontal line are presented in Fig. **??**. In this case all models are able to correctly replicate the effect, although in the case of (**??**) and (**??**) the perceptual response is not perfect, due to the presence of some oscillations. We also remark that the correct replication of this illusion is extremely sensitive to the chosen parameters.

Dungeon illusion. Profiles of the central section (3 middle squares) of each target are shown in Fig. **??**. The first part of the plot (left to right) represents the profile of the rectangle on black

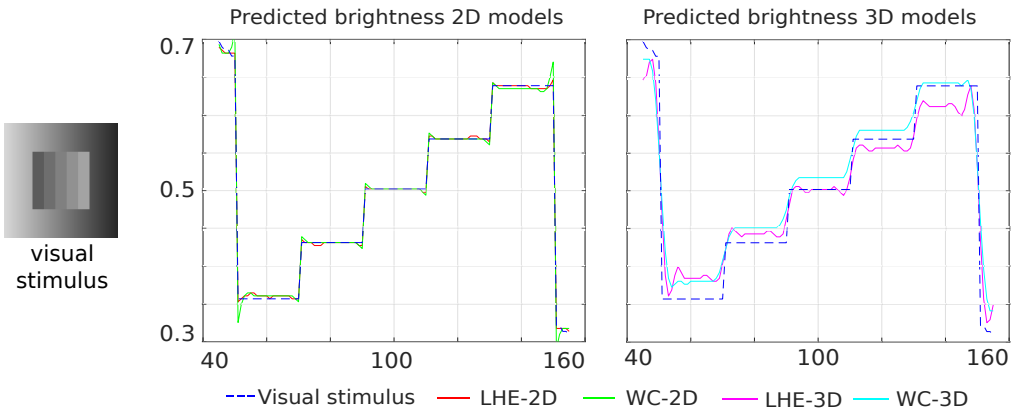


Figure 8: Predicted brightness in Chevreul cancellation

background. The second plot shows the target on white background. As these profiles show, our four proposed models replicate human perception (first target is predicted as darker than the second). Nevertheless, the assimilation effect (target intensity goes towards surrounding) is stronger in the 3D models.

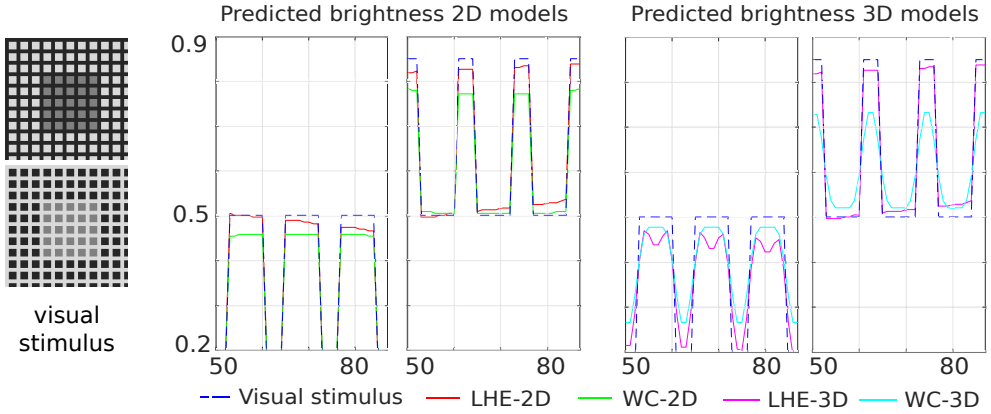


Figure 9: Predicted brightness in Dungeon illusion

Grating induction. In Fig. **??** the continuous and dashed blue lines respectively show the profile of the grating and of the central horizontal line row of the visual stimulus. Then, the line profiles of the central horizontal line of the outputs have been plotted. We observe that for both 2D and 3D models a counter-phase grating appears in the middle row, which successfully coincides with human perception. Notice that LHE methods have a higher amplitude in both cases.

Hong-Shevell illusion. Fig. **??** shows the line profiles of the central horizontal line around the target (gray ring) neighbourhood rings in the first half of the image. As in the case of the Dungeon illusion, we present in the first half of the plot (left to right) the output of the first stimulus (light background) and in the second half the output of the second (dark background). We see how our four proposed models replicate the assimilation effect. Hence, the gray ring in the first image is predicted as brighter than the gray ring in the second visual stimulus.

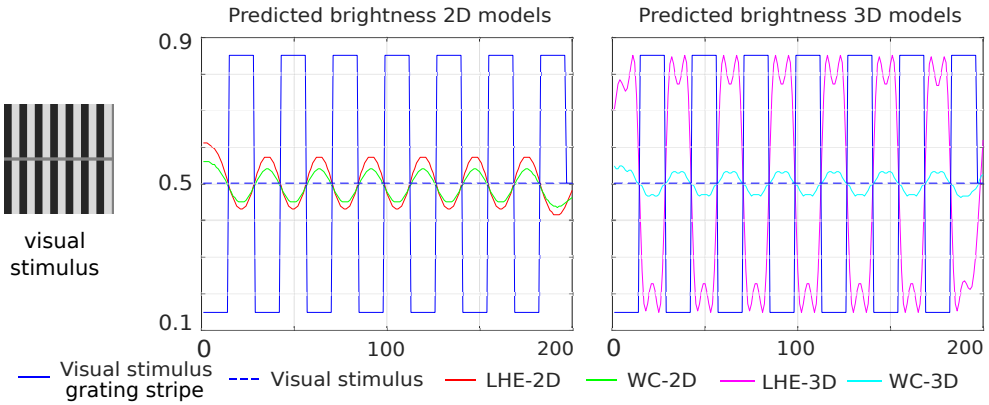


Figure 10: Predicted brightness in grating induction

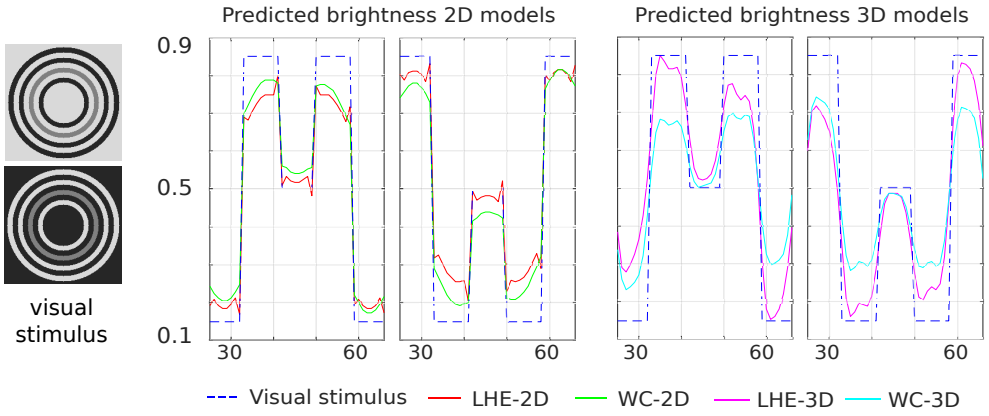


Figure 11: Predicted brightness in Hong-Shevell illusion

Luminance illusion. Horizontal profiles crossing top left and right targets (gray circles) are depicted in Fig. **??**. For each target our four models reconstruct the left target as brighter than the right one. Hence, all models correctly predict this contrast effect. In this case, LHE presents a higher contrast response in both responses (2D and 3D).

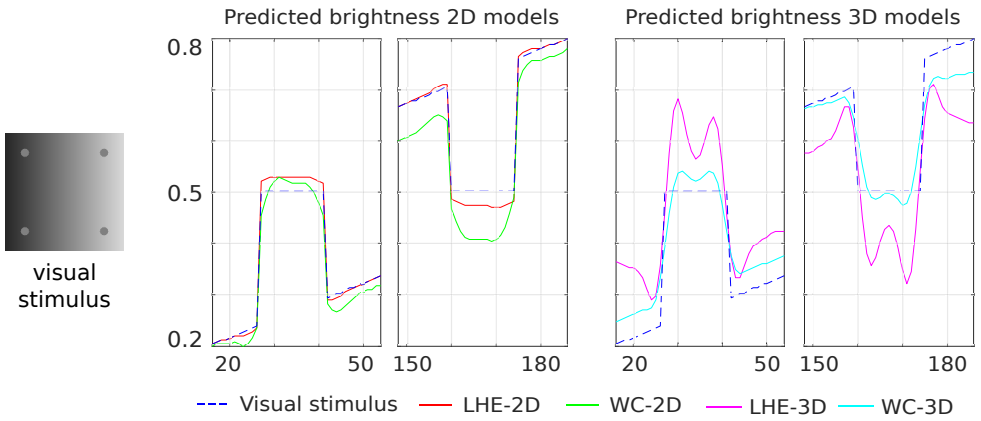


Figure 12: Predicted brightness in luminance gradient illusion

We observe that in all the considered brightness illusions both the 3D methods present neighbourhood-dependent oscillations.

3.2 Orientation-dependent illusions

Poggendorff illusion. The output images and a zoom of the target gray middle area are presented in Fig. **??**. In this case (**??**), (**??**), and (**??**) are not able to completely replicate the illusion, since induced white lines on the gray area are not connected. On the other hand, (**??**) successfully replicates the perceptual completion over the gray middle stripe.

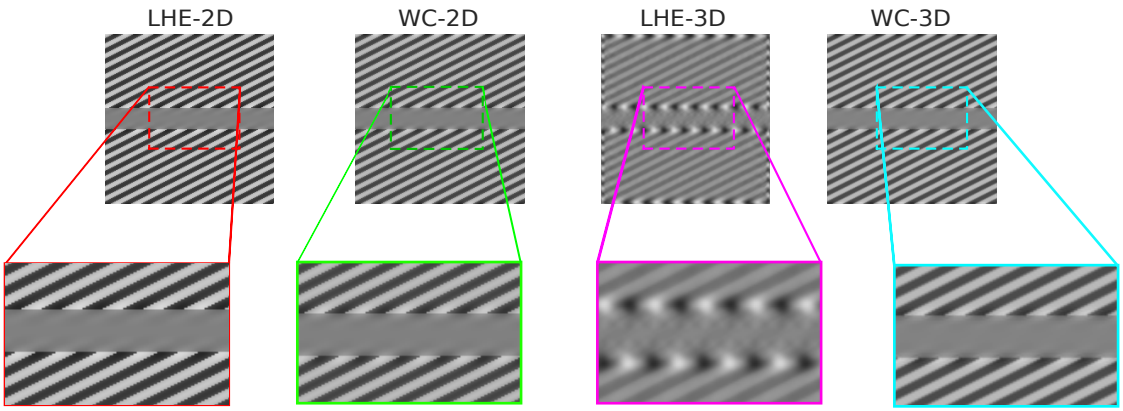


Figure 13: Zoom of the predicted completion for Poggendorff illusion

Tilt illusion. In Fig. **??** we present line profiles, for both visual stimuli, for a diagonal line starting at the bottom left corner of the image and ending at the top right one. In order to be able to correctly compare the two images, the line profile of the second image (from top to bottom) has been extracted after flipping the outer circle along the vertical axis, so that the responses to both stimulus have the same background. Although there is a noticeable effect, such as a reduction in contrast for the (**??**), the difference between the responses to the two stimuli is very mild for all models with the exception of (**??**).

The fact that indeed this model is replicating the effect can be better appreciated looking at Fig. **??**, which shows a composite of the inner circle for the responses to the two visual stimuli of the two orientation-dependent models. It is then evident that the (**??**) model yields a stronger result than the (**??**) one. In fact, the former shows increased visibility (measured here as the contrast) for the half of the circle corresponding to the second stimulus than the one corresponding to the first stimulus. On the other hand, in the case of the (**??**) model (or of 2D models, not depicted here), the circle shows no difference among its two halves. This justifies our claim that the (**??**) model can increase the visibility of the inner circle (replicate the illusion) based on the orientation of the outer circle.

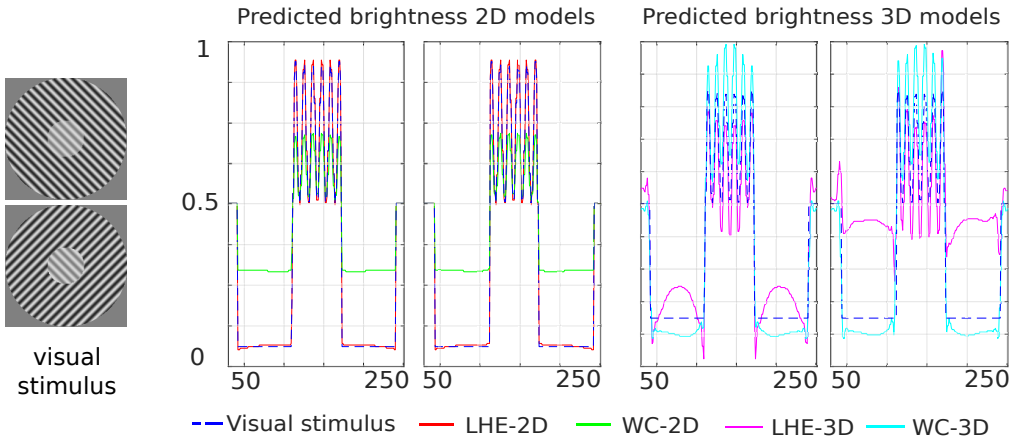


Figure 14: Predicted brightness in Tilt illusion

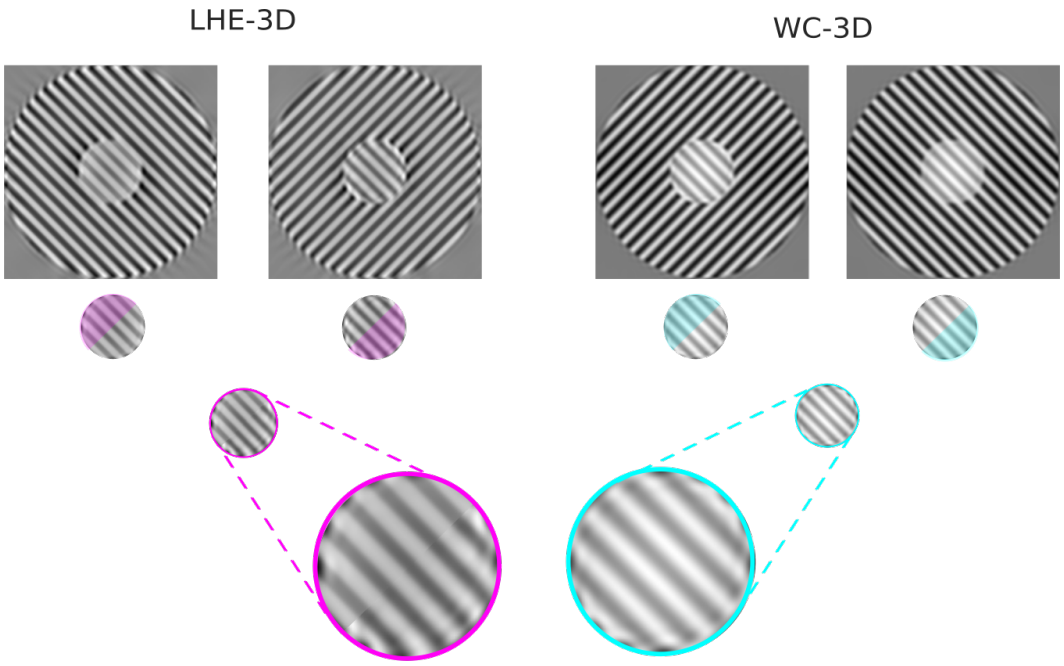


Figure 15: Detail in predicted brightness in Tilt illusion

4 Discussion

The results presented in the previous section show that the four models are able to reproduce several brightness illusions. Concerning orientation-dependent illusions we observe that, as expected, 2D models cannot reproduce them, while the only 3D model that correctly reproduces the perceptual outcome is the (**??**). However we stress that determining replication or lack thereof in the Tilt illusion is subtle, as the observed effects are very mild.

As already mentioned, the parameters of the presented results are chosen independently from one illusion to the other in order to qualitatively optimise the perceptual replication in terms of suitable line profiles. Empirical observations show that the value of the model parameters involved are indeed related with the size of the target and the spatial frequency of the background. Nevertheless, if one settles for milder replications, it would be possible to choose more uniform parameters. For instance, this happens for the (**??**) model in the Chevreul and Chevreul cancellation illusions, which can be reproduced simultaneously with parameters $\boldsymbol{\sigma}_\mu = 3$ and $\boldsymbol{\sigma}_\omega = 30$, although with less striking results.

Regarding the 3D models, we want to point out that we have chosen to use $K = 30$ orientations whereas this number commonly takes values in the 12-18 range in the literature (e.g. [**?**, **?**, **?**]). Our selection of 30 orientations is motivated by some preliminary tests (which we are not presenting here) showing that a coarser orientation discretisation seems not to be sufficient to reproduce most of the orientation-dependent illusions. As future research we will test whether or not a different selection of parameters allows to reproduce those illusions with less orientations, but we should also mention that some works in the literature actually use a high number of orientations in cortical models (e.g. 64 orientations in [**?**]).

Finally, we notice that the output of 3D models often shows oscillations. For some illusions (white and dungeon), the (**??**) model produces more oscillatory solutions than (**??**), and for others (Chevreul brightness, grating induction, and luminace gradient), the (**??**) have stronger oscillations than (**??**). The relation between the model parameters and possible dependence of the target surrounding is a matter of future research.

5 Conclusions and future work

We consider Wilson-Cowan-type models describing neuronal dynamics and apply them to the study of replication of brightness visual illusions.

We show that Wilson-Cowan equations are able to replicate a number of brightness illusions and that their variational modification, accounting for changes in the local contrast and performing local histogram equalisation, outperforms them. We consider also extensions of both models accounting for explicit local orientation dependence, in agreement with the architecture of V1. Although in the case of pure brightness illusions we found no real advantage in considering models taking into account orientations, these turned out to be necessary for the replication of two exemplary orientation-dependent illusions, which only the 3D LHE variational model is able to reproduce.

In order to understand and fully exploit the potential of the orientation-dependent LHE model, further research should be done. In particular, a more accurate modelling reflecting the actual structure of V1 should be addressed. This concerns first the lift operation, where the cake wavelet should be replaced by the more physiologically plausible Gabor filters, as well as the interaction weight ω which could be taken to be the anisotropic heat kernel of [**?**, **?**, **?**]. The design of appropriate psychophysics experiments testing the visual illusions considered in this work and their match with our models' outputs is clearly a further important research direction, which would turn our qualitative study into a quantitative one. The problem of matching computational models of perception with psychophysical data is in fact not trivial, but necessary to provide insights about how visual perception works and to identify the computational parameters able to reproduce the perceptual bias induced by these phenomena.

Acknowledgements and Grants

M. B. would like to thank the organizers of the conference to celebrate Jack Cowan's 50 years at the University of Chicago for their kind invitation to attend that meeting, which served as inspiration for this work, and also acknowledges the support of 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 of the Spanish government and FEDER Fund, grant ref. PGC2018-099651-B-I00 (MCIU/AEI/FEDER, UE). L. C., V. F. and D. P. acknowledge the support of a public grant overseen by the French National Research Agency (ANR) as part of the *Investissement d'avenir program*, through the iCODE project funded by the IDEX Paris-Saclay, ANR-11-IDEX-0003-02 and of the research project *LiftME* funded by INS2I, CNRS. L. C. and V. F. acknowledge the support provided by the *Fondation Mathématique Jacques Hadamard*. V. F. acknowledges the support received from the European Union's Horizon 2020 research and innovation programme under the *Marie Skłodowska-Curie grant No 794592*. V. F. and D. P. also acknowledge the support of ANR-15-CE40-0018 project *SRGI - Sub-Riemannian Geometry and Interactions*. B. F. acknowledges the support of the Fondation Asile des Aveugles.

Disclosures

All authors declare that there is no commercial relationship relevant to the subject matter of presentation.

Author contributions

All authors equally contributed to this work.

A Non-variational nature of Wilson-Cowan equation

In this section we show that, for non-trivial choices of weight and sigmoid functions, Wilson-Cowan equations do not admit a variational formulation.

For the sake of simplicity, we will consider only a finite dimensional variant of Wilson-Cowan equations, with constant input. Namely, for $a : \mathbb{R} \to \mathbb{R}^n$, we consider

$$\frac{d}{dt}a(t) = -\mu a(t) + W\sigma(a(t)) + h. \tag{A.1}$$

Here, $h \in \mathbb{R}^n$ is the input, $\mu > 0$ is a parameter, $\sigma \in C^1(\mathbb{R})$ is any function (we denoted $\sigma(v) = (\sigma(v_i))_i$ for $v \in \mathbb{R}^n$), and $W \in \mathbb{R}^{n\times n}$ is a symmetric interaction kernel. For a proof in the infinite-dimensional setting we refer to [**?**]

Equation (**??**) admits a variational formulation if it can be written as the steepest descent associated with a functional $J : \mathbb{R}^n \to \mathbb{R}$, i.e.,

$$\frac{d}{dt}a(t) = -\nabla J(a(t)). \tag{A.2}$$

We have the following.

Theorem. *The Wilson-Cowan equation* (**??**) *admits the variational formulation* (**??**) *only if either W is a diagonal matrix, or σ is an affine function, i.e., $\sigma(x) = \alpha x + \beta$ for some $\alpha, \beta \in \mathbb{R}$.*

Proof. Writing (**??**) and (**??**) componentwise, we find the following relation for J:

$$\partial_i J(v) = \mu v_i - \sum_k W_{\ell,k}\sigma(v_\ell) - h_i, \qquad v = (v_1, \dots, v_n) \in \mathbb{R}^n, \quad i = 1, \dots, n.$$

By differentiating again the above, and letting δ_{ij} denote the Kroenecker delta symbol, we have

$$\partial_{ij} J(v) = \mu\delta_{ij} - \sum_k W_{\ell,k}\sigma'(v_\ell)\delta_{j\ell} = \mu\delta_{ij} - W_{ij}\sigma'(v_j), \qquad i, j = 1, \dots, n. \tag{A.3}$$

Namely, $\operatorname{Hess} J(v) = (\mu\delta_{ij} - W_{ij}\sigma'(v_j))_{ij}$. Assume that W is not a diagonal matrix. Then, since both the Hessian matrix and W are symmetric, by choosing $i \neq j$ such that $W_{ij} \neq 0$ we get

$$\sigma'(v_i) = \sigma'(v_j) \qquad \forall v \in \mathbb{R}^n. \tag{A.4}$$

This clearly implies that σ' is constant, thus showing that σ must be an affine function. □

We observe that the above reasoning does not apply to the LHE algorithm. Indeed, the discrete form of the latter is

$$\frac{d}{dt}a(t) = -\mu a(t) + \sum_\ell W_{i\ell}\sigma\big(a_i(t) - a_\ell(t)\big) + h. \tag{LHE}$$

Then, the corresponding variational equation (for $\mu = 0$ and $h = 0$) is

$$\partial_i J(v) = -\sum_{\ell\neq i} W_{i\ell}\sigma(v_i - v_\ell), \qquad v \in \mathbb{R}^n. \tag{A.5}$$

This yields

$$\partial_{ji} J(v) = W_{ij}\sigma'(v_i - v_j), \qquad \text{for } v \in \mathbb{R}^n, \quad i \neq j. \tag{A.6}$$

This does not contradict the symmetry of the Hessian, as σ was chosen to be odd an thus σ' is even. Indeed, we know by [**?**] that we can let

$$J(v) := \sum_{k,\ell} W_{k\ell}\Sigma(v_k - v_\ell), \tag{A.7}$$

where Σ is such that $\Sigma' = \sigma$.

B Encoding orientation-dependence via cortical-inspired models

Orientation dependence of the visual stimulus is encoded via cortical inspired techniques, following e.g., [**?**, **?**, **?**, **?**, **?**]. The main idea at the base of these works goes back to the 1959 paper [**?**] by Hubel and Wiesel (Nobel prize in 1981) who discovered the so-called *hypercolumn functional architecture* of the visual cortex V1. Following [**?**], each neuron ξ in V1 detects couples (x, θ) where $x \in \mathbb{R}^2$ is a retinal position and θ is a direction at x. Orientation preferences θ are then organised in hypercolumns over the retinal position x, see [**?**, Section 2].

Let $Q \subset \mathbb{R}^2$ be the visual plane. To a visual stimulus $f : Q \to [0, 1]$ is associated a cortical activation $Lf : Q \times [0, \pi) \to \mathbb{R}$ such that $Lf(\xi)$ encodes the response of the neuron $\xi = (x, \theta)$. Letting $\psi_\xi \in L^2(\mathbb{R}^2)$ be the receptive profile (RP) of the neuron ξ, such response is assumed to be given by

$$Lf(\xi) = \langle \psi_\xi, f \rangle_{L^2(\mathbb{R}^2)} = \int_Q \overline{\psi_\xi(x)} f(x)\, dx. \tag{B.1}$$

Motivated by neuro-phyisiological evidence, we assume that RPs of different neurons are "deducible" one from the other via a linear transformation. As detailed in [**?**, **?**], see also [**?**, Section 3.1], this amounts to the fact that the linear operator $L : L^2(Q) \to L^2(Q \times [0, \pi))$ is a continuous wavelet transform (also called *invertible orientation score transform*). That is, there exists a *mother wavelet* $\Psi \in L^2(\mathbb{R}^2)$ such that $Lf(x, \theta) = \big[f * (\Psi^* \circ R_{-\theta})\big](x)$. Here, $f * g$ denotes the standard convolution on $L^2(\mathbb{R}^2)$ and $R_{-\theta}$ is the counter-clock-wise rotation of angle θ. Notice that, although images are functions of $L^2(\mathbb{R}^2)$ with values in $[0, 1]$, it is in general not true that $Lf(x, \theta) \in [0, 1]$.

Concerning the choice of the mother wavelet, we remark that neuro-physiological evidence suggests that a good fit for the RPs is given by Gabor filters, whose Fourier transform is the product of a Gaussian with an oriented plane wave [**?**]. However, these filters are quite challenging to invert, and are parametrised on a bigger space than $\mathcal{M}$, which takes into account also the frequency of the plane wave and not only its orientation. For this reason, in this work we instead considered *cake wavelets*, introduced in [**?**, **?**]. These are obtained via a mother wavelet Ψ^{cake} whose support in the Fourier domain is concentrated on a fixed slice, depending on the number of orientations one aims to consider in the numerical implementation. For the sake of integrability, the Fourier transform of this mother wavelet is then smoothly cut off via a low-pass filtering, see [**?**, Section 2.3] for details. Observe, however, that, since we are considering orientations on $[0, \pi)$ and not directions on $[0, 2\pi)$, we choose a non-oriented version of the mother wavelet, given by $\tilde{\psi}^{cake}(\omega) + \tilde{\psi}^{cake}(e^{i\pi}\omega)$, in the notations of [**?**].

An important feature of cake wavelets is that, in order to recover the original stimulus from its cortical activation, it suffices to simply "project" the cortical activations along hypercolumns. This yields

$$f(x) := \frac{1}{\pi} \int_0^\pi Lf(x, \theta)\, d\theta. \tag{B.2}$$

This justify the assumption, implicit in equation (**??**), that the projection of a cortical activation F (not necessarily given by a visual stimulus) to the visual plane is given by

$$PF(x) = \frac{1}{\pi} \int_0^\pi F(x, \theta)\, d\theta. \tag{B.3}$$