



# Cortical Divisive Normalization from Wilson–Cowan Neural Dynamics

Jesús Malo<sup>1</sup> · José Juan Esteve-Taboada<sup>1</sup> · Marcelo Bertalmío<sup>2</sup>

Received: 26 October 2022 / Accepted: 21 December 2023 / Published online: 15 February 2024  
© The Author(s) 2024

## Abstract

Divisive Normalization and the Wilson–Cowan equations are well-known influential models of nonlinear neural interaction (Carandini and Heeger in *Nat Rev Neurosci* 13(1):51, 2012; Wilson and Cowan in *Kybernetik* 13(2):55, 1973). However, they have been always treated as different approaches and have not been analytically related yet. In this work, we show that Divisive Normalization can be derived from the Wilson–Cowan dynamics. Specifically, assuming that Divisive Normalization is the steady state of the Wilson–Cowan differential equations, we find that the kernel that controls neural interactions in Divisive Normalization depends on the Wilson–Cowan kernel but also depends on the signal. A standard stability analysis of a Wilson–Cowan model with the parameters obtained from our relation shows that the Divisive Normalization solution is a stable node. This stability suggests the appropriateness of our steady state assumption. The proposed theory provides a mechanistic foundation for the suggestions that have been done on the need of signal-dependent Divisive Normalization in Coen-Cagli et al. (*PLoS Comput Biol* 8(3):e1002405, 2012). Moreover, this theory explains the modifications that had to be introduced ad hoc in Gaussian kernels of Divisive Normalization in Martínez-García et al. (*Front Neurosci* 13:8, 2019) to reproduce contrast responses in V1 cortex. Finally, the derived relation implies that the Wilson–Cowan dynamics also reproduce visual masking and subjective image distortion, which up to now had been explained mainly via Divisive Normalization.

**Keywords** Neural networks · Visual cortex · Nonlinear neural interactions · Divisive Normalization · Wilson–Cowan equations · Visual masking · Contrast perception

**Mathematics Subject Classification** 34 · 35 · 37 · 94

---

Communicated by Kyle Wedgwood.

---

✉ Jesús Malo  
jesus.malo@uv.es

<sup>1</sup> Image Processing Lab, Universitat de València, Valencia, Spain

<sup>2</sup> Spanish National Research Council (CSIC), Madrid, Spain

## 1 Introduction

A number of perceptual experiences in different modalities can be described with the Divisive Normalization interaction among the outputs of linear neurons (Carandini and Heeger 1994, 2012). In particular, in vision, the perception of color (Hillis and Brainard 2005), texture (Watson and Solomon 1997), and motion (Simoncelli and Heeger 1998) seem to be mediated by this nonlinear interaction. Intuitively, this *divisive* interaction modifies the response of a sensor by normalizing it with the responses of the neighbor neurons, thus explaining inhibition by the surround.

The discussion on the circuits underlying the Divisive Normalization in Carandini and Heeger (2012) suggests that there may be different architectures leading to this specific computation. Recent results suggest specific mechanisms for Divisive Normalization in certain situations (Sato et al. 2016), but the debate on the physiological implementations is still open. On the other hand, a number of functional advantages (Schwartz et al. 2009; Schwartz and Rieke 2011; Coen-Cagli et al. 2012; Coen-Cagli and Schwartz 2013) suggest that the kernel that describes the interaction in Divisive Normalization should be adaptive (i.e., signal or context dependent). Moreover, the match between the linear receptive fields and the interaction kernel in the Divisive Normalization is not trivial: The conventional Gaussian kernels in Watson and Solomon (1997), Malo and Laparra (2010) had to be tuned by hand to reproduce contrast responses (Martinez-Garcia et al. 2019).

These open questions imply that it is interesting to relate Divisive Normalization to other models of neural interaction for a better understanding of its implementation, the structure of the interaction kernel, and its eventual dependence with the signal. Interesting possibilities to consider are the classical dynamic neural field models of Wilson–Cowan (Wilson and Cowan 1972, 1973; Bressloff and Cowan 2003), Amari (Amari 1977), or Grossberg (Grossberg 1968, 1988), all of which have a similar subtractive nature (Chow and Karimipanih 2020): In subtractive models, the surround modifies the activity of a given sensor by subtracting a weighted average of its neighbor's responses, as opposed to the division made in normalization models.

Subtractive and divisive adaptation models have been qualitatively related before (Wilson and Humanski 1993; Bressloff et al. 2002). Both models have been shown to have similar advantages in information-theoretic terms: The Wilson–Cowan interaction in a neural layer uniformizes the probability density of the responses (Bertalmio 2014), and there is less redundancy among them (Gomez-Villa et al. 2020). Similarly, Divisive Normalization layers reduce the relations between the responses (Malo et al. 2006), factorize the joint probability of the responses (Malo and Laparra 2010), and maximize the transmitted information from the input (Malo 2020, 2022). Additionally, both models provide similar descriptions of pattern discrimination (Wilson and Humanski 1993; Bertalmio et al. 2017). However, despite all these similarities, no direct analytical correspondence has been established between these models yet.

In this paper, we assume that the psychophysical behavior described by Divisive Normalization comes from underlying neural interactions that follow the Wilson–Cowan equations. In particular, we identify the Divisive Normalization response with the stationary regime of a Wilson–Cowan system. From this identification, we derive

an expression for the Divisive Normalization kernel in terms of the interaction kernel of the Wilson–Cowan equations.

This analytically derived relation has the following interesting consequences:

- (1) It has been suggested that Divisive Normalization should depend on the input because of functional reasons (Coen-Cagli et al. 2012; Coen-Cagli and Schwartz 2013), but no physiological mechanism was proposed to implement this statistical adaptation. The proposed relation of the Divisive Normalization with a dynamical system with fixed wiring among neurons provides a mechanistic explanation for this dependence with the input.
- (2) The relation explains the modifications that had to be introduced ad hoc in the kernel of Divisive Normalization in Martinez-Garcia et al. (2019) to reproduce contrast responses. This implies that the Wilson–Cowan dynamics reproduce visual masking, which up to now had been mainly explained via Divisive Normalization (Foley 1994; Watson and Solomon 1997).
- (3) The relation allows to build effective image quality metrics based on the Wilson–Cowan model, something which, to the best of our knowledge, has not been considered before in the literature, as opposed to the many examples of metrics based on Divisive Normalization (Teo and Heeger 1994; Laparra et al. 2010; Malo and Laparra 2010; Laparra et al. 2017; Hepburn et al. 2020).
- (4) A standard stability analysis of a Wilson–Cowan model with the parameters obtained from our relation shows that the Divisive Normalization solution is a stable node of the dynamic model. This shows the appropriateness of our steady-state assumption. Moreover, this stability justifies the straightforward use of Divisive Normalization with time-varying stimuli, as in Simoncelli and Heeger (1998).

The structure of the paper is as follows. The *Materials and Methods* section reviews the retina-V1 neural path and the contrast perception of visual patterns. We also introduce the notation of the models: the Divisive Normalization and the Wilson–Cowan equations. Besides, we recall some experimental facts that will be used to illustrate the performance of the proposed relation: (1) Contrast responses curves imply certain interactions between subbands (Cavanaugh 2000; Watson and Solomon 1997), (2) the Divisive Normalization kernel must have a specific structure [identified in Martinez-Garcia et al. (2019)] to reproduce contrast response curves, and (3) the shape of the Divisive Normalization kernel should have a specific dependence with the surrounding signal (Cavanaugh et al. 2002a, b). In the *Analytical Results* section, we derive the relation between the Divisive Normalization and the Wilson–Cowan equations based on the steady-state assumption. The *Numerical Experiments* section illustrates with simulations the mathematical properties and the perceptual consequences of the proposed relation. First, we experimentally check the convergence of the Wilson–Cowan solution to the Divisive Normalization response in a wide range of model parameterizations. We quantify the error introduced by the approximations done to get the analytical results. Moreover, we illustrate the appropriateness of the steady-state assumption by showing that the Divisive Normalization is a stable node of the Wilson–Cowan system. Then, we address contrast perception facts using the proposed relation to build a psychophysically meaningful Wilson–Cowan model: We theoretically derive the specific structure of the kernel that was previously inferred empirically (Martinez-Garcia et al.

2019), we show that the proposed interaction kernel adapts with the signal, and as a result, we reproduce general trends of contrast response curves. Finally, we discuss the use of the derived kernel in predicting the perceptual metric of the image space. The *Final Remarks* section concludes the paper.

## 2 Materials and Methods

In this work, the theory is illustrated in the context of models of the retina-cortex pathway. The considered framework follows the approach suggested in Carandini and Heeger (2012): a cascade of four isomorphic *linear+nonlinear* modules. These four modules sequentially address brightness, contrast, frequency filtered contrast masked in the spatial domain, and orientation/scale masking. An example of the transforms of the input in such models is shown in Fig. 1.

In this general context, we focus on the cortical (fourth) layer: a set of linear sensors with wavelet-like receptive fields modeling simple cells in V1, and the nonlinear interaction between the responses of these linear sensors. Divisive Normalization has been the conventional model used for the nonlinearity to describe contrast perception psychophysics (Watson and Solomon 1997), but here we will explore the application of the Wilson–Cowan model in the contrast perception context.

Below we introduce the notation of both models of neural interaction and the facts on contrast perception that should be explained by the models.

### 2.1 Modeling Cortical Interactions

Our focus here is the last *linear+nonlinear* module of the retina-V1 cascade in Fig. 1, and specifically the nonlinear layer that describes the interactions in the primary visual cortex V1. The *driving input* of this final nonlinear layer is the vector of energies,  $\mathbf{e}$ , of the responses of linear wavelet-like simple cells, and the *output* of this interaction is the vector of nonlinear responses  $\mathbf{x}$ :

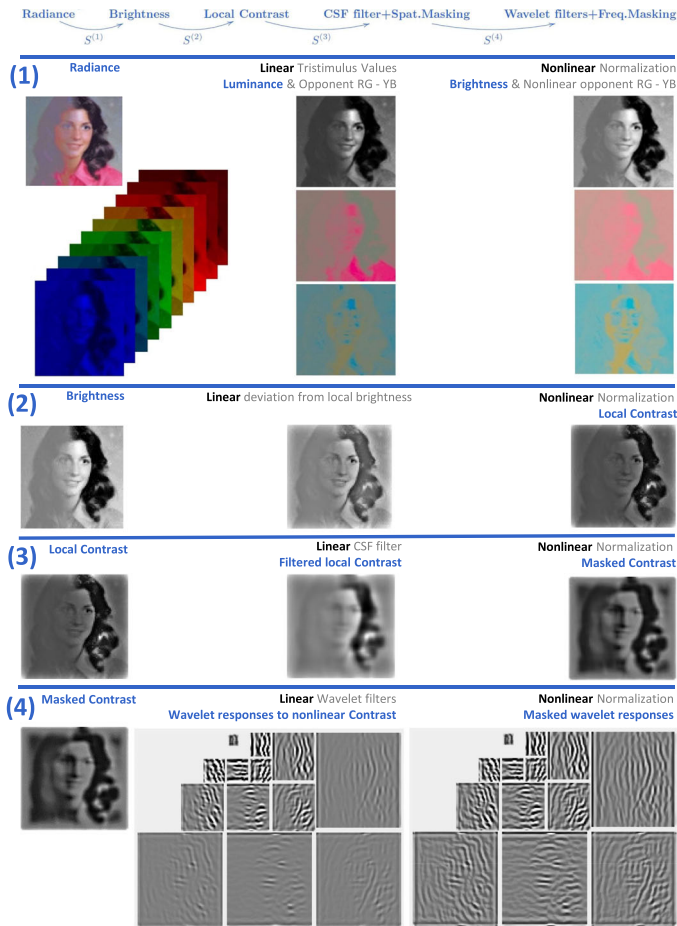
$$\mathbf{e} \xrightarrow{\mathcal{N}} \mathbf{x} \quad (1)$$

In this work, the two models considered describe the interaction  $\mathcal{N}$  between the linear simple cells in V1. The Wilson–Cowan equations model neural firing rate dynamics that may converge to a steady state. If that is the case, the long-term behavior of the Wilson–Cowan equations may be similar to the Divisive Normalization model, since the latter models static neural firing rates.

### 2.2 The Divisive Normalization model

*Forward transform.*

The conventional expression of the canonical Divisive Normalization (Carandini and Heeger 2012) uses an element-wise formulation:



**Fig. 1** Signal transforms in the retina-cortex pathway: a cascade of linear + nonlinear modules [example from Martinez-Garcia et al. (2018)]. The input is the spatial distribution of the *spectral irradiance* at the retina. (1) The linear part of the first layer consist of three positive spectral sensitivities (Long, Medium, Short, LMS, wavelengths) and a linear recombination of the LMS values with positive/negative weights. This leads to three tristimulus values in each spatial location: One of them is proportional to the luminance, and the other two have opponent chromatic meaning (red-green and yellow-blue). These linear tristimulus responses undergo adaptive saturation transforms. Perception of *brightness* is mediated by an adaptive Weber-like nonlinearity applied to the luminance at each location. This nonlinearity enhances the response in the regions with small linear input (low luminance). (2) The linear part of the second layer computes the deviation of the brightness at each location from the local brightness. Then, this deviation is nonlinearly normalized by the local brightness to give the local contrast. (3) The responses to local contrast are convolved by center surround receptive fields (or filtered by the Contrast Sensitivity Function). Then, the linearly filtered contrast is nonlinearly normalized by the local contrast. Again normalization increases the response in the regions with small input (low contrast). (4) After linear wavelet transform, each response is nonlinearly normalized by the activity of the neurons in the surround. Again, the activity relatively increases in the regions with low input. The common effect of the nonlinear modules throughout the network is response equalization (Color figure online)

$$x_i = k_i \frac{e_i}{b_i + \sum_j H_{ij} e_j} \quad (2)$$

where the output vector of nonlinear activations,  $\mathbf{x}$ , depends on the energy of the input linear wavelet responses,  $\mathbf{e}$ , which are dimension-wise normalized by a sum of neighbor energies of the input. For convenience for the derivations below, the transform can be rewritten in matrix form Martinez-Garcia et al. (2018), Martinez-Garcia et al. (2019):

$$\mathbf{x} = \mathbb{D}_{\mathbf{k}} \cdot \mathbb{D}_{(\mathbf{b} + \mathbf{H} \cdot \mathbf{e})}^{-1} \cdot \mathbf{e} \quad (3)$$

where  $\mathbb{D}_{\mathbf{v}}$  are diagonal matrices with the vector  $\mathbf{v}$  in the diagonal. The non-diagonal nature of the interaction kernel  $\mathbf{H}$  which is in the denominator,  $\mathbf{b} + \mathbf{H} \cdot \mathbf{e}$ , implies that the  $i$ -th element of the response is attenuated by the activity of the neighbor sensors,  $e_j$  with  $j \neq i$ . Each row of the kernel  $\mathbf{H}$  describes how the energies of the neighbor simple cells attenuate each simple cell after the interaction. Each element of the vectors  $\mathbf{b}$  and  $\mathbf{k}$  represents the semisaturation and the dynamic range of the nonlinear response of each sensor, respectively. This nonlinear interaction only affects the amplitude of the responses, not its sign. As a result, for simplicity in the notation, throughout the work  $\mathbf{x}$  refers to the vector of absolute values of the responses. The sign of the normalized responses is inherited from the sign of the linear wavelet responses.

*Inverse transform.*

The matrix notation (Martinez-Garcia et al. 2018, 2019) is convenient to derive the analytical inverse of the Divisive Normalization, which will be used to obtain the relation between the two models considered in this work. The inverse is given by Malo et al. (2006), Martinez-Garcia et al. (2018), Martinez-Garcia et al. (2019):

$$\mathbf{e} = \left( \mathbf{I} - \mathbb{D}_{\mathbf{k}}^{-1} \cdot \mathbb{D}_{\mathbf{x}} \cdot \mathbf{H} \right)^{-1} \cdot \mathbb{D}_{\mathbf{b}} \cdot \mathbb{D}_{\mathbf{k}}^{-1} \cdot \mathbf{x} \quad (4)$$

This inverse, originally proposed in the context of image coding (Malo et al. 2006), has been used in other applications that require the reconstruction of the image (Camps et al. 2008; Laparra et al. 2017).

### 2.3 The Wilson–Cowan Model

The Wilson–Cowan dynamical system was proposed as a general model of the inhibitory/excitatory interactions between neural populations, and as application, it can be used to model the signal at specific stages in the visual pathway (Wilson and Cowan 1972, 1973; Bressloff and Cowan 2003). In Wilson–Cowan models, part of the neural population (part of the coefficients in the vectors  $\mathbf{e}$  and  $\mathbf{x}$ ) is *excitatory* and part is *inhibitory*, meaning how their magnitude affects the neighbors (in additive or subtractive way, respectively). Excitatory and inhibitory populations will be referred to as  $\mathbf{e}^e$ ,  $\mathbf{x}^e$ , and  $\mathbf{e}^i$ ,  $\mathbf{x}^i$ , respectively. We will consider that these excitatory and inhibitory neurons (or coefficients) are interleaved in the vectors that describe the responses. Or,

for simplicity, one may also represent them as separate rows in the response vectors:

$$\begin{pmatrix} e^e \\ e^i \end{pmatrix} \xrightarrow{\mathcal{N}} \begin{pmatrix} x^e \\ x^i \end{pmatrix} \quad (5)$$

In any case, the arbitrary arrangement of the neurons in the vectors does not restrict the generality of the formulation. The only effect of this choice is the interpretation of the elements of the matrices that will represent the interaction between the neurons.

*Dynamical system.*

In Wilson–Cowan models (Wilson and Cowan 1972, 1973; Bressloff and Cowan 2003), the transform  $\mathcal{N}$  is defined by differential equations that describe the temporal variation of the activity of the populations. In particular, this variation is driven by three factors:

- An external driving input (either  $e^e$  or  $e^i$ ), in our case the responses of the linear V1 cells.
- The variation of the response of a population is auto-attenuated due to its own activity.
- The variation of the response is amplified by the excitatory responses and is moderated by the inhibitory responses.

Specifically, if in the notation of Bressloff et al. (2002), Bressloff and Cowan (2002, 2003), which considers no refractory period in V1 neurons, we explicitly identify the excitatory and inhibitory populations as done originally in Wilson and Cowan (1973), for a neuron tuned to the feature  $p$ , we have one of these (excitatory or inhibitory) equations:

$$\begin{aligned} \frac{\partial x_p^e(t)}{\partial t} &= e_p^e(t) - \alpha_p^e x_p^e(t) + \int W_{pp'}^{ee} f(x_{p'}^e(t)) dp' - \int W_{pp'}^{ei} f(x_{p'}^i(t)) dp' \\ \frac{\partial x_p^i(t)}{\partial t} &= e_p^i(t) - \alpha_p^i x_p^i(t) + \int W_{pp'}^{ie} f(x_{p'}^e(t)) dp' - \int W_{pp'}^{ii} f(x_{p'}^i(t)) dp' \end{aligned} \quad (6)$$

or, in matrix notation:

$$\begin{aligned} \dot{x}^e &= e^e - \mathbb{D}_{\alpha^e} \cdot x^e + W^{ee} \cdot f(x^e) - W^{ei} \cdot f(x^i) \\ \dot{x}^i &= e^i - \mathbb{D}_{\alpha^i} \cdot x^i + W^{ie} \cdot f(x^e) - W^{ii} \cdot f(x^i) \end{aligned} \quad (7)$$

where  $W^{ee}$ ,  $W^{ei}$ ,  $W^{ie}$ ,  $W^{ii}$  are the matrices that describe the excitatory and inhibitory relations between sensors, the activation function  $f(\cdot)$  is a dimension-wise saturating nonlinearity, and the elements of the vectors  $\alpha^e$  and  $\alpha^i$  are the auto-attenuation parameters. The above matrices are usually considered to be a fixed set of connections (wired relations), made of positive and negative Gaussian neighborhoods, that represent the local interaction between sensors (Bressloff and Cowan 2002, 2003; Chossat and Faugeras 2009). Also, note that if in Eq. 7, the inhibitory and the excitatory components are stacked together into a single vector (with some sort of arrangement as in Eq. 5), the two equations in the traditional Wilson–Cowan formulation can be

represented by a single expression, as in Bressloff and Cowan (2003), Chow and Karimipanh (2020), here in matrix form:

$$\dot{\mathbf{x}} = \mathbf{e} - \mathbb{D}_{\alpha} \cdot \mathbf{x} - \mathbf{W} \cdot f(\mathbf{x}) \quad (8)$$

where

$$\alpha = \begin{pmatrix} \alpha^e \\ \alpha^i \end{pmatrix}, \quad f(\mathbf{x}) = f \begin{pmatrix} x^e \\ x^i \end{pmatrix}, \quad \mathbf{W} = \begin{pmatrix} -W^{ee} & W^{ei} \\ -W^{ie} & W^{ii} \end{pmatrix}$$

The above single-equation matrix formulation of the Wilson–Cowan model, Eq. 8, is convenient to get the relation between the models and clearly shows the subtractive nature of the interactions in the kernel  $\mathbf{W}$  as opposed to the divisive nature of the interactions due to the kernel  $\mathbf{H}$  in Eq. 3.

*Steady state and inverse.*

The stationary solution of the above differential equation (obtained by taking  $\dot{\mathbf{x}} = 0$  in Eq. 8) leads to the following analytical inverse for static inputs:

$$\mathbf{e} = \mathbb{D}_{\alpha} \cdot \mathbf{x} + \mathbf{W} \cdot f(\mathbf{x}) \quad (9)$$

As we will see in the *Analytical Results* section, the identification of the decoding equations corresponding to both models, Eqs. 4 and 9, is the key to obtain simple relations between their corresponding parameters.

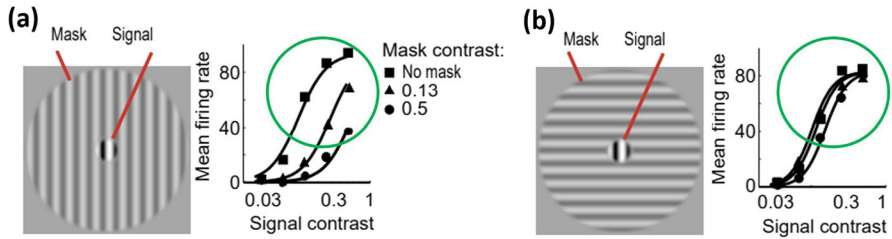
## 2.4 Experimental Facts

### 2.4.1 Adaptive Contrast Response Curves

In the considered spatial vision context, the models should reproduce the fundamental trends of contrast perception. Thus, the slope of the contrast response curves should depend on the spatial frequency, so that the sensitivity at threshold contrast is different for different spatial frequencies according to the Contrast Sensitivity Function (CSF) (Campbell and Robson 1968). Also, the response curves should saturate with contrast (Legge and Foley 1980; Legge 1981). Finally, the responses should attenuate with the energy of the background or surround, and this additional saturation should depend on the texture of the background (Foley 1994; Watson and Solomon 1997): If the frequency/orientation of the test is similar to the frequency/orientation of the background, the decay should be stronger. This background-dependent adaptive saturation, or *masking*, is mediated by cortical sensors tuned to spatial frequency with responses that saturate depending on the background, as illustrated in Fig. 2.

The above trends are key to discard too simple models, and also to propose the appropriate modifications in the model architecture to get reasonable results (Martinez-Garcia et al. 2019).





**Fig. 2** Adaptive contrast response curves. Mean firing rate (response) of V1 neurons tuned to the signal as a function of the signal contrast in two masking situations (adapted from Schwartz and Simoncelli (2001), Cavanaugh (2000)). Note the decay in the response when signal and mask do have the same spatio-frequency characteristics (a), as opposed to the case where they do not (b). For visualization, the differences in the curves are highlighted by the green circles (Color figure online)

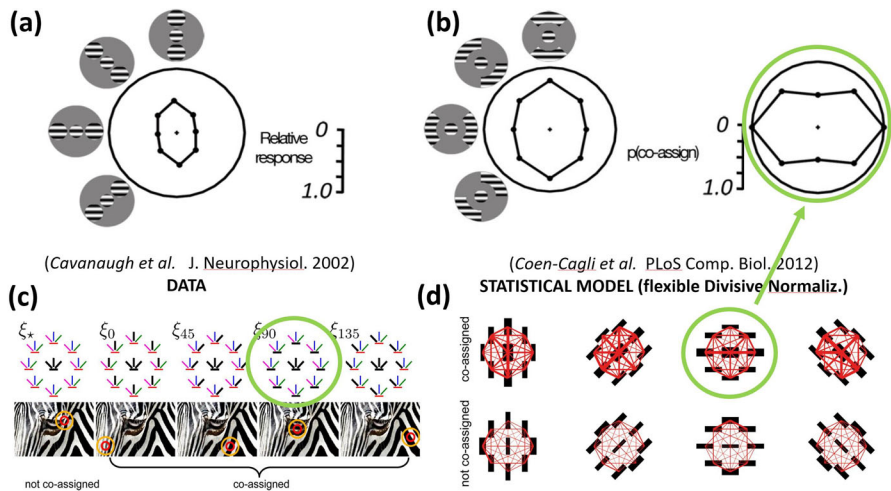
## 2.4.2 Unexplained Kernel Structure in Divisive Normalization

In the Divisive Normalization setting, the masking interaction between tests and backgrounds of different textures is classically described by using a Gaussian kernel in the denominator of Eq. 3 in wavelet-like domains: the effect of the  $j$ -th wavelet sensor on the attenuation of the  $i$ -th wavelet sensor decays with the distance in space between the  $i$ -th and  $j$ -th sensors, but also with the spatial frequency and orientation (Watson and Solomon 1997). We will refer to these unit-norm Gaussian kernels as Watson and Solomon kernels (Watson and Solomon 1997) and will be represented by  $H^{ws}$ . Gaussian kernels are useful to describe the general behavior shown in Fig. 2: Activity in close neighbors leads to strong decays in the response, while activity in neighbors tuned to more distant features has smaller effect.

However, in order to have well-behaved responses in every subband with every possible background, a *special balance* between the wavelet representation and the Gaussian kernels is required. When using reasonable log-polar Gabor basis or steerable filters to model V1 receptive fields, as in Watson and Solomon (1997), Schwartz and Simoncelli (2001), the energies of the sensors tuned to low frequencies are notably higher than the energy of high-frequency sensors. Moreover, the smaller number of sensors in low-frequency subbands in this kind of wavelet representations implies that unit-norm Gaussian kernels have bigger values in coarse subbands. These two facts overemphasize the impact of low-frequency responses on high-frequency responses. Thus, in Martinez-Garcia et al. (2019) we found that classical unit-norm Gaussian kernels require ad hoc extra modulation to avoid excessive effect of low-frequency backgrounds on high-frequency tests. The appropriate wavelet-kernel balance was then reestablished by introducing extra high-pass filters in the Gaussian kernel  $H^{ws}$ , with the aim to moderate the effect of low frequencies (Martinez-Garcia et al. 2019):

$$H = \mathbb{D}_l \cdot H^{ws} \cdot \mathbb{D}_r \quad (10)$$

In this new definition of the kernel: (1) the diagonal matrix at the right,  $\mathbb{D}_r$ , pre-weights the subbands of  $e$  to moderate the effect of low frequencies before computing the interaction; and (2) the diagonal matrix at the left,  $\mathbb{D}_l$ , sets the relative weight of



**Fig. 3** Experimental context-dependent interaction (Cavanaugh et al. 2002b) and statistical model (Coen-Cagli et al. 2012). **a** Results of Cavanaugh et al. (2002b): images with a gray background represent the stimuli. Cell relative responses are shown as points inside the black normalization circle. The distance from the origin indicates the magnitude of the response, while its angle represents the location of the surrounding stimulus. **b** Cell response predicted from the statistical model of Coen-Cagli et al. (2012), and probability that center and surround receptive fields are *co-assigned* to the same normalization pool and contribute to the Divisive Normalization of the model response. The probability of co-assignment depends on the covariance with the surround, as shown below. **c** Different center-surround visual neighborhoods in a natural scene. In each case, the activity of the sensors in the surround can be co-assigned to the activity in the center (i.e., considered in the normalization pool) if the orientation of maximally responding sensors is consistent (which is the case for four of the considered regions and not the case for the first). The horizontal surround that is *co-assigned* with the corresponding center is highlighted in bold. **d** Covariance matrices learned from natural images determine co-assignment: The orientation and relative position of the receptive fields are represented by the black bars. (The thickness of the bar is proportional to the variance, while the thickness of the red lines is proportional to the covariance between the two connected bars.) This figure has been adapted from Coen-Cagli et al. (2012) (Color figure online)

the masking for each sensor, moderating low frequencies again. The vectors  $\mathbf{r}$  and  $\mathbf{l}$  were tuned ad hoc in Martinez-Garcia et al. (2019) to get reasonable contrast response curves, both for low- and high-frequency tests.

However, *what is the explanation for this specific structure of the kernel matrix in Eq. 10? And where do these two high-pass diagonal matrices come from?*

### 2.4.3 Adaptive Nature of Kernel in Divisive Normalization

Previous physiological experiments on cats and macaques demonstrated that the effect of the surround on each cell does not come equally from all peripheral regions, showing up the existence of a spatially asymmetric surround (Nelson and Frost 1985; Deangelis et al. 1994; Walker et al. 1999; Cavanaugh et al. 2002a,b). As shown in Fig. 3a, the experimental cell response is suppressed due to the surround, and this attenuation is greater when the grating patches are *iso-oriented* and at the ends of the receptive field (as defined by the axis of preferred orientation) (Cavanaugh et al. 2002b).

In the Divisive Normalization context, this asymmetry could be explained with non-isotropic interaction kernels. Depending on the texture of the surround, the interaction strength in certain direction may change. This would change the denominator, and hence the gain in the response.

Coen-Cagli et al. (2012) proposed a specific statistical model to account for these contextual dependencies. This model includes grouping and segmentation of neighboring oriented features and leads to a flexible generalization of the Divisive Normalization. Representative center-surround configurations considered in the statistical model are shown in Fig. 3c. A surround orientation can be either *co-assigned* with the center group or *not co-assigned*. In the first case, the model assumes dependence between center and surround and includes them both in the normalization pool for the center. In the second case, the model assumes center-surround independence and does not include the surround in the normalization pool. Figure 3d shows the covariance matrices learned from natural images between the variables associated with center and surround in the proposed statistical model. As expected, the variances of the center and its *co-linear* neighbors, and also the covariance between them, are larger, due to the predominance of *co-linear* structures in natural images. The cell response that is computationally obtained assuming their statistical model is shown in Fig. 3b, together with the probability that center and surround receptive fields are *co-assigned* to the same normalization pool and contribute then to the divisive normalization of the model response. Note that the higher the probability of *co-assignment* between the center and surround, the higher the suppression (or the lower the signal) in the cell response.

This flexible (or adaptive) Divisive Normalization model based on image statistics (Coen-Cagli et al. 2012) allows to explain the experimental asymmetry in the center-surround modulation (Cavanaugh et al. 2002b). However, no direct mechanistic approach has been proposed yet to describe how this adaptation in the Divisive Normalization kernel may be implemented.

### 3 Analytical Results: Relation Between Models

The kernels that describe the relation between sensors in the Divisive Normalization and the Wilson–Cowan models,  $\mathbf{H}$  and  $\mathbf{W}$ , have similar qualitative roles: both moderate the response, either by division or subtraction, taking into account the activity of the neighbor sensors.

In this section, we derive the relation between both models assuming that the Divisive Normalization behavior corresponds to the steady-state solution of the Wilson–Cowan dynamics. This leads to an interesting analytical relation between both kernels,  $\mathbf{H}$  and  $\mathbf{W}$ .

Under the steady-state assumption, it is possible to identify the different terms in the decoding equations in both cases (Eqs. 4 and 9). However, just to get a simpler analytical relation between both kernels, we make one extra simplification on each model. Numerical experiments in the next section with natural inputs and a wide range of model parameterizations show that these simplifications are acceptable in practice.

First, in the Divisive Normalization model, Eq. 4, the identification may be simpler by taking the series expansion of the inverse. This expansion was used in Malo et al. (2006) because it clarifies the condition for invertibility:

$$\left(I - \mathbb{D}_k^{-1} \cdot \mathbb{D}_x \cdot \mathbf{H}\right)^{-1} = I + \sum_{n=1}^{\infty} \left(\mathbb{D}_k^{-1} \cdot \mathbb{D}_x \cdot \mathbf{H}\right)^n$$

The inverse exist if the eigenvalues of  $\mathbb{D}_k^{-1} \cdot \mathbb{D}_x \cdot \mathbf{H}$  are smaller than one so that the series converges. In fact, if the eigenvalues are small, the inverse can be well approximated by a small number of terms in the Maclaurin series. Taking into account this approximation, Eq. 4 may be written as:

$$\begin{aligned} \mathbf{e} &= \mathbb{D}_b \cdot \mathbb{D}_k^{-1} \cdot \mathbf{x} + \left(\mathbb{D}_k^{-1} \cdot \mathbb{D}_x \cdot \mathbf{H}\right) \cdot \mathbb{D}_b \cdot \mathbb{D}_k^{-1} \cdot \mathbf{x} + \\ &\quad + \left(\mathbb{D}_k^{-1} \cdot \mathbb{D}_x \cdot \mathbf{H}\right)^2 \cdot \mathbb{D}_b \cdot \mathbb{D}_k^{-1} \cdot \mathbf{x} + \\ &\quad + \left(\mathbb{D}_k^{-1} \cdot \mathbb{D}_x \cdot \mathbf{H}\right)^3 \cdot \mathbb{D}_b \cdot \mathbb{D}_k^{-1} \cdot \mathbf{x} + \dots \\ \mathbf{e} &\approx \left(\mathbb{D}_b \cdot \mathbb{D}_k^{-1} + \mathbb{D}_k^{-1} \cdot \mathbb{D}_x \cdot \mathbf{H} \cdot \mathbb{D}_b \cdot \mathbb{D}_k^{-1}\right) \cdot \mathbf{x} \end{aligned} \quad (11)$$

Second, in the case of the Wilson–Cowan model (Eq. 9) we approximate the saturation function  $f(\mathbf{x})$  so that we can isolate the vector  $\mathbf{x}$ . This can be done by expressing  $f(\mathbf{x})$  through an Euler integration of  $n$  terms:  $f(\mathbf{x}) = f(\mathbf{0}) + \int_0^{\mathbf{x}} \frac{df}{dx}(\mathbf{x}') d\mathbf{x}' \approx f(\mathbf{0}) + \sum_{\beta=0}^{n-1} \frac{df}{dx}(\frac{\beta}{n}\mathbf{x}) \frac{\mathbf{x}}{n}$ . Note that along the integration the derivatives are computed at different points from  $\mathbf{0}$  up to  $\frac{n-1}{n}\mathbf{x}$ . If  $n = 1$ , we have the (in principle poor) Maclaurin approximation, and if  $n \rightarrow \infty$ , the result is perfect. In between, for finite  $n$ , we have an approximation with certain accuracy. In this case, taking into account that in the activation functions  $f(\mathbf{0}) = \mathbf{0}$ , and calling  $g_n(\mathbf{x}) = \frac{1}{n} \sum_{\beta=0}^{n-1} \frac{df}{dx}(\frac{\beta}{n}\mathbf{x})$ , we can write:

$$\mathbf{e} \approx (\mathbb{D}_\alpha + \mathbf{W} \cdot \mathbb{D}_{g_n(\mathbf{x})}) \cdot \mathbf{x} \quad (12)$$

Now, the identification between the approximated versions of the decoding equations, Eqs. 11 and 12, is straightforward. As a result, we get the following relations between the parameters of both models:

$$\begin{aligned} \alpha &= \frac{b}{k} \\ \mathbf{W} &= \mathbb{D}_{(\frac{x}{k})} \cdot \mathbf{H} \cdot \mathbb{D}_{\left(\frac{k}{b} \odot g_n(\mathbf{x})\right)}^{-1} \end{aligned} \quad (13)$$

where the symbol  $\odot$  denotes the element-wise (or Hadamard) product, and the ratios between vectors are also Hadamard divisions.

Note that the Divisive Normalization kernel which is compatible with Eq. 13,  $\mathbf{H} = \mathbb{D}_{\left(\frac{k}{x}\right)} \cdot \mathbf{W} \cdot \mathbb{D}_{\left(\frac{k}{b} \odot g_n(\mathbf{x})\right)}$ , has exactly the same structure as the one in Eq. 10. Therefore,

both models agree if the Divisive Normalization kernel *inherits* the structure from the Wilson–Cowan kernel left- and right-multiplied by these diagonal matrices,  $\mathbb{D}\left(\frac{k}{x}\right)$  and  $\mathbb{D}\left(\frac{k}{b \odot g_n(x)}\right)$ , respectively.

This theoretical result suggests an explanation for the structure that had to be introduced ad hoc in Martinez-Garcia et al. (2019) just to reproduce contrast masking. Note that the interaction in the Wilson–Cowan case may be understood as wiring between sensors tuned to similar features, so a unit-norm Gaussian,  $\mathbf{W} = \mathbf{H}^{ws}$ , is a reasonable choice (Wilson and Cowan 1973; Chossat and Faugeras 2009). Note also that the weights before and after  $\mathbf{W}$  (the diagonal matrices) are signal dependent. Therefore, a fixed wiring  $\mathbf{W}$  implies that the kernel in Divisive Normalization should be *adaptive*. The one in the left,  $\mathbb{D}\left(\frac{k}{x}\right)$ , has a direct dependence on the inverse of the signal, while the one in the right,  $\mathbb{D}\left(\frac{k}{b \odot g_n(x)}\right)$ , depends on the derivatives of the activation  $f(x)$ . Next section shows that these vectors  $\frac{k}{x}$  and  $\frac{k}{b} \odot g_n(x)$  do have the high-pass frequency nature that explains why the low frequencies in  $\mathbf{e}$  had to be attenuated ad hoc by introducing  $\mathbb{D}_l$  and  $\mathbb{D}_r$ . We also show that the term of the right,  $\mathbb{D}\left(\frac{k}{b \odot g_n(x)}\right)$ , produces the shape changes needed on the interactions.

It is important to stress that the simplifications made in the decoding equations to get the analytical relations in Eq. 13 were done only for the sake of simplicity in the final relations obtained. In summary, the expressions in Eq. 13 are exact for the simplified versions of the models. Considering the full version of the models, Eq. 13 would be an approximation. However, the experiments below point out the validity of this approximation, because: (a) we explicitly check that the errors are small in a range of scenarios, and (b) we check that plugging these expressions into the full versions of the models also leads to consistent results.

## 4 Numerical Experiments

The analysis of the proposed relation between the Divisive Normalization (DN) and the Wilson–Cowan (WC) models is a three stage process. *First*, one should take biologically plausible parameters (either in DN, in WC, or in both) and then use the proposed expressions to build versions of the models supposed to behave similarly. *Second*, one should check if the models obtained in that way actually behave similarly. So that finally, *third*, one can elaborate on the consequences of this correspondence.

In this experimental analysis, in Sect. 4.1, we build a psychophysically inspired Wilson–Cowan model for V1 from a Divisive Normalization with psychophysically tuned parameters (Malo and Simoncelli 2015; Martinez-Garcia et al. 2018, 2019). This model also preserves the basic properties of the interaction kernel and the saturation function of the Wilson–Cowan literature (Wilson and Cowan 1973; Bressloff and Cowan 2003; Chossat and Faugeras 2009). This Wilson–Cowan model should behave similarly to the corresponding Divisive Normalization model.

Then, Sect. 4.2 experimentally checks the mathematical relation between the models. In particular, for a wide range of parameters: (a) we show that the integration of

the Wilson–Cowan equation certainly converges to a solution which is close to its corresponding Divisive Normalization; (b) we quantify the accuracy of the approximations required to get the relation between the models; and (c) we show that the Divisive Normalization solution is a stable node of the dynamical system governed by the Wilson–Cowan equations.

Finally, in Sect. 4.3, we address different *consequences on contrast perception* using the proposed relation: (a) We analyze the signal-dependent behavior of the theoretically derived kernel and the benefits of the high-pass behavior to moderate the weight of the low-frequency components; (b) we show that the shape of the interactions between sensors changes depending on the surround; (c) we reproduce the contrast response curves with the proposed signal-dependent kernel; and (d) we discuss the use of the derived kernel in predicting the subjective metric of the image space.

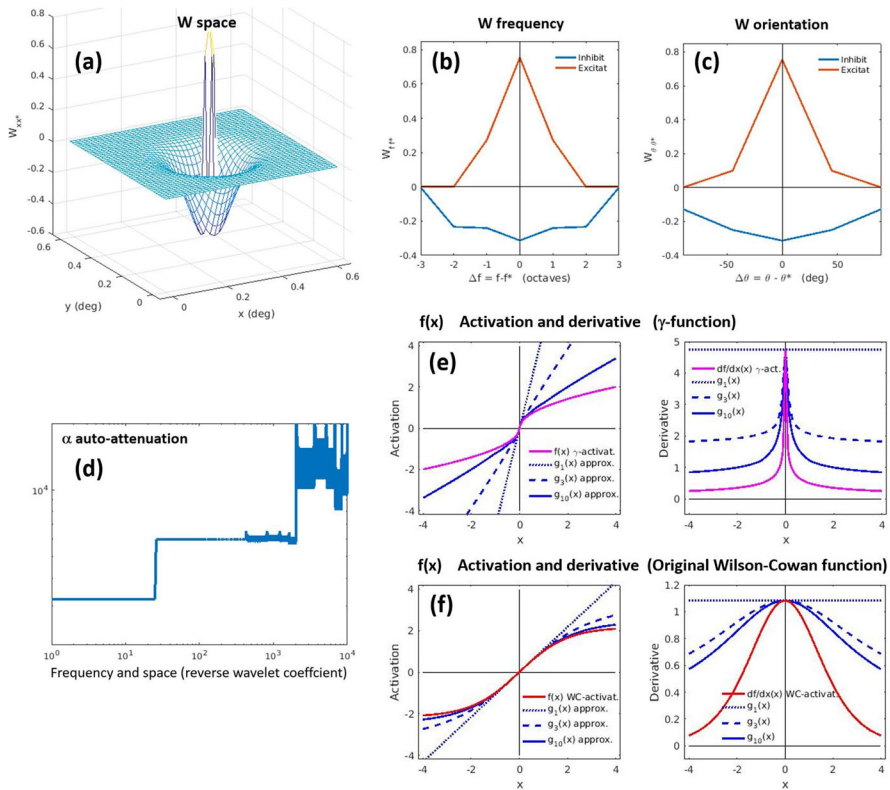
#### 4.1 Psychophysically Plausible Parameters for a Wilson–Cowan Model in V1

A possible way to check the relation between the models in V1 consists of starting from the (lower-level/mechanistic/physiological) Wilson–Cowan model and let it evolve to see if it converges to the (psychophysical) Divisive Normalization response. To this end, for our Wilson–Cowan model, we need reasonable  $\alpha$ ,  $\mathbf{W}$ , and  $f(x)$ , for  $e$  and  $x$  defined in certain wavelet representation.

For the wavelet representation here, we assume 4-orientation steerable transforms (Simoncelli et al. 1992) as a convenient model of the simple cells [as done in Schwartz and Simoncelli (2001), Martinez-Garcia et al. (2018, 2019)]. In the experiments involving the (computationally intensive) integration of the Wilson–Cowan differential equation, Sect. 4.2, we used wavelets with 3 scales in  $40 \times 40$  images to speed up the computation. But in the psychophysical illustrations, Sect. 4.3, we used 4 scales in  $64 \times 64$  images.

The reference parameters for the nonlinearity are taken from the Divisive Normalization model in Martinez-Garcia et al. (2019). In that case, the parameters corresponding to contrast computation, contrast sensitivity, and masking in the spatial domain were directly measured using *Maximum Differentiation* psychophysics (Malo and Simoncelli 2015), while the parameters related to brightness and to masking in the wavelet domain were tuned to reproduce subjective image quality data (Martinez-Garcia et al. 2018) and contrast perception curves (Martinez-Garcia et al. 2019).

As stated after Eq. 13, we took  $\mathbf{W}$  as a Watson–Solomon separable Gaussian kernel (Watson and Solomon 1997) with widths in space/frequency/orientation taken from the psychophysically plausible values in Martinez-Garcia et al. (2019). In order to include both excitatory and inhibitory populations, we complemented this initial kernel with narrow excitatory neighborhoods whose width was a fraction of the original inhibitory neighborhoods. Finally, we normalized the absolute amplitude of the neighborhoods to have unit-norm center-surround interactions. Figure 4a–c illustrates the psychophysically sensible separable kernels  $\mathbf{W}$ . These unit-norm kernels  $\mathbf{W}$  scaled as in Watson and Solomon (1997), Martinez-Garcia et al. (2019) are consistent with the shapes used in the Wilson–Cowan literature (Bressloff and Cowan 2003; Chossat and Faugeras 2009).



**Fig. 4** Psychophysically inspired parameters for Wilson–Cowan model. Illustration of the Gaussian neighborhoods  $W$  with excitatory and inhibitory parts: separable interactions depending on departure in space (a), frequency (b), and orientation (c). The spatial component of  $W$  is shown for the central location of the considered visual field and the 24 cpd vertical subband. Following Watson and Solomon (Watson and Solomon 1997), the Gaussian kernels (here difference of Gaussians) are separable in space, frequency, and orientation. Therefore, lower-frequency subbands have coarser sampling (and thus higher amplitude) but the same shape in space. The shape is also the same for subbands tuned to different orientations. Equivalent separability applies for variations in frequency and orientation: The interactions in the plots (b) and (c) are the same for every spatial location (and orientation and frequency, respectively). The above plots display the (more intuitive)  $-W$ , where positive and negative values mean excitatory and negative interaction, respectively, but note that, according to the sign in Eq. 8, positive weights are inhibitory. (d) Auto-attenuation factor  $\alpha$ . In this plot, the horizontal axis (wavelet coefficients in reverse order) can be qualitatively understood as frequency so high frequencies display bigger attenuation. The panels (e) and (f) show different options for the pointwise activation nonlinearity,  $f(x)$ . In pink, we have the  $\gamma$ -function proposed in Martinez-Garcia et al. (2018), and in red, we have the original proposal of Wilson–Cowan (Wilson and Cowan 1973). In both cases, we show the approximations of these functions using  $g_n(x)$  (for increasing number of terms  $n$ ), and the corresponding derivatives, which have an impact in the relation between the kernels in the different models (Eq. 13) (Color figure online)



Regarding the auto-attenuation we simply took the constants  $k$  and  $b$  from Martinez-Garcia et al. (2019) and used the first equation of the proposed relation, Eq. 13, to obtain  $\alpha$ . Figure 4d shows the  $\alpha$  vector for the 3-scale wavelet (coefficients ordered from low to high frequency). Note that the response of sensors tuned to higher frequencies is more attenuated in the evolution of the differential equation, while low frequencies have lower auto-attenuation.

Finally, Fig. 4e, f, displays different activations  $f(x)$  that we used in the experiments together with representative Euler approximations,  $g_n(x)x$ , and the functions related to their derivatives,  $\frac{df}{dx}(x)$  and  $g_n(x)$ . These activation functions include the *original*-activation in Wilson and Cowan (1973), Bressloff and Cowan (2003), and the so-called  $\gamma$ -activation inspired in retinal transduction (Martinez-Garcia et al. 2018). Appendix A gives the expressions of these activation functions. In our wavelet case, the horizontal and vertical axes of the function  $f(\cdot)$  to be applied to each coefficient  $x$  of certain subband are scaled by the average amplitude of the responses of the corresponding linear sensors to natural images. With that scaling, the nonlinearities preserve the relative scales of the input subbands in the vector  $e$  that comes from the linear filters.

In the next experiments, the above psychophysically sensible parameters (the reference values) are modified in several ways to show that the proposed relation works for a wide range of model parameterizations. Specifically, (a) we explored different widths of the interaction kernels by using five scaling factors applied to the reference widths: from unrealistically narrow (zero-width, identity kernels that disregard interactions), to unrealistically wide kernels (where the reference widths are increased by an order of magnitude); (b) we considered kernels with the above mentioned excitatory-inhibitory nature, and kernels with just-inhibitory nature; and (c) we considered two possible activations (the original-activation and the  $\gamma$ -activation). We considered a total of 12 parameterizations of the models: 5 kernel widths  $\times$  2 excit-inhib configurations  $\times$  1 activation (original-activ.) + 1 kernel width (the reference one)  $\times$  2 excit-inhib configurations  $\times$  1 activation (the  $\gamma$ -activ.).

The interested reader has access to the specific values of the parameters in Appendix A for the activations, and in the code that reproduces all the simulations of the paper (described in Appendix 1).

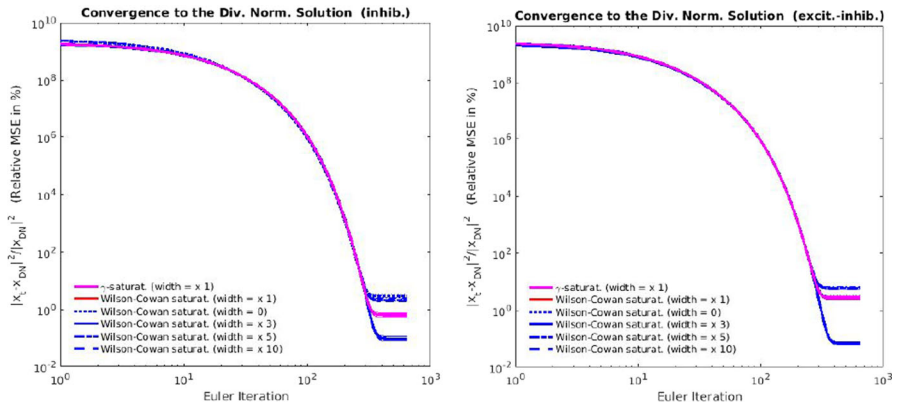
## 4.2 Experimental Check of Mathematical Properties

### 4.2.1 Wilson–Cowan Systems Converge to the Divisive Normalization

The Wilson–Cowan expression, Eq. 8, defines an initial value problem where the response at time zero evolves (or is updated) according to the right-hand side of the differential equation. In our case, we assume that the initial value of the output is just the input  $x(0) = e$ . Moreover, as we deal with static images, we assume that the input is constant. And then, we solve this first-order differential equation by the simplest (Euler) integration method:

$$x(t + \Delta t) = x(t) + \left( e - \mathbb{D}_\alpha \cdot x(t) - W \cdot f(x(t)) \right) \Delta t \quad (14)$$

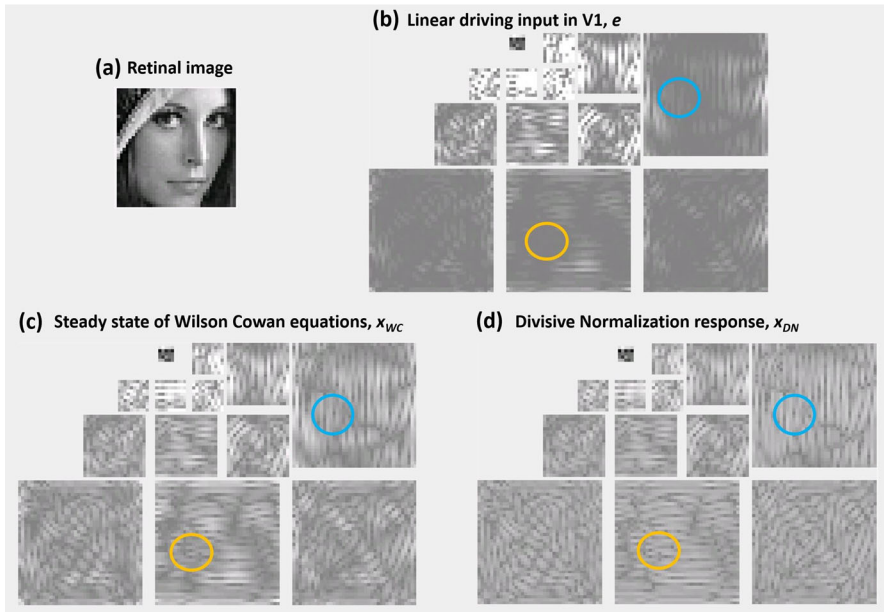




**Fig. 5** Convergence to the Divisive Normalization solution (I): quantitative error. Percentage of relative MSE between the solution of the Wilson–Cowan equations and the Divisive Normalization response as a function of discrete time (steps in Euler integration). *Left*: relative MSE for different kernel widths and activation/saturation functions in case of just inhibitory interactions. *Right*: relative MSE for different kernel widths and saturation functions in case of excitatory+inhibitory interactions. The configurations in pink and red are the psychophysically meaningful ones, which achieve 0.6% and 2.7% relative MSE, respectively. In all cases, the update of the solution tends to zero (results not shown) indicating the approach to a steady state. The curves in bold style represent the median (50 quantile) over 45 natural images, and the curves in light style (very close to the median) represent the 25 and 75 quantiles. This result suggests the successful convergence of WC to DN for natural images over a wide range of model parameterizations (Color figure online)

Figure 5 shows the evolution of the response obtained from this integration, applied to 45 natural images taken from calibrated databases (Hateren and Schaaf 1998; Laparra et al. 2012), using the biologically sensible parameters  $\alpha$ ,  $W$  and  $f(\cdot)$  presented in Fig. 4 (Malo and Simoncelli 2015; Martinez-Garcia et al. 2018, 2019; Wilson and Cowan 1973; Bressloff and Cowan 2003; Chossat and Faugeras 2009), and the mentioned variations to cover a wide range of model parameterizations. Our Euler integration used a small enough discrete time step,  $\Delta t = 10^{-5}$ , and the initial responses, the vectors  $\mathbf{e}$ , were computed using the first 3 layers of the model in Fig. 1 (Martinez-Garcia et al. 2018, 2019) followed by a linear steerable wavelet of 4 orientations and 3 scales. The integration requires no approximation of the WC model, and the evolving solution is checked against the corresponding DN response that uses the proposed Eq. 13.

As can be seen, the solution of the Wilson–Cowan integration converges to the Divisive Normalization solution because their difference (percentage of relative mean squared error) decreases as it evolves in *all* the 12 considered parameterizations. The relative MSE in the psychophysically meaningful situations ( $\times 1$  width) is below 3% (lines in pink and red), and for the other configurations, it is always below 6%. Therefore, the Divisive Normalization always explains more than 94% of the energy of the Wilson–Cowan solution. Moreover, these results represent the steady states because the updates of the solutions in the integrals always tend to zero (results not shown).



**Fig. 6** Convergence to the Divisive Normalization solution (II): qualitative similarity. **a** Input image (stimulus at the retina), **b** Responses of the linear simple cells of V1. Responses are spatially non-stationary, see regions of very low amplitude highlighted in yellow and blue. **c** Steady state of the Wilson–Cowan equation after 650 iterations of Eq. 14 with inhibitory kernel  $W$  of psychophysically sensible width and  $\gamma$ -activation, **d** Corresponding Divisive Normalization response using the kernel given by Eq. 10. In this example, the relative MSE between  $x_{WC}$  and  $x_{DN}$  is 0.93%, but, more importantly, note how the amplitude of the response of the highlighted neurons has increased similarly in the nonlinear cases leading to a more stationary response (Color figure online)

Figure 6 illustrates the qualitative similarity of the responses of the two models and their comparable equalization effect in the wavelet domain. The nonlinear response  $x_{WC}$  was computed by integrating Eq. 14, and  $x_{DN}$  was computed with Eq. 3. We used the parameters introduced in Sect. 4.1 and the corresponding parameters for Divisive Normalization using Eq. 13.

Note how the nonlinearities substantially increase the amplitude of the signal in the regions where the linear response is low. The regions highlighted in blue and orange in  $e$  display low activity compared to their neighbors because there are no edges in those regions of the image. However, the corresponding neurons after Wilson–Cowan or Normalization have increased their activity. The amplitude of the signal after the nonlinearities is more stationary across the subbands. Moreover, the nonlinearities lead to responses where the image structure is less apparent: The activity of a neuron is more independent from the activity of the neighbors. Equalization and increased independence qualitatively suggested in Fig. 6 are consistent with previous (quantitative) studies that report redundancy reduction both in Divisive Normalization (Schwartz and Simoncelli 2001; Malo and Laparra 2010; Malo 2020), and in the Wilson–Cowan model (Gomez-Villa et al. 2020).

#### 4.2.2 Quantification of the Accuracy of the Approximations

The proposed relation, Eq. 13, is based on two approximations:

- The approximation of the *inverse* of Divisive Normalization to obtain Eq. 11, namely:  $(I - \mathbb{D}_k^{-1} \cdot \mathbb{D}_x \cdot \mathbf{H})^{-1} \approx I + \mathbb{D}_k^{-1} \cdot \mathbb{D}_x \cdot \mathbf{H}$ .
- The approximation of  $f(\mathbf{x})$  in Wilson–Cowan to obtain Eq. 12, namely:  $f(\mathbf{x}) \approx g_n(\mathbf{x}) \mathbf{x} = \left( \frac{1}{n} \sum_{\beta=0}^{n-1} \frac{df}{dx}(\beta \frac{\mathbf{x}}{n}) \right) \mathbf{x}$ .

The accuracy of such approximations depends on the model parameters, e.g., the shape and magnitude of  $\mathbf{H}$  or  $f(\cdot)$ , and on the responses  $\mathbf{x}$  to natural images. The low amplitude of the coefficients of natural images in wavelet representations (Olshausen and Field 1996; Malo et al. 2000) and the accuracy of similar approximations for psychophysically sensible parameters (Malo et al. 2006) suggest that errors will be small. However, in this section we explicitly compute both sides of the above expressions (with and without approximation) for a range of representative images and model parameterizations, and we compute the difference between both sides. This difference is the error due to the approximation. We express the energy of the difference (the mean squared error) in terms of percentage of the energy of the function with no approximation: the relative MSE in %.

In Table 1, we show the relative MSE (in %) for both approximations (*inverse* and  $f(\mathbf{x})$ ) together with the error in *convergence*, also in relative MSE, for 12 different model parameterizations. The approximations of  $f(\mathbf{x})$  were done using  $g_{10}(\mathbf{x})$ , i.e., computing derivatives in 10 points.

The approximations generally explain more than 90% of the energy of the original magnitudes. The only exception is the approximation of the inverse of the Divisive Normalization for the (unrealistic) zero-width kernel, where the relative MSE amounts to  $\sim 30\%$ . The deviation in this unrealistic case makes sense because reducing the width of unit-volume kernels increases their height and hence the magnitude of  $\mathbf{H}$  increases so the term summed to the identity in the expression under consideration is not as small. This leads to an increased error in the approximation.

Interestingly, in the whole range of parameterizations considered, the approximations do not have a big impact in the convergence error, which is the actual measure of correspondence between the two models.

#### 4.2.3 Stability Analysis of the Divisive Normalization Response

The stability of a dynamical system at the steady state is determined by the Jacobian with regard to perturbations in the response: If the eigenvalues of this Jacobian are all negative for this response, it is a stable node of the system (Logan 2015). In that situation, the evolution of the perturbations is a vector field oriented toward the stable node.

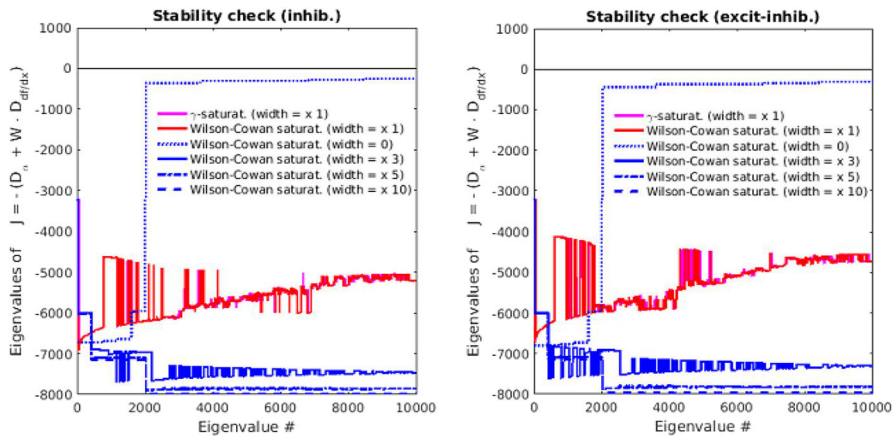
In our case, the Jacobian with regard to the output signal of the right-hand side of the Wilson–Cowan differential equation, Eq. 8, is:

$$J = -(\mathbb{D}_\alpha + \mathbf{W} \cdot \mathbb{D}_{\frac{df}{dx}}(\mathbf{x})) \quad (15)$$

**Table 1** Accuracy of the approximations and convergence error.

Inhibitory W		Excitatory-Inhibitory W		$f(\mathbf{x})$	Convergence
Width $\times$	Inverse	Inverse	Inverse		
$\gamma$ -activation					
<b>Width <math>\times</math></b>	<b>0.019 <math>\pm</math> 0.009</b>		<b>10 <math>\pm</math> 6</b>	<b>0.0005 <math>\pm</math> 0.0001</b>	<b>2.6 <math>\pm</math> 0.3</b>
<i>Original-activation</i>					
$\times 0$	32.74 $\pm$ 0.04		31.04 $\pm$ 0.05	0.06 $\pm$ 0.02	2.9 $\pm$ 0.3
$\times 1$	0.019 $\pm$ 0.009		10 $\pm$ 6	0.04 $\pm$ 0.03	<b>2.7 <math>\pm</math> 0.4</b>
$\times 3$	0.00022 $\pm$ 0.00004		0.00014 $\pm$ 0.00002	0.06 $\pm$ 0.04	0.067 $\pm$ 0.007
$\times 5$	6.0 $\times 10^{-8} \pm 0.3 \times 10^{-8}$		1.1 $\times 10^{-7} \pm 0.6 \times 10^{-7}$	0.05 $\pm$ 0.01	6.0 $\pm$ 0.5
$\times 10$	1.6 $\times 10^{-10} \pm 0.1 \times 10^{-10}$		1.5 $\times 10^{-8} \pm 0.1 \times 10^{-8}$	0.10 $\pm$ 0.03	5.8 $\pm$ 0.5

Relative Mean Squared Error (rel. MSE in %) of three magnitudes: (a) the error in the approximation of the *inverse* of Divisive Normalization, (b) the error in the approximation of  $f(\mathbf{x})$  in Wilson-Cowan, and (c) the final error in the *convergence* shown in Fig. 5. The convergence errors for the psychophysically sensible configurations are highlighted in bold style and color. Values represent the median relative MSE in % over 45 images, and uncertainty intervals represent the 50% interquartile distance



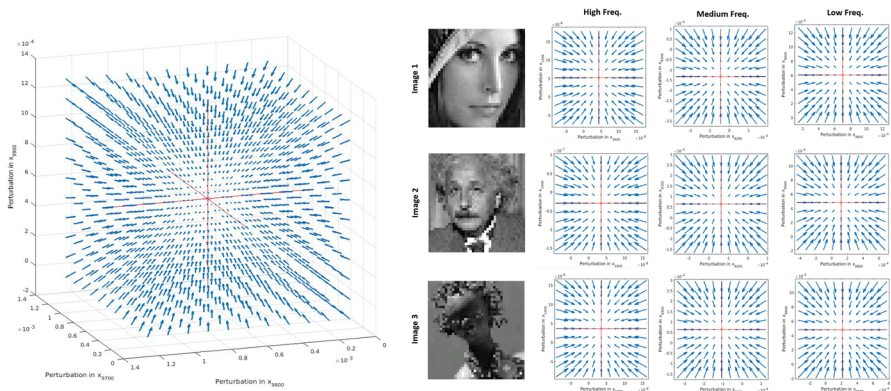
**Fig. 7** Stability of the Divisive Normalization solution (I). Eigenspectrum of the Jacobian of the right-hand side of the Wilson–Cowan differential equation with psychophysically tuned parameters on natural images (curves in pink and red), and different additional configurations (different widths and activation functions), including just inhibitory interactions (Left) and excitatory–inhibitory interaction (right). The curves refer to the median of the eigenvalues over 45 representative images extracted from the calibrated datasets (Hateren and Schaaf 1998; Laparra et al. 2012). The standard deviation and quartile distance are so small that cannot be seen in the plot. The result shows that the eigenvalues are *all negative*. This suggests that the Divisive Normalization is a stable node of the system for natural images for a wide range of model parameterizations (Color figure online)

Figure 7 shows the eigenvalues of this Jacobian using a wide range of parameters (the 12 configurations obtained through variations of the reference values presented in Sect. 4.1), with responses from a set of 45 representative natural images from colorimetrically calibrated datasets (Hateren and Schaaf 1998; Laparra et al. 2012). This result shows that *all* the eigenvalues are negative, thus suggesting that the Divisive Normalization solution is a stable node of the dynamical system.

The stability of the system can be further illustrated by the visualization of the vector field of perturbations in the phase space of the system (Logan 2015). In this case, we visualize this vector field for the Divisive Normalization solution. As the signals in our problem live in very high-dimensional spaces (the wavelet vectors in this section have dimension 10,025), it is not possible to visualize the complete phase space, so we just select some illustrative 3-dimensional and 2-dimensional examples.

Figure 8 (left) shows an example taking just 3 neurons of the V1 layer. In this case, we took a particular image (the standard image *Lena*) and we focused on the response of 3 specific sensors of the low-frequency scale of the Divisive Normalization vector: the 9700, 9800, and 9900-th responses. In that way, we get the red circle in Fig. 8 (left). Arbitrary perturbations of the responses of these neurons lead to the dynamics shown in the phase space: The vector field induced by the Jacobian implies that any perturbation is sent back to the original (no-perturbation) response, which is, then, a stable node of the system.

Similar behavior is obtained for coefficients of other subbands or other images. See Fig. 8 (right), where, for simplicity, we consider perturbations in pairs of neurons.



**Fig. 8** Stability of the Divisive Normalization solution (II). Vector fields in the phase space generated by the Jacobian of the psychophysically tuned Wilson–Cowan model. The example at the left (red dot) corresponds to the response of 3 low-frequency sensors at the Divisive Normalization solution. This vector field describes the evolution of the response if it is perturbed in arbitrary directions (the cardinal directions, in red, or any combination of them). The result is general: Examples at the right show similar results for pairs of sensors tuned to different frequencies for different input stimuli (Color figure online)

In summary, the Divisive Normalization solutions are stable nodes of the corresponding Wilson–Cowan systems. This conclusion confirms the assumption under the proposed relation: Divisive Normalization as a steady state of the Wilson–Cowan dynamics.

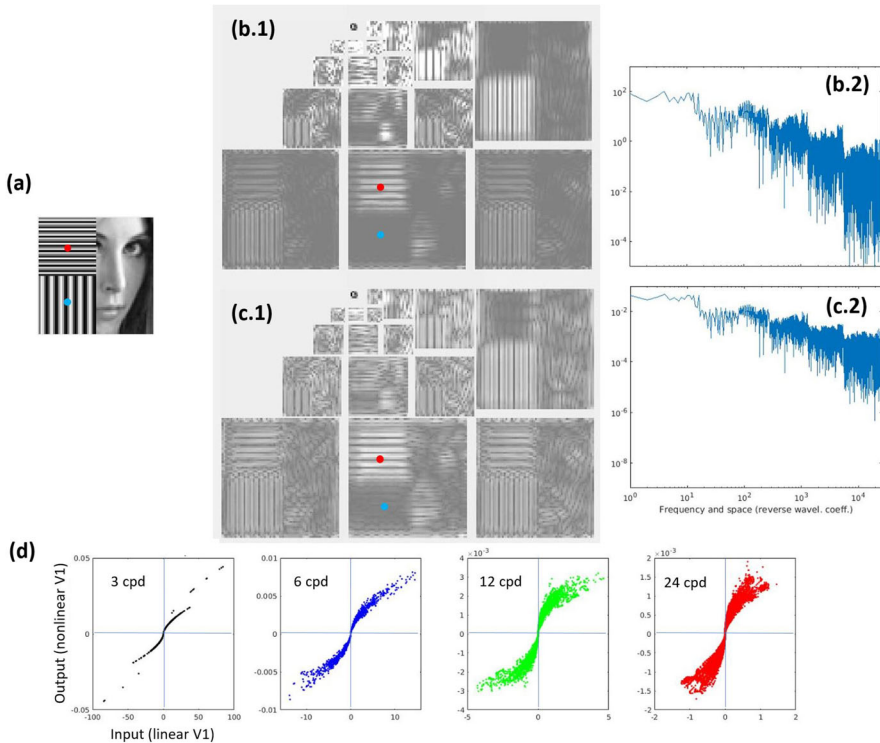
### 4.3 Consequences on Contrast Perception

The proposed relation implies that the Divisive Normalization kernel *inherits* the structure of the Wilson–Cowan interaction matrix (typically Gaussian (Wilson and Cowan 1973; Chossat and Faugeras 2009)), modified by some specific signal dependent diagonal matrices, as seen after Eq. 13, and allows to explain a range of contrast perception phenomena.

First, regarding the structure of the kernel, we show that our prediction is consistent with previously required modifications of the Gaussian kernel in Divisive Normalization to reproduce contrast perception (Martinez-Garcia et al. 2019). Second, we show that the kernel in Divisive Normalization modifies its shape depending on the signal, thus explaining the behavior previously reported in Cavanaugh et al. (2002b), Coen-Cagli et al. (2012). Third, we use the predicted signal-dependent kernel to simulate contrast response curves consistent with Foley (1994), Watson and Solomon (1997). And finally, the proposed relation is also applied to reproduce the experimental visibility of spatial patterns in more general contexts as subjective image quality assessment (Ghadiyaram and Bovik 2016; Ponomarenko et al. 2008, 2009).

In this section, we do not integrate the Wilson–Cowan differential equation, but we use the expression for the steady-state solution with the kernel obtained from the proposed relation. This alleviates computation so, as opposed to the previous Section, in the following examples we use a wavelet representation of higher dimensionality,





**Fig. 9** Linear and nonlinear responses in V1 for an illustrative stimulus. **a** Retinal image composed of a natural image and two synthetic patches of frequencies 24 and 12 cpd. This image goes through the first stages of the model (see Fig. 1) up to the cortical layer, where a set of linear wavelet filters leads to the responses with energy  $e$ , which are nonlinearly transformed into the responses  $x$ . **b.1** Wavelet panel that represents  $e$ . **c.1** Wavelet panel that represents  $x$ . The highlighted sensors in red and blue (tuned to different locations of the 24 cpd scale, horizontal orientation) have characteristic responses given the image patterns in those locations. The plots **b.2** and **c.2** show the vector representation of the wavelet responses arranged according to the MatlabPyrTools convention (Simoncelli et al. 1992). These plots show how natural images typically have bigger energy in the low-frequency sensors. **d** Input–output scatter plots at different spatial frequencies, in cycles per degree (cpd), and demonstrate that Divisive Normalization (and the Wilson–Cowan solution) imply adaptive saturating nonlinearities depending on the neighbors (i.e., a family of sigmoid functions) (Color figure online)

with 4 scales and 4 orientations, applied on bigger images,  $64 \times 64$ . Regarding the parameters, we use unit-norm Gaussian kernels in  $H^{ws}$  or  $W$ , and constants  $k$  and  $b$  also defined over 4 scales and 4 orientations, directly taken from Martinez-Garcia et al. (2019).

### 4.3.1 Structure of the Kernel in Divisive Normalization

Here, we compare the empirical filters  $\mathbb{D}_l$  and  $\mathbb{D}_r$  that had to be introduced ad hoc in Martinez-Garcia et al. (2019), with the theoretical ones obtained through Eq. 13.

Before going into the details of the kernel, let's get some intuition on the typical structure of the vectors  $x$  and  $g_n(x)$ . Figure 9 shows an illustrative stimulus with

oriented textures and the corresponding responses of linear and nonlinear V1-like sensors based on steerable wavelets. Typical responses for natural images are low-pass signals (see the vectors in Fig. 9b.2, c.2). The response in each subband is an adaptive (context dependent) nonlinear transduction (Fig. 9d). Each point in Fig. 9d represents the input–output relation for each neuron in the subbands of the different scales (from coarse to fine). As each neuron has a different neighborhood, there is no simple input–output transduction function, but a scatter plot representing different instances of an adaptive transduction.

The considered image is designed to lead to specific excitations in certain sensors (subbands and locations in the wavelet domain). Note, for instance, the high- and low-frequency synthetic patterns (24 and 12 cycles per degree, cpd, horizontal and vertical, respectively) in the image regions highlighted with the red and blue dots. In the wavelet representations, we also highlighted some specific sensors in red and blue corresponding to the same spatial locations and the horizontal subband tuned to 24 cpd. Given the tuning properties of the neurons highlighted in red and blue, it makes sense that wavelet sensor in red has bigger response than the sensor in blue.

With this knowledge of the signal in mind: (1) low-pass trend in  $\mathbf{x}$  shown in Fig. 9, (2) bigger derivative  $g_n(\mathbf{x})$  for high frequencies because the derivative is higher for low amplitude signals (see Fig. 4), and (3) the vector  $\mathbf{b}$  is bigger for low frequencies (Martinez-Garcia et al. 2019), we can understand the high-pass nature of the vectors included in the diagonal matrices that appear at the left and right sides of the theoretically derived kernel  $\mathbf{H} = \mathbb{D}\left(\frac{\mathbf{k}}{\mathbf{x}}\right) \cdot \mathbf{W} \cdot \mathbb{D}\left(\frac{\mathbf{k}}{\mathbf{b}} \odot g_n(\mathbf{x})\right)$ .

Figure 10 compares the empirical *left* and *right* vectors,  $\mathbf{l}$  and  $\mathbf{r}$  that were adjusted ad hoc to reproduce contrast curves in Martinez-Garcia et al. (2019), with those based on the theoretical relation proposed here. In this case, we only consider the comparison with the psychophysically sensible parameterization since the ad hoc tuning was done for that specific scenario. As these empirical filters were just qualitatively adjusted in Martinez-Garcia et al. (2019), the reproduction of their high-pass nature and their order of magnitude is more important than the specific MSE values.

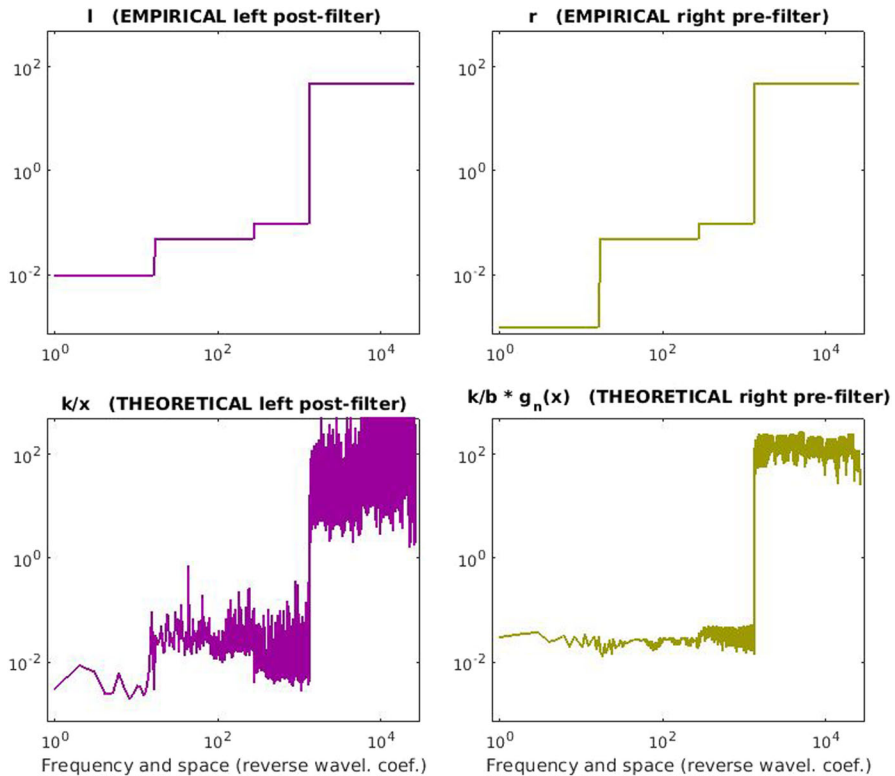
The similarity of the structure of the empirical and theoretical interaction matrices (Eqs. 10, 13) and the coincidence of empirical and theoretical filters (Fig. 10) suggest that the proposed theory explains the modifications that had to be introduced in classical unit-norm kernels in Divisive Normalization to explain contrast response.

#### 4.3.2 Shape Adaptation of the Kernel Depending on the Signal

Once we have shown the global high-pass nature of the vectors  $\frac{\mathbf{k}}{\mathbf{x}}$  and  $\frac{\mathbf{k}}{\mathbf{b}} \odot g_n(\mathbf{x})$ , let's see in more detail the signal-dependent adaptivity of the kernel. In order to do so, let's consider the interaction neighborhood of two particular sensors in the wavelet representation of an illustrative stimulus with easy to understand features. Specifically, the sensors are highlighted in red and blue in Fig. 9.

Figure 11 compares different versions of the two individual neighborhoods displayed in the same wavelet representation: *left* the unit-norm Gaussian kernels,  $\mathbf{H}^{ws}$ , and *right* the empirical kernel modulated by ad hoc pre- and post-filters, Eq. 10. In these diagrams, lighter gray in each  $j$ -th sensor corresponds to bigger interaction with



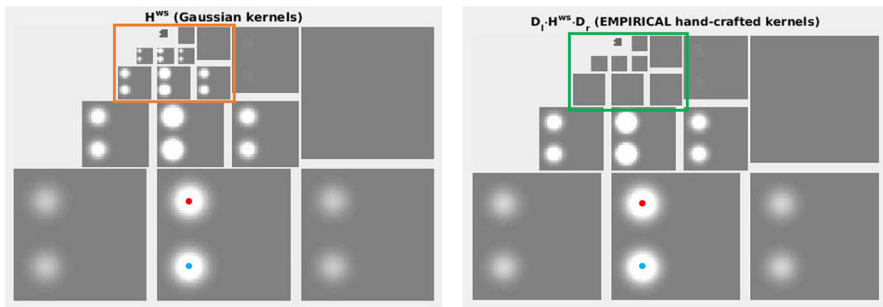


**Fig. 10** Empirical and theoretical modulation of the Divisive Normalization kernel. Vectors in diagonal matrices  $D_l$  and  $D_r$  that multiply the Gaussian kernel in the empirical tuning represented by Eq. 10 (top), and in the theoretically derived Eq. 13 (bottom). These theoretical filters correspond to a specific natural image and using the  $g_{10}(x)$  approximation of the  $\gamma$ -activation. The median relative RMSEs for the predicted filters over 45 natural images are 10.2% (for the left filter) and 5.4% (for the right filter). The difference for the right filter using the original Wilson–Cowan activation and also  $g_{10}(x)$  is 5.1% (curve not shown). However, as the empirical filters were just ad hoc adjusted in Martinez-Garcia et al. (2019), here the relevant is the reproduction of the required high-pass structure, not the MSE

the considered  $i$ -th sensor (highlighted in color). The gray values are normalized to the global maximum in each case. Each subband displays two Gaussians. Obviously, each Gaussian corresponds to only one of the sensors (the one highlighted in red or in blue, depending on the spatial location of the Gaussian). We used a single wavelet diagram since the two neighborhoods do not overlap, and there is no possible confusion between them.

In the base-line unit-norm Gaussian case,  $H^{ws}$ , a unit-volume Gaussian in space is defined centered in the spatial location preferred by the  $i$ -th sensor. Then, the corresponding Gaussians at every subband are weighted by a factor that decays as a Gaussian over scale and orientation from the maximum, centered at the subband of the  $i$ -th sensor.

The *problem* with the unit-norm Gaussian in every scale is that the reduced set of sensors for low-frequency scales lead to higher values of the kernel so that it has the



**Fig. 11** Gaussian and empirical interaction kernels for the sensors highlighted in red and light blue in Fig. 9. Gaussian kernel (left) with overestimated contribution of low-frequency subbands (highlighted in orange). Handcrafted kernel (right) to reduce the influence of low-frequency subbands (highlighted in green) (Color figure online)

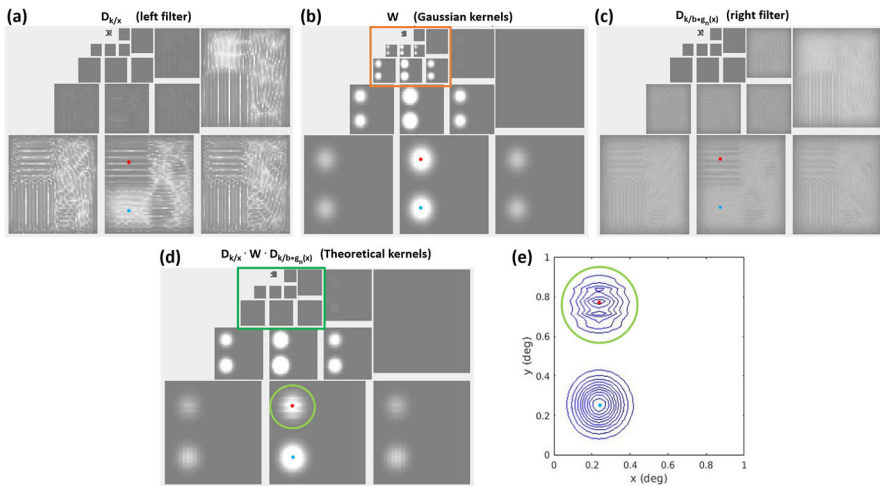
required volume. In that situation, the impact of activity in low-frequency subbands is substantially higher. This fact, combined with the low-pass trend of wavelet signals, implies a strong bias of the response and ruins the contrast masking curves. This *problem* is represented by the relatively high values of the neighborhoods in the low-frequency subbands highlighted in orange.

This overemphasis in the low-frequency scales was corrected ad hoc using right and left multiplication in Eq. 10 by handcrafted high-pass filters. The effect of these filters is to reduce the values for the Gaussian neighborhoods at the low-frequency scales, as seen in the empirical kernel in Fig. 11-right. The positive effect of the high-pass filters is reducing the impact of the neighborhoods at low-frequency subbands (highlighted in green).

In both cases (the classical  $H^{ws}$ , and the handcrafted  $H = \mathbb{D}_l \cdot H^{ws} \cdot \mathbb{D}_r$ ), the size of the interaction neighborhood (the interaction length) is signal independent. Note that the neighborhoods for both sensors (red and blue) are the same, regardless of the different stimulation that can be seen in Fig. 9.

Figure 12 shows the kernels obtained from Eq. 13. The three components of  $H$  are: in Fig. 12a the term proportional to  $\frac{1}{x}$ , in Fig. 12b the term based on Gaussian neighborhoods  $W$ , and in Fig. 12c the term proportional to  $g_n(x)$ . Finally, Fig. 12d shows the global result of the product of the three terms and Fig. 12e zooms on the high-frequency horizontal subband that contains the co-linear situation considered in the physiological experiments (Cavanaugh et al. 2002b).

These three terms have the following positive results: (1) the product by the high-pass terms moderates the effect of the unit-norm Gaussian at low-frequency subbands as in the empirical kernel tuned in Martinez-Garcia et al. (2019) shown in Fig. 11-right, (2) the term proportional to  $\frac{1}{x}$  scales the interaction length according to the signal, and (3) the shape of the kernel depends on the signal because  $H_{ij}$  is modulated by  $(g_n(x))_j$ , and this implies that when the surround is aligned with the sensor, the kernel elongates in that direction (as the probability of co-assignment in Fig. 3b). This will lead to smaller responses when the sensor is flanked by co-linear stimuli [as in Cavanaugh et al. results (Cavanaugh et al. 2002b)].



**Fig. 12** Changes in the shape of the interaction in the theoretically derived kernel. Panels **a–c** show the isolated factors in the kernel matrix  $H$ , assuming a Gaussian wiring in  $W$ . The Gaussian component implies interactions with low frequencies (highlighted in orange). **d** Shows the interaction kernel resulting from the product of the three factors  $H = \mathbb{D}\left(\frac{k}{x}\right) \cdot W \cdot \mathbb{D}\left(\frac{k}{b} \odot g_n(x)\right)$ , for the two highlighted points. Here, we used  $g_{10}(x)$ . Note the high-pass effect of the left- and right-matrix product over  $W$ , that removed the interaction in the low-frequency subbands, now highlighted in dark green. **e** Zoom on the high-frequency horizontal subband. The term depending on the derivatives implies changes of the shape of the kernel (from circular to horizontal ellipses) when the context is a high contrast horizontal pattern. This is compatible with the probabilities of co-assignment (Coen-Cagli et al. 2012) recalled in Fig. 3b (Color figure online)

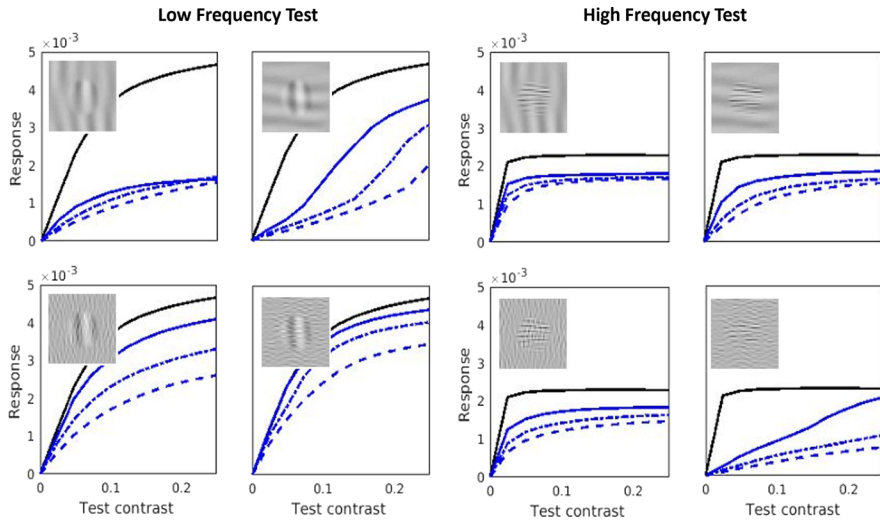
In summary, deriving the Divisive Normalization as the steady state of a Wilson–Cowan system with Gaussian unit-norm wiring explains two experimental facts: (1) the high-pass filters that had to be added to the structure of the kernel in Divisive Normalization to reproduce contrast responses (Martinez-Garcia et al. 2019), and (2) the adaptive asymmetry of the kernel that changes its shape depending on the background texture (Nelson and Frost 1985; Deangelis et al. 1994; Walker et al. 1999; Cavanaugh et al. 2002a, b).

#### 4.3.3 Contrast Response Curves from the Wilson–Cowan Model

The above results suggest that the Wilson–Cowan model could successfully reproduce contrast response curves and masking, which have not yet been addressed through this model. Here, we explicitly check this hypothesis.

We can use the proposed relation, Eq. 13, to plug successful parameters of Divisive Normalization tuned for contrast perception into the corresponding Wilson–Cowan model. We can avoid the integration of the differential equation using the knowledge of the steady state. The only problem to compute the response through the steady-state solution is that the kernel of the Divisive Normalization depends on the (still unknown) response.

In this case, we compute a first guess of the response,  $\hat{x}$ , using the fixed handcrafted kernel tuned in Martinez-Garcia et al. (2019), and then, this first guess is used to



**Fig. 13** Contrast response curves obtained from the Wilson–Cowan model. Contrast response curves for low spatial frequency vertical tests (left) and high spatial frequency horizontal tests (right) seen on top of backgrounds of different spatial frequencies, orientations, and contrasts (see representative stimuli in the insets). The backgrounds include: (1) two spatial frequencies (low and high, corresponding to the top and bottom rows, respectively); (2) two orientations (vertical and horizontal, as seen in the insets); and (3) four different contrasts represented by the line styles (0.0, 0.15, 0.30, and 0.45, corresponding to the black solid line, blue solid line, dotted blue line, and dashed blue line, respectively). The responses display the qualitative trends of contrast perception: frequency selectivity, saturation with contrast, and cross-masking depending on spatio-frequency similarity between test and background (Color figure online)

compute the proposed signal-dependent kernel, which in turn is used to compute the actual response,  $\mathbf{x}$ .

Figure 13 shows the response curves corresponding to neurons that are tuned to low and high spatial frequency tests, as a function of the contrast of these tests located on top of backgrounds of different contrast, spatial frequency, and orientation. In each case, we considered four different contrasts for the background (represented by the different line styles). Representative stimuli are shown as image patches inside each plot. The results in this figure display the expected qualitative properties of contrast perception:

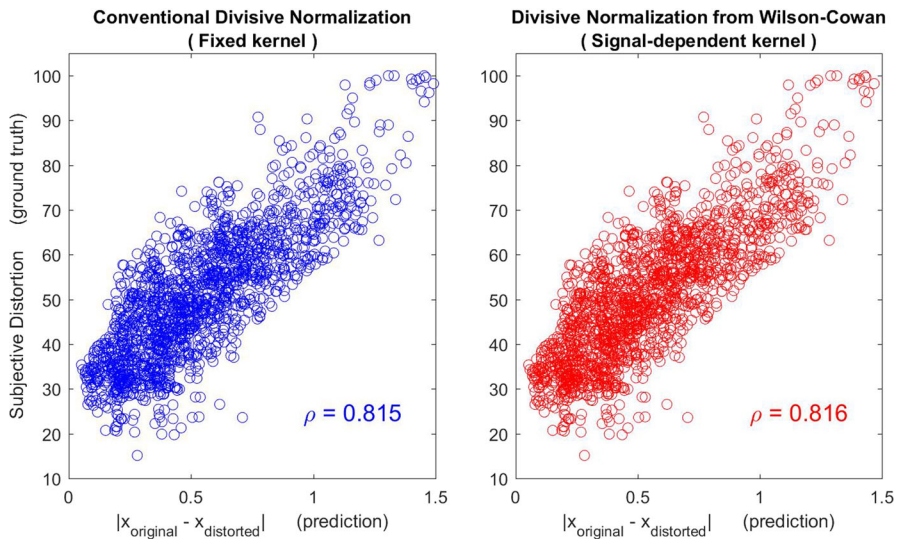
#### *Frequency selectivity.*

The magnitude of the response depends on the frequency of the test: Responses for the low-frequency test are bigger than the responses for the high-frequency test. This frequency-dependent behavior in Fig. 13 is consistent with the Contrast Sensitivity Function (Campbell and Robson 1968).

#### *Saturation.*

The responses increase with the contrast of the test, but this increase is nonlinear (saturates), and the responses decrease with the contrast of the background. This behavior in Fig. 13 is consistent with the contrast discrimination results in Legge and Foley (1980), Legge (1981).

#### *Cross-masking.*



**Fig. 14** Subjective image distortion using the handcrafted kernel (Martinez-Garcia et al. 2019) (left), and the kernel based on Wilson–Cowan equations (right). For each scatter plot, the Pearson correlation between Mean Opinion Scores (ordinates) and predicted image distortions (abscissas) is given. Differences in the correlations are not statistically significant indicating the validity of the proposed relation

Reduction of the responses depends on the frequency similarity between test and background. Note that the low-frequency test is more attenuated by the low-frequency background of the same orientation than by the high-frequency background of orthogonal orientation. Similarly, the high-frequency test is more affected by the high-frequency background of the same orientation. This behavior in Fig. 13 is consistent with cross-masking results in Foley (1994), Watson and Solomon (1997).

#### 4.3.4 Metric in the Image Space from the Wilson–Cowan Model

As a result of the derived relation between models, Eq. 13, the Wilson–Cowan model may also be used to predict subjective image distortion scores. In this section, we explicitly check the performance of the Wilson–Cowan response to compute the visibility of distortions from neural differences following the same approach detailed in the previous section regarding the computation of the signal-dependent kernel and its use to obtain the steady state.

The TID database (Ponomarenko et al. 2008, 2009) contains natural images modified with many kinds of degradation and has the experimental subjective distortion for each degraded image. Given a model, the theoretical prediction of the subjective distortion is obtained from the modulus  $|x_{\text{orig}} - x_{\text{distort}}|$ , i.e., the Euclidean difference of the model responses to the original and to the degraded images.

Figure 14 compares these predictions (abscissas) with the experimental distortions (ordinates) for the responses with a fixed interaction kernel (the conventional Divisive Normalization approach, in blue) and with the proposed signal-adaptive kernel obtained from the Wilson–Cowan model in red.

The high values obtained for the Pearson's correlation coefficients in both cases, and the close similarities between the plots, prove the good performance of the models and the validity of the proposed relation between them.

## 5 Final Remarks

In this paper, we derived an analytical relation between two well-known models of nonlinear neural interaction: the Wilson–Cowan model (Wilson and Cowan 1972, 1973) and the Divisive Normalization model (Carandini and Heeger 1994, 2012). Specifically, assuming that the Divisive Normalization is the steady-state solution of the Wilson–Cowan differential equations, the Divisive Normalization interaction kernel may be derived from the Wilson–Cowan kernel weighted by two signal-dependent contributions.

We showed the appropriateness of the proposed relation in a range of model parameterizations by checking the convergence of the Wilson–Cowan solution to the Divisive Normalization solution, and by proving that the Divisive Normalization solution is a stable node of the Wilson–Cowan system.

Moreover, the derived relation has the following implications in contrast perception: (a) the specific structure obtained for the interaction kernel of Divisive Normalization explains the need of high-pass filters for unit-norm Gaussian interactions to describe contrast masking found in Martinez-Garcia et al. (2019); (b) the signal-dependent kernel predicts elongations of the interaction neighborhood in backgrounds aligned with the sensor, thus providing a mechanistic explanation to the adaptation facts found in Cavanaugh et al. (2002a, b); and (c) low-level Wilson–Cowan dynamics may also explain behavioral aspects that have been classically explained through Divisive Normalization, such as contrast response curves (Foley 1994; Watson and Solomon 1997), or image distortion metrics (Laparra et al. 2010; Berardino et al. 2017). This is the first work that justifies why the Wilson–Cowan interaction successfully reproduces image distortion metrics and contrast response curves. As stated in Bertalmio et al. (2020a), there are not many works that explore the use of Wilson–Cowan equations to model psychophysics, so the examples presented in this work are relevant to fill this gap.

The assumption of the discrete time step  $\Delta t$  in the Euler integration of the Wilson–Cowan equations, Eq. 14, has implications which were not considered in this work. Here, the specific value for  $\Delta t$  was just an arbitrary choice done for computational convenience. If one could assume that this  $\Delta t = 10^{-5}$  is measured in seconds, as the process converges in about 400–500 Euler steps (as shown in Fig. 5), this would mean 4–5 ms to arrive at the steady-state. As an illustration, 4–5 ms would not be a relevant time delay for visual processing of motion, because the cutoff frequency of the temporal Contrast Sensitivity is about 70–100 Hz (so events below 10–15 ms are disregarded), see Kelly (1979). However, here, the choice for this  $\Delta t$  is not based in the biophysics of the neural interactions, and it may indeed be larger (Zeraati 2023). If the interaction time is in fact larger, the convergence will take longer than 4–5 ms. This would imply that the actual behavior would be given by the dynamic Wilson–Cowan model and not by the Divisive Normalization approximation. This may imply that the use of static models like DN should be limited to slow-varying stimuli, or that

the use of DN is more correct on a certain region of spatiotemporal frequencies (or speeds). Nevertheless, the detailed analysis of that region from a sensible biophysical estimation of  $\Delta t$  is out of the scope here and a matter for further work.

Finally, the relation between models proposed here opens the possibility to analyze Divisive Normalization from new perspectives, following methods that have been developed for Wilson–Cowan systems (Destexhe and Sejnowski 2009). Similarly, mechanisms that generalize the Wilson–Cowan equation such as the neurons with intrinsically nonlinear receptive fields (Bertalmío et al. 2020b) could be analyzed via information theoretic tools that have been used to quantify the performance of Divisive Normalization (Malo 2020, 2022; Saproo and Serences 2014).

**Acknowledgements** This work was partially funded by the Spanish Ministerio de Ciencia e Innovación (MICIIN/FEDER/UE) projects PID2020-118071GB-I00 and PDC2021-121522-C21 and by the Generalitat Valenciana Grants GrisoliaP/2019/035 and CIPROM/2021/056.

**Author Contributions** The study was conceived by JM and discussed with MB. The simulations were carried out by JM and JJE-T. The first draft of the manuscript was written by JM and JJE-T, and all authors commented on posterior versions of the manuscript. All authors read and approved the final manuscript.

**Funding** Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature.

**Data Availability** The datasets generated during and/or analyzed during the current study and the code are available in [http://isp.uv.es/docs/DivNorm\\_from\\_Wilson\\_Cowan.zip](http://isp.uv.es/docs/DivNorm_from_Wilson_Cowan.zip).

## Declarations

**Conflict of interest** The authors declare no conflict of interest.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## A Activation Functions

**A.1 Original Activation.** Wilson and Cowan (Wilson and Cowan 1972, 1973) proposed the use of *typical sigmoids* such as the logistic function as appropriate activation functions:

$$f(x) = C \left( \frac{1}{1 + e^{-\frac{x}{e^*}}} - \frac{1}{2} \right) \quad (16)$$

where  $C = e^* \left( \frac{1}{1+e^{-1}} - \frac{1}{2} \right)^{-1}$  is just a scaling factor of the response and the domain so that  $f(e^*) = e^*$ . In the work, we selected the constants  $e^*$ , i.e., we scaled  $f(\cdot)$ ,



using the standard deviation of the linear responses of the V1 cells over the subbands. In this way, the activation preserves the scale between the different subbands.

The derivative of this original activation function, invoked  $n$  times in  $g_n(\mathbf{x})$ , is given by:

$$\frac{df}{d\mathbf{x}}(\mathbf{x}) = \frac{C}{e^*} \frac{e^{-\frac{\mathbf{x}}{e^*}}}{\left(1 + e^{-\frac{\mathbf{x}}{e^*}}\right)^2} \quad (17)$$

**A.2  $\gamma$ -activation.** Following (Turner and Rieke 2016) that uses power laws to model activation functions, we also used another sigmoid inspired in the luminance-brightness nonlinearities of the retinal photoreceptors (Martinez-Garcia et al. 2018). Due to its similarity to the exponential  $\gamma$ -correction transforms, it is called  $\gamma$ -activation or  $\gamma$ -saturation throughout the work. Regular exponential functions were modified around the origin in Martinez-Garcia et al. (2018) in order to avoid and control the singularity of the derivative in the origin:

$$f(\mathbf{x}) = \begin{cases} \text{sign}(\mathbf{x}) C |\mathbf{x}|^\gamma, & \text{if } |\mathbf{x}| \geq \varepsilon \\ \text{sign}(\mathbf{x}) (a |\mathbf{x}| + b |\mathbf{x}|^2), & \text{if } |\mathbf{x}| < \varepsilon \end{cases} \quad (18)$$

where  $\gamma < 1$  so that the function saturates, and the constant  $C = (e^*)^{1-\gamma}$  ensures that  $f(e^*) = e^*$ . In this work,  $\gamma = 0.6$  and, as stated for the original activation, the scaling  $e^*$  was selected to preserve the relation between the standard deviations of the different subbands in the linear responses. The neighborhood,  $\varepsilon$ , was selected to be close to the origin  $\varepsilon = 1 \times 10^{-3} e^*$ . The constants,  $a$  and  $b$  are chosen to ensure continuity of the derivative at  $\varepsilon$ :  $a = (2 - \gamma) C \varepsilon^{\gamma-1}$ , and  $b = (\gamma - 1) C \varepsilon^{\gamma-2}$ .

The derivative of this function, useful to compute  $g_n(\mathbf{x})$ , is:

$$\frac{df}{d\mathbf{x}}(\mathbf{x}) = \begin{cases} \gamma C |\mathbf{x}|^{\gamma-1}, & \text{if } |\mathbf{x}| \geq \varepsilon \\ a + 2b |\mathbf{x}|, & \text{if } |\mathbf{x}| < \varepsilon \end{cases} \quad (19)$$

Figure 4 in the main text shows that both functions share the same qualitative properties: They are saturating sigmoids, with derivatives that peak at the origin and decrease with the signal.

## B MATLAB Code

This appendix lists the main MATLAB routines associated with each experiment described in the main text. All the material is in [http://isp.uv.es/docs/DivNorm\\_from\\_Wilson\\_Cowan.zip](http://isp.uv.es/docs/DivNorm_from_Wilson_Cowan.zip). Detailed parameters of the models and the instructions on how to use these routines are given in the corresponding \*.m files.

- *The Divisive Normalization retina-cortex model* The MATLAB toolbox that implements the 4-layer network for spectral or color images considered in Fig. 1 is



in `BioMultiLayer_L_NL_color.zip`. This toolbox includes the model, its inverse and Jacobians, and a distortion metric based on the model.

The file `demo_deep_DN_iso_color_spectral.m` shows how to choose the parameters of the model, how to apply it to spectral images and images in opponent color representations, and how to compute the responses, the Jacobians and the inverse.

The demo function `demo_metric_deep_DN_iso_color.m` shows how to represent conventional digital images in the appropriate opponent color representation.

- *Psychophysically sensible parameters for the Wilson–Cowan model* The toolbox includes the functions `saturation_f.m` and `inv_saturation_f.m` that compute and invert the dimension-wise saturating response of the Wilson–Cowan model depicted in Sect. 4.1 of the main text. These functions also compute the corresponding derivative with regard to the stimuli. The routine `Converg_Stability_ND_excit_inhibit.m` defines and plots the parameters of the Wilson–Cowan model based on psychophysically tuned Divisive Normalization.
- *Experiments on convergence* The connectivity  $\mathbf{W}$  and the activation  $f$  are applied together in `integrability_WC_with_f_after_KindReview.m` to check the convergence of the system. That script applies Euler integration and shows the convergence of the dynamic solution to the corresponding Divisive Normalization solution.
- *Experiments on stability* The stability of the dynamic Wilson–Cowan system is studied through the eigendecomposition of the Jacobian that controls the amplification of the perturbations in (`Converg_Stability_ND_excit_inhibit.m`), which includes visualizations of the phase diagram.
- *Signal-dependent kernel* The script `signal_dependent_kernel_with_f.m` generates an illustrative image made of high contrast patterns with selected frequencies to stimulate specific subbands of the models. Then, it computes the responses to such stimulus and the corresponding signal dependent-filters according to the relations derived in the main text, Eq. 13. These theoretical filters are compared to the empirical filters found in Martinez-Garcia et al. (2019). Finally, in environments where the surround is aligned with the wavelet sensors, the shape of the interaction kernel is found to change as in Cavanaugh et al. (2002b).
- *Contrast response curves* The script `contrast_response_WC.m` generates a series of noisy Gabor patterns of controlled frequency and contrast displayed on top of noisy sinusoids of different frequencies, orientations, and contrasts. It computes the visibility of these patterns seen on top of the backgrounds by applying the Divisive Normalization model with the signal-dependent kernel derived from the Wilson–Cowan model. The visibility was computed from the response of the neurons tuned to the tests.
- *Image distortion metric* The series of scripts `images_TID_atd_thorough_WC_model_x.m` compute the Divisive Normalization response with the signal-dependent kernel derived from the Wilson–Cowan model for the original and distorted images of the TID database (previously expressed in the appropriate ATD color space). Then, the Euclidean distance is applied to compute the visibility of the distortions. The distances are computed by applying `metric_deep_DN_iso_colorWC.m` that computes the responses by calling `deep_model_DN_iso_colorWC.m`.

## References

- Amari, S.I.: Dynamics of pattern formation in lateral-inhibition type neural fields. *Biol. Cybern.* **27**(2), 77–87 (1977)
- Berardino, A., Laparra, V., Ballé, J., Simoncelli, E.: Eigen-distortions of hierarchical representations. *Adv. Neural Inf. Process. Syst.* **30**, 3533–3542 (2017)
- Bertalmio, M.: From image processing to computational neuroscience: a neural model based on histogram equalization. *Front. Comput. Neurosci.* **8**(71), 1–7 (2014)
- Bertalmio, M., Cyriac, P., Batard, T., Martinez-Garcia, M., Malo, J.: The Wilson–Cowan model describes contrast response and subjective distortion. *J. Vis.* **17**(10), 657 (2017)
- Bertalmio, M., Calatroni, L., Franceschi, V., Franceschiello, B., Gomez Villa, A., Prandi, D.: Visual illusions via neural dynamics: Wilson–Cowan-type models and the efficient representation principle. *J. Neurophysiol.* **123**(5), 1606–1618 (2020)

- Bertalmío, M., Gomez-Villa, A., Martín, A., Vazquez-Corral, J., Kane, D., Malo, J.: Evidence for the intrinsically nonlinear nature of receptive fields in vision. *Sci. Rep.* **10**, 16277 (2020). <https://doi.org/10.1038/s41598-020-73113-0>
- Bressloff, P.C., Cowan, J.D.: An amplitude equation approach to contextual effects in visual cortex. *Neural Comput.* **14**(3), 493–525 (2002)
- Bressloff, P.C., Cowan, J.D.: The functional geometry of local and horizontal connections in a model of V1. *J. Physiol. Paris* **97**(2), 221–236 (2003)
- Bressloff, P., Cowan, J., Golubitsky, M., Thomas, P., Wiener, M.C.: What geometric visual hallucinations tell us about the visual cortex. *Neural Comput.* **14**, 473–491 (2002). <https://doi.org/10.1162/089976602317250861>
- Bressloff, P.C., Cowan, J.D., Golubitsky, M., Thomas, P.J., Wiener, M.C.: What geometric visual hallucinations tell us about the visual cortex. *Neural Comput.* **14**(3), 473–491 (2002)
- Campbell, F., Robson, J.: Application of Fourier analysis to the visibility of gratings. *J. Physiol.* **197**, 551–566 (1968)
- Camps, G., Gutiérrez, J., Gómez, G., Malo, J.: On the suitable domain for SVM training in image coding. *J. Mach. Learn. Res.* **9**(3), 49–66 (2008)
- Carandini, M., Heeger, D.: Summation and division by neurons in visual cortex. *Science* **264**(5163), 1333–1336 (1994)
- Carandini, M., Heeger, D.J.: Normalization as a canonical neural computation. *Nat. Rev. Neurosci.* **13**(1), 51–62 (2012)
- Cavanaugh, J.R.: Properties of the receptive field surround in macaque primary visual cortex. PhD Thesis, Center for Neural Science, New York Univ. (2000)
- Cavanaugh, J.R., Bair, W., Anthony Movshon, J.: Nature and interaction of signals from the receptive field center and surround in Macaque V1 neurons. *J. Neurophysiol.* **88**, 2530–2546 (2002)
- Cavanaugh, J.R., Bair, W., Movshon, J.A.: Selectivity and spatial distribution of signals from the receptive field surround in Macaque V1 neurons. *J. Neurophysiol.* **88**(5), 2547–2556 (2002)
- Chossat, P., Faugeras, O.: Hyperbolic planforms in relation to visual edges and textures perception. *PLoS Comput. Biol.* **5**, e1000625 (2009)
- Chow, C., Karimipani, Y.: Before and beyond the Wilson–Cowan equations. *J. Neurophysiol.* **123**, 1645–1656 (2020). <https://doi.org/10.1152/jn.00404.2019>
- Coen-Cagli, R., Schwartz, O.: The impact on midlevel vision of statistically optimal divisive normalization in V1. *J. Vis.* **13**(8), 13 (2013)
- Coen-Cagli, R., Dayan, P., Schwartz, O.: Cortical surround interactions and perceptual salience via natural scene statistics. *PLoS Comput. Biol.* **8**(3), e1002405 (2012)
- Deangelis, G., Freeman, R., Ohzawa, I.: Length and width tuning of neurons in the cat's primary visual cortex. *J. Neurophysiol.* **71**, 347–374 (1994). <https://doi.org/10.1152/jn.1994.71.1.347>
- Destexhe, A., Sejnowski, T.J.: The Wilson–Cowan model, 36 years later. *Biol. Cybern.* **101**(1), 1–2 (2009)
- Foley, J.: Human luminance pattern mechanisms: masking experiments require a new model. *J. Opt. Soc. Am. A* **11**(6), 1710–1719 (1994)
- Ghadiyaram, D., Bovik, A.C.: Massive online crowdsourced study of subjective and objective picture quality. *IEEE Trans. Image Process.* **25**(1), 372–387 (2016). <https://doi.org/10.1109/TIP.2015.2500021>
- Gomez-Villa, A., Bertalmio, M., Malo, J.: Visual information flow in Wilson–Cowan networks. *J. Neurophysiol.* **123**, 2249–2268 (2020). <https://doi.org/10.1152/jn.00487.2019>
- Grossberg, S.: Some nonlinear networks capable of learning a spatial pattern of arbitrary complexity. *Proc. Natl. Acad. Sci. USA* **59**, 368–372 (1968). <https://doi.org/10.1073/pnas.59.2.368>
- Grossberg, S.: Nonlinear neural networks: principles, mechanisms, and architectures. *Neural Netw.* **1**, 17–61 (1988). [https://doi.org/10.1016/0893-6080\(88\)90021-4](https://doi.org/10.1016/0893-6080(88)90021-4)
- Hepburn, A., Laparra, V., Malo, J., McConville, R., Santos-Rodriguez, R.: Perceptnet: a human visual system inspired neural network for estimating perceptual distance. In: *IEEE ICIP*, pp. 121–125 (2020)
- Hillis, J.M., Brainard, D.: Do common mechanisms of adaptation mediate color discrimination and appearance? *JOSA A* **22**(10), 2090–2106 (2005)
- Kelly, D.: Motion and vision II: stabilized spatio-temporal threshold surface. *J. Opt. Soc. Am.* **69**(10), 1340–1349 (1979)
- Laparra, V., Muñoz-Marí, J., Malo, J.: Divisive normalization image quality metric revisited. *JOSA A* **27**(4), 852–864 (2010)
- Laparra, V., Jiménez, S., Camps-Valls, G., Malo, J.: Nonlinearities and adaptation of color vision from sequential principal curves analysis. *Neural Comput.* **24**(10), 2751–2788 (2012)

- Laparra, V., Berardino, A., Balle, J., Simoncelli, E.: Perceptually optimized image rendering. *JOSA A* **34**(9), 1511–1525 (2017)
- Legge, G.: A power law for contrast discrimination. *Vis. Res.* **18**, 68–91 (1981)
- Legge, G., Foley, J.: Contrast masking in human vision. *J. Opt. Soc. Am.* **70**, 1458–1471 (1980)
- Logan, J.D.: *A First Course in Differential Equations, A First Course in Differential Equations*, 3rd edn. Springer, Incorporated (2015)
- Malo, J.: Spatio-chromatic information available from different neural layers via Gaussianization. *J. Math. Neurosci.* (2020). <https://doi.org/10.1186/s13408-020-00095-8>
- Malo, J.: Information flow in biological networks for color vision. *Entropy* **24**(10), 1442 (2022). <https://doi.org/10.3390/e24101442>
- Malo, J., Simoncelli, E.: Geometrical and statistical properties of vision models obtained via maximum differentiation. In: *SPIE Electronic Imaging. International Society for Optics and Photonics*, pp. 93,940L–93,940L (2015)
- Malo, J., Laparra, V.: Psychophysically tuned divisive normalization approximately factorizes the PDF of natural images. *Neural Comput.* **22**(12), 3179–3206 (2010)
- Malo, J., Ferri, F., Albert, J., Soret, J., Artigas, J.: The role of perceptual contrast non-linearities in image transform quantization. *Image Vis. Comput.* **18**(3), 233–246 (2000)
- Malo, J., Epifanio, I., Navarro, R., Simoncelli, E.P.: Nonlinear image representation for efficient perceptual coding. *IEEE Trans. Image Process.* **15**(1), 68–80 (2006)
- Martinez-Garcia, M., Cyria, P., Batard, T., Bertalmio, M., Malo, J.: Derivatives and inverse of cascaded linear+nonlinear neural models. *PLoS ONE* **13**(10), e0201326 (2018). <https://doi.org/10.1371/journal.pone.0201326>
- Martinez-Garcia, M., Bertalmio, M., Malo, J.: In praise of artifice reloaded: caution with natural image databases in modeling vision. *Front. Neurosci.* **13**, 8 (2019)
- Nelson, J.I., Frost, B.J.: Intracortical facilitation among co-oriented, co-axially aligned simple cells in cat striate cortex. *Exp. Brain Res.* **61**(1), 54–61 (1985)
- Olshausen, B., Field, D.: Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature* **281**, 607–609 (1996). <https://doi.org/10.1038/381607a0>
- Ponomarenko, N., Carli, M., Lukin, V., Egiiazarian, K., Astola, J., Battisti, F.: Color image database for evaluation of image quality metrics. In: *Proceedings of the International Workshop on Multimedia Signal Processing*, pp. 403–408 (2008)
- Ponomarenko, N., Lukin, V., Zelensky, A., Egiiazarian, K., Astola, J., Carli, M., Battisti, F.: TID2008—a database for evaluation of full-reference visual quality assessment metrics. *Adv. Mod. Radioelectron.* **10**, 30–45 (2009)
- Sapoor, S., Serences, J.T.: Attention improves transfer of motion information between V1 and MT. *J. Neurosci.* **34**(10), 3586–3596 (2014)
- Sato, T., Haider, B., Hausser, M., Carandini, M.: An excitatory basis for divisive normalization in visual cortex. *Nat. Neurosci.* (2016). <https://doi.org/10.1038/nn.4249>
- Schwartz, G., Rieke, F.: Nonlinear spatial encoding by retinal ganglion cells: when  $1 + 1 \neq 2$ . *J. Gen. Physiol.* **138**(3), 283–290 (2011). <https://doi.org/10.1085/jgp.201110629>
- Schwartz, O., Simoncelli, E.: Natural signal statistics and sensory gain control. *Nat. Neurosci.* **4**(8), 819–825 (2001)
- Schwartz, O., Sejnowski, T.J., Dayan, P.: Perceptual organization in the tilt illusion. *J. Vis.* **9**(4), 19 (2009). <https://doi.org/10.1167/9.4.19>
- Simoncelli, E.P., Heeger, D.J.: A model of neuronal responses in visual area MT. *Vis. Res.* **38**(5), 743–761 (1998)
- Simoncelli, E.P., Freeman, W.T., Adelson, E.H., Heeger, D.J.: Shiftable multi-scale transforms. *IEEE Trans. Inf. Theory* **38**(2), 587–607 (1992). <https://doi.org/10.1109/18.119725>
- Teo, P., Heeger, D.: Perceptual image distortion. In: *IEEE Proceedings of 1st International Conference on Image Processing ICIP94*, vol. 2, pp. 982–986 (1994)
- Turner, M., Rieke, F.: Synaptic rectification controls nonlinear spatial integration of natural visual inputs. *Neuron* **90**(6), 1257–1271 (2016)
- van Hateren, J.H., van der Schaaf, A.: Independent component filters of natural images compared with simple cells in primary visual cortex. *Proc. Biol. Sci.* **265**(1394), 359–366 (1998)
- Walker, G.A., Ohzawa, I., Freeman, R.D.: Asymmetric suppression outside the classical receptive field of the visual cortex. *J. Neurosci.* **19**, 10536–10553 (1999)

- Watson, A.B., Solomon, J.A.: Model of visual contrast gain control and pattern masking. *JOSA A* **14**(9), 2379–2391 (1997)
- Wilson, H., Cowan, J.: Excitatory and inhibitory interactions in localized populations of model neurons. *Biophys. J.* **12**, 1–24 (1972)
- Wilson, H.R., Cowan, J.D.: A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue. *Kybernetik* **13**(2), 55–80 (1973)
- Wilson, H.R., Humanski, R.: Spatial frequency adaptation and contrast gain control. *Vis. Res.* **33**(8), 1133–1149 (1993)
- Zeraati, R., et al.: Intrinsic timescales in the visual cortex change with selective attention and reflect spatial connectivity. *Nat. Commun.* **14**(1), 1858 (2023)

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.