SCENE STATISTICS AND DIVISIVE NORMALIZATION

JESÚS MALO

VNIVERSITAT DE VALÈNCIA

SPAIN

http://isp.uv.es

Workshop on Computational Models for Visual Image Processing 2021

https://www.cfp.upv.es/formacion-permanente/curso/taller-modelos-computacionales-procesamiento-visual-imagenes_71578.html

So far (in previous talks)...

FACTS:

METHODS:

APPLICATIONS:

So far (in previous talks)...

FACTS:

- Cone Sensitivities & Adaptation — (Stockman)
- LMS noisy response — (Wandell)
- Tristimulus colorimetry — (Huertas)
- Contrast Sensitivity — (Mantiuk)
- Opponent Color Spaces & CAMs } – (Fairchild)
- Spatial Masking }

METHODS:

APPLICATIONS:

So far (in previous talks)...

FACTS:

- Cone Sensitivities & Adaptation —— (Stockman)
- LMS noisy response —— (Wandell)
- Tristimulus colorimetry —— (Huertas)
- Contrast Sensitivity —— (Mantiuk)
- Opponent Color Spaces & CAMs
- Spatial Masking

} — (Fairchild)

METHODS:

- Psychophysics —— (Párraga & García-Pérez)
- Display Calibration —— (Murdoch)
- Data Analysis —— (Camacho)

APPLICATIONS:

So far (in previous talks)...

FACTS:

- Cone Sensitivities & Adaptation —— (Stockman)
- LMS noisy response —— (Wandell)
- Tristimulus colorimetry —— (Huertas)
- Contrast Sensitivity —— (Mantiuk)
- Opponent Color Spaces & CAMs
- Spatial Masking —— (Fairchild)

METHODS:

- Psychophysics —— (Párraga & García-Pérez)
- Display Calibration —— (Murdoch)
- Data Analysis —— (Camacho)

APPLICATIONS: - Perceptually Rated Images —— (Pedersen)

So far (in previous talks)...

FACTS:

- Cone Sensitivities & Adaptation — (Stockman)
- LMS noisy response — (Wandell)
- Tristimulus colorimetry — (Huertas)
- Contrast Sensitivity — (Mantiuk)
- Opponent Color Spaces & CAMs
- Spatial Masking

} – (Fairchild)

YES, THE VISUAL BRAIN BEHAVES THAT WAY...

So far (in previous talks)...

FACTS:

- Cone Sensitivities & Adaptation —— (Stockman)
- LMS noisy response —— (Wandell)
- Tristimulus colorimetry —— (Huertas)
- Contrast Sensitivity —— (Mantiuk)
- Opponent Color Spaces & CAMs
- Spatial Masking

} – (Fairchild)

YES, THE VISUAL BRAIN BEHAVES THAT WAY...

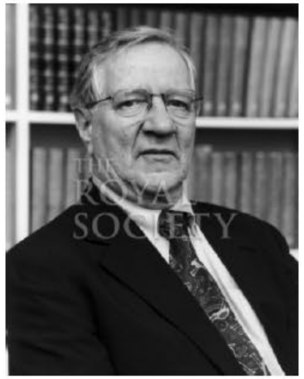
WHY ?

OUTLINE:

0. The WHY question
1. One example: empirical models
2. Information theory tools
3. Natural scenes
4. The two sides of Efficient Coding
5. Open issues

OUTLINE:

BARLOW



0 The WHY question

1 One model → Martinez et al. **PLOS** 2018
https://isp.uv.es/code/visioncolor/vistamodels.html

2 Info. theory tools → SPCA
Malo & Gutiérrez **Network** 2006
Laparra et al. **Neural Comp.** 2012
Laparra et al. **IEEE J. Sel. Top. Sign. Proc.** 2015

RBIG
Laparra et al. **IEEE Trans. Neur. Nets.** 2011
https://isp.uv.es/RBIG4IT.htm

3 Natural scenes → https://isp.uv.es/data_color.htm

4 The two sides of Efficient Coding
Gutmann, Laparra, Hyvarinen & Malo **PLOS** 2014
Laparra & Malo **Front. Neurosci.** 2015
Gomez-Villa et al. **Vision Res.** 2020

Malo & Simoncelli **IEEE Trans. Im. Proc.** 2006
Malo & Laparra **Neural Comp.** 2010
Gómez-Villa et al. **J. Neurophysiol.** 2020
Malo **J. Math. Neurosci.** 2020

5 Open issues →
Martinez et al. **PLOS** 2017
Martínez et al. **Front. Neurosci.** 2019
Li, Gómez, Bertalmío & Malo *Submitted* **JoV.** 2021
Bertalmío et al. **Scientific Reports** 2020
Esteve et al. **Arxiv.** 2020

① ONE EXAMPLE :

Frequency sensors and non-linearities
Image-computable models

- Different behaviors described by A SINGLE MODEL

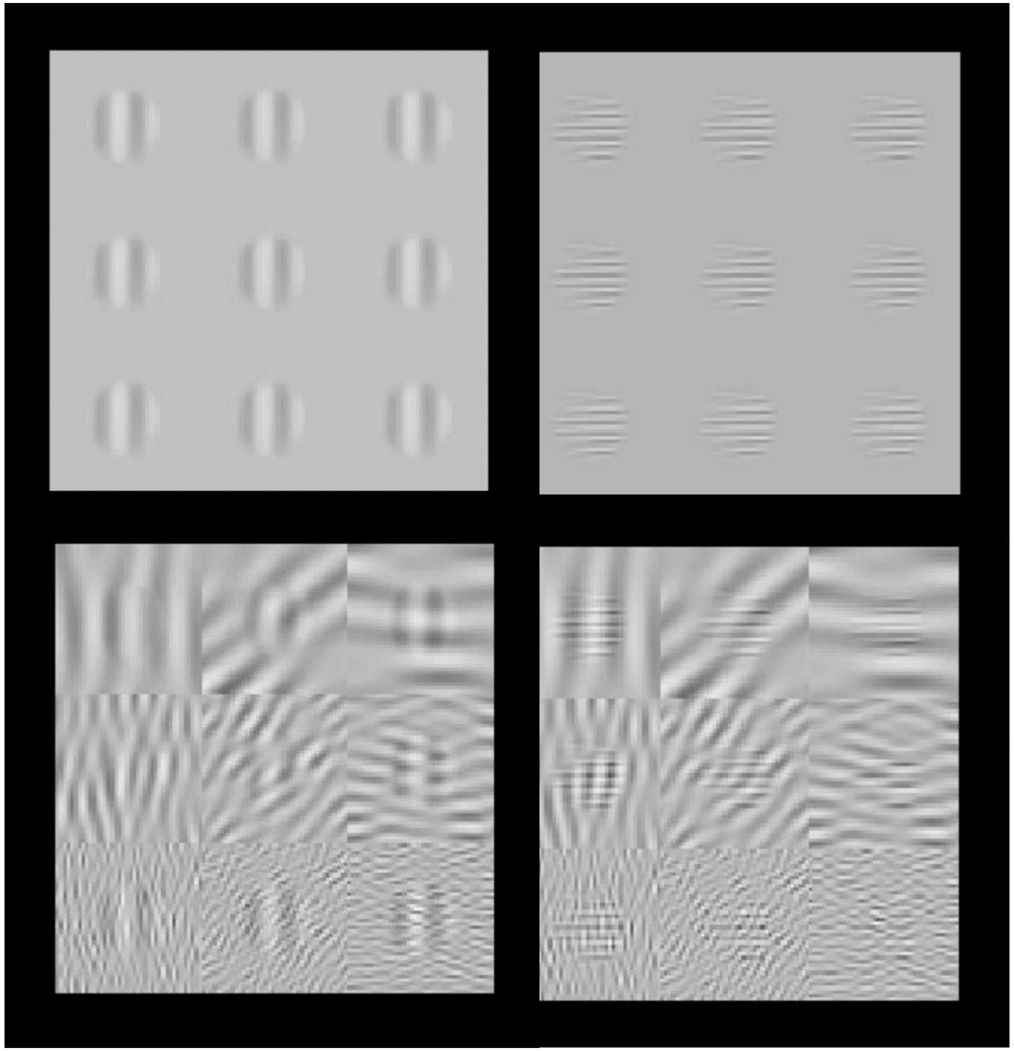
- Linear Sensors + Divisive Normalization

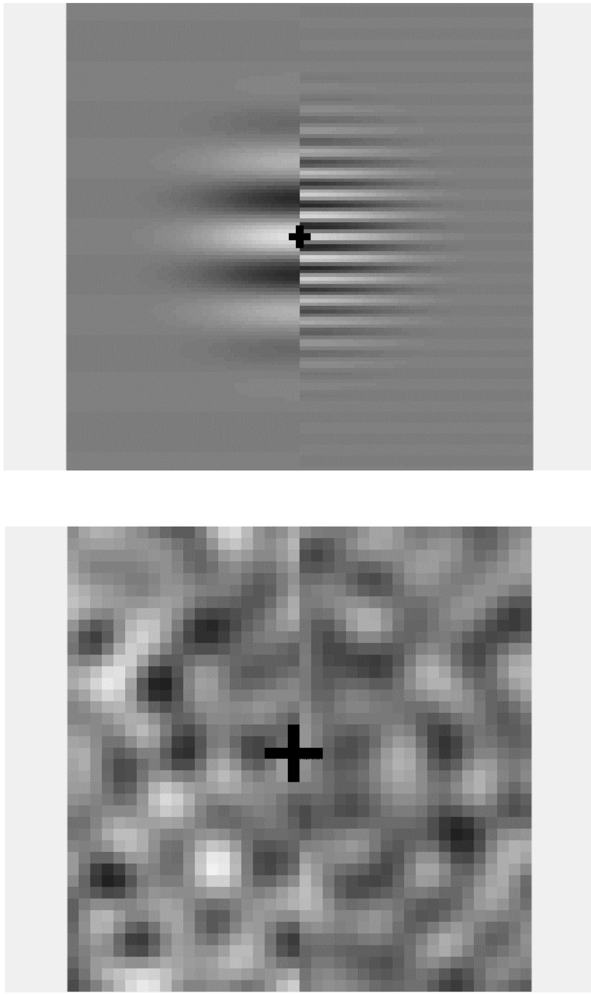
- Empirical image-computable models

① ONE EXAMPLE :

Frequency sensors and non-linearities

Image-computable models

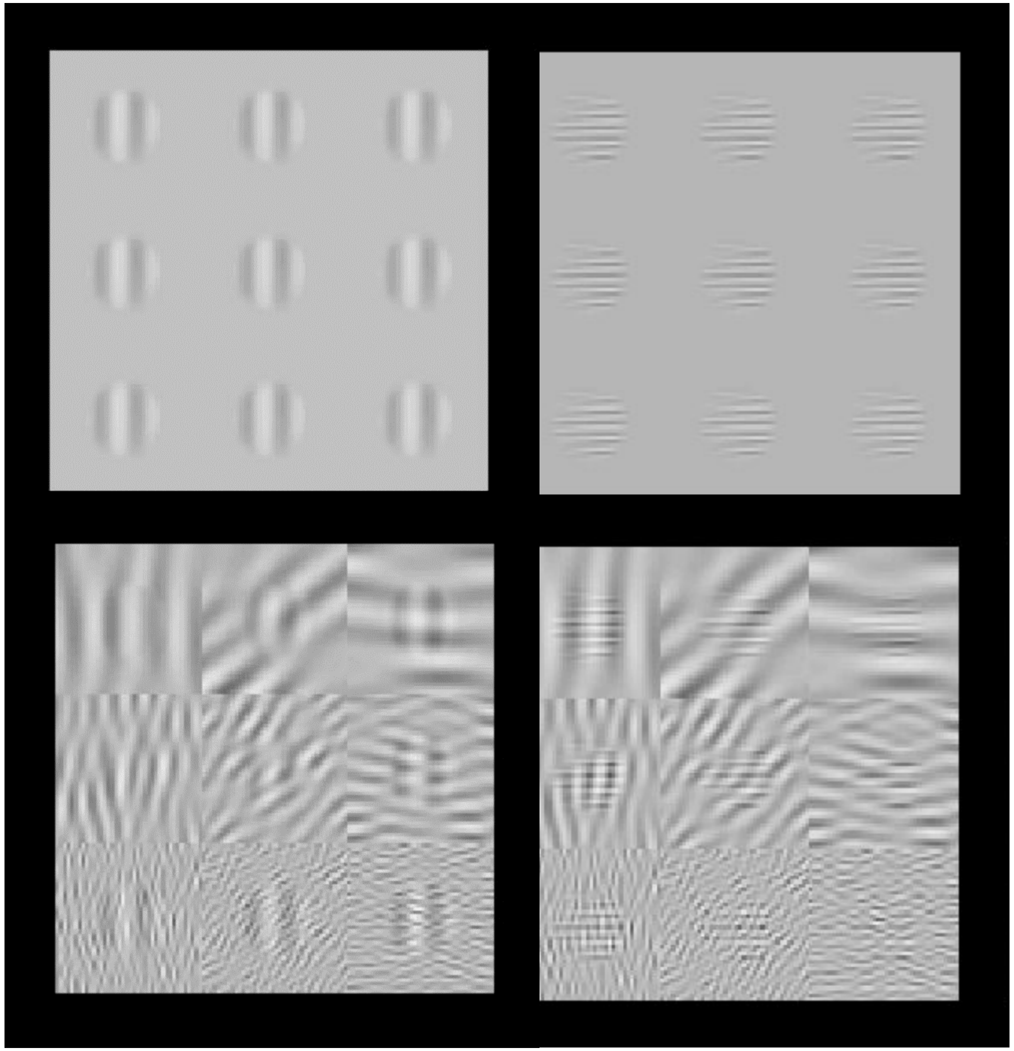


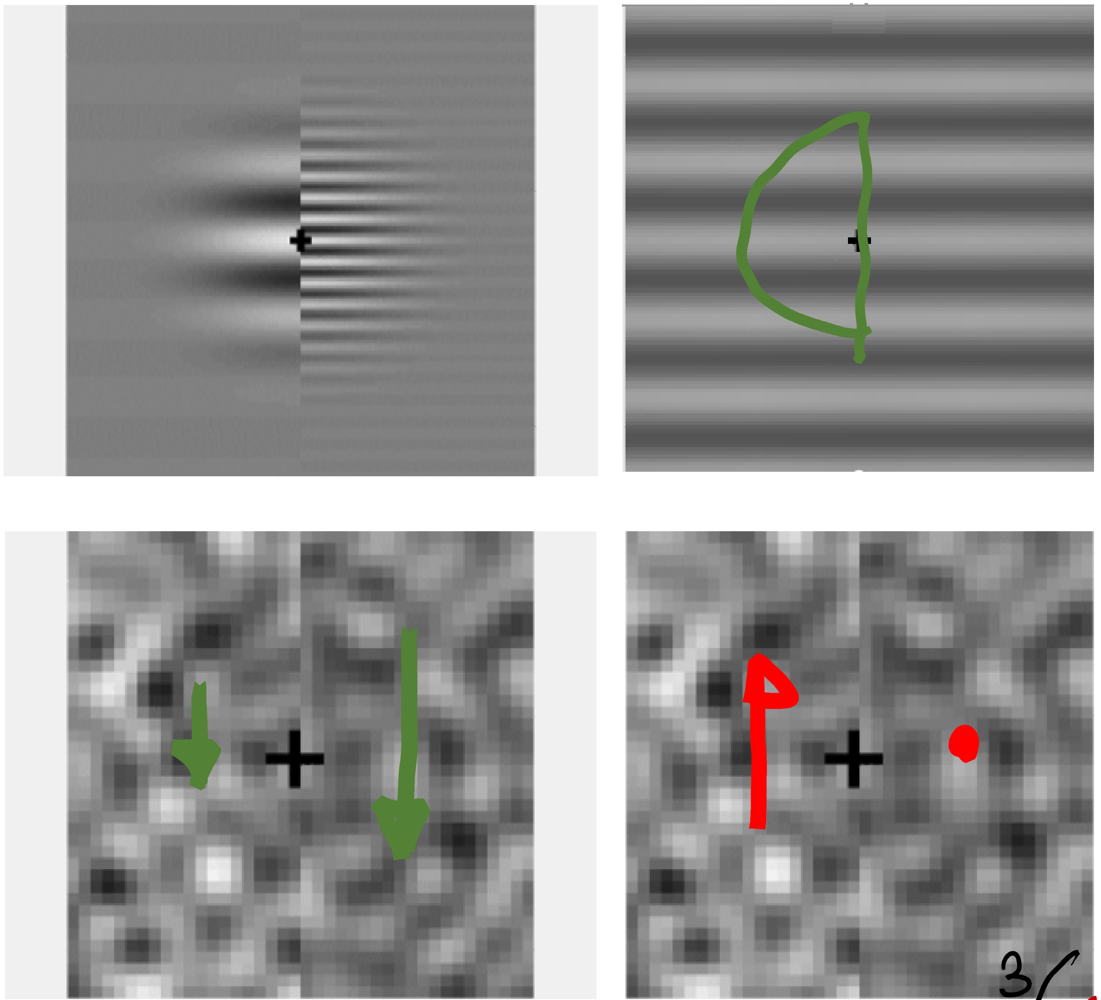


① ONE EXAMPLE :

Frequency sensors and non-linearities

Image-computable models

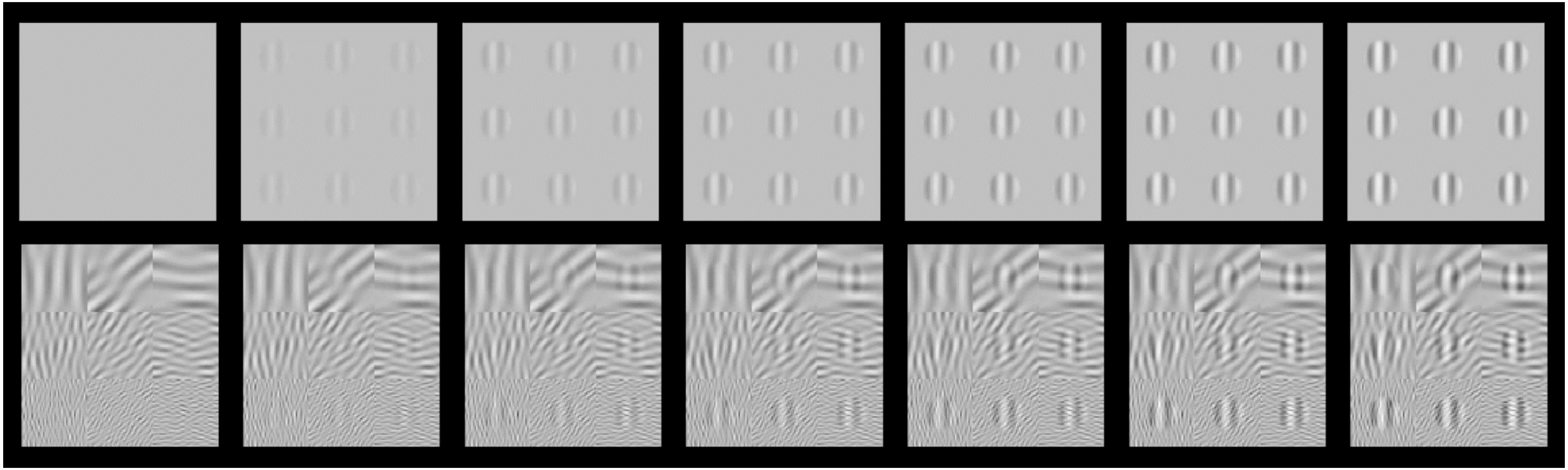




1 ONE EXAMPLE:

Frequency sensors and non-linearities

Image-computable models



NO BACKGROUND

SIMILAR BACKGROUND

DIFFERENT BACKGROUND

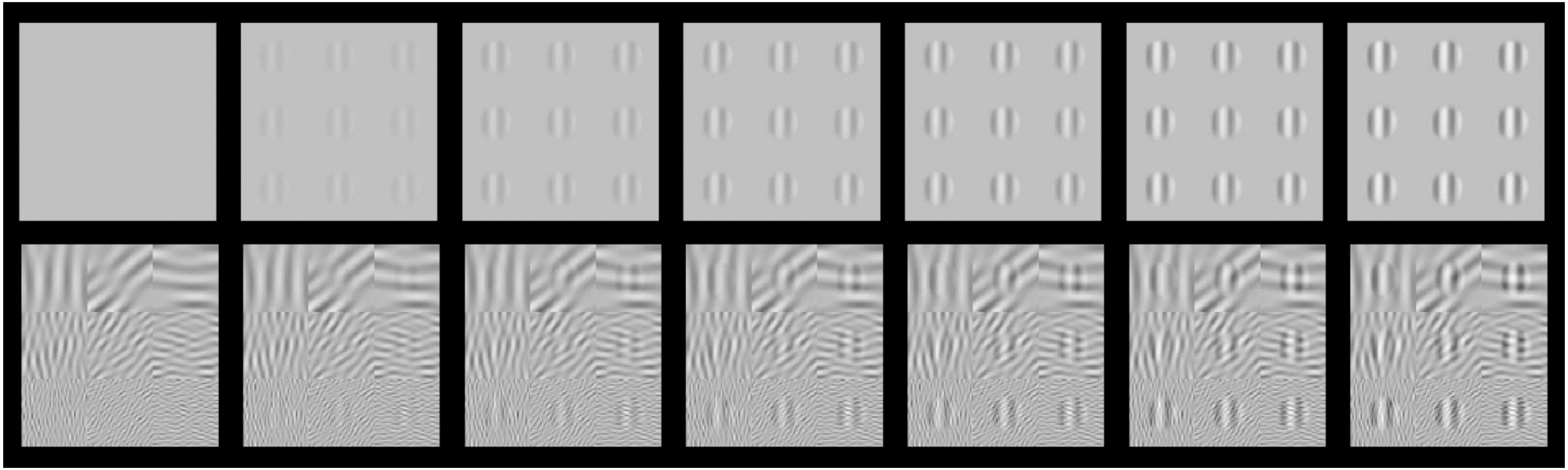
Martinez et al. Front. Neurosci 2019

Watson & Solomon JOSA 1997

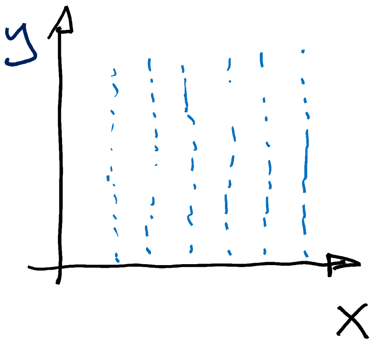
(1) ONE EXAMPLE:

Frequency sensors and non-linearities

Image-computable models



NO BACKGROUND



SIMILAR BACKGROUND

DIFFERENT BACKGROUND

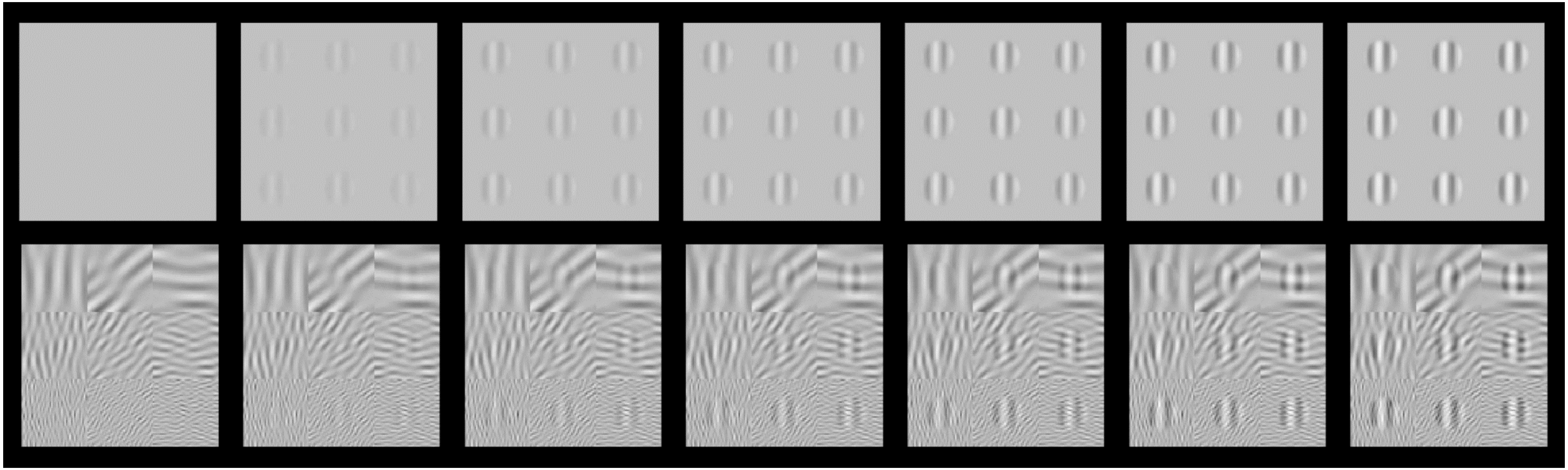
Martinez et al. Front. Neurosci 2019

Watson & Solomon JOSA 1997

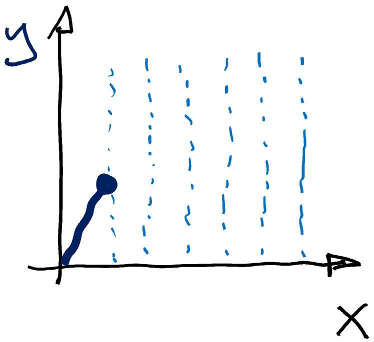
1 ONE EXAMPLE :

Frequency sensors and non-linearities

Image-computable models



NO BACKGROUND



SIMILAR BACKGROUND

DIFFERENT BACKGROUND

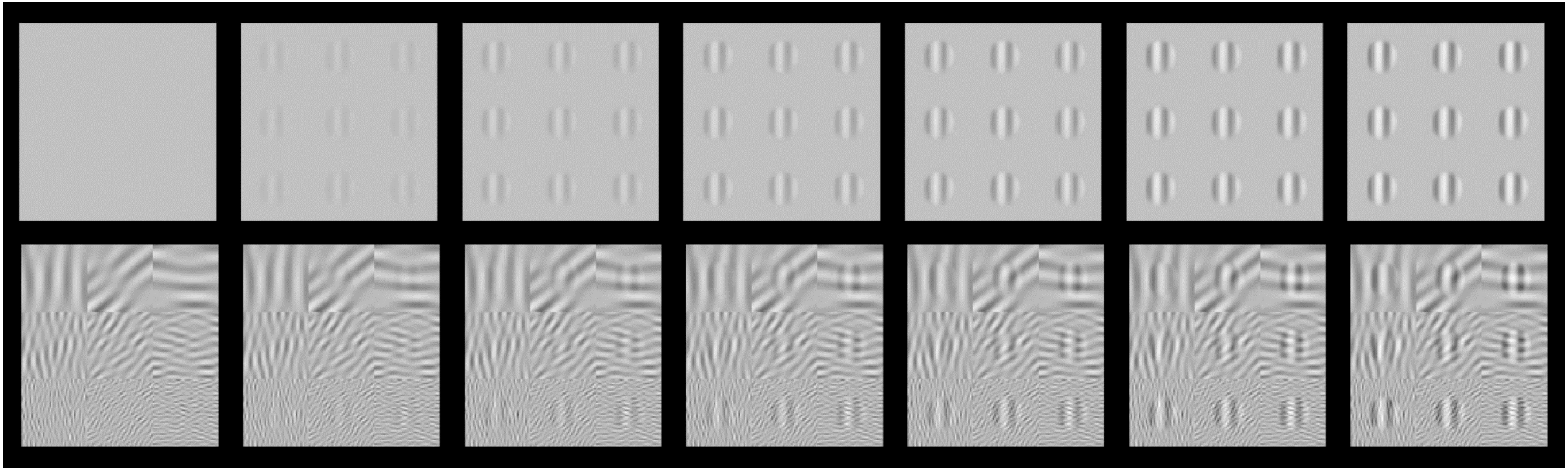
Martinez et al. Front. Neurosci 2019

Watson & Solomon JOSA 1997

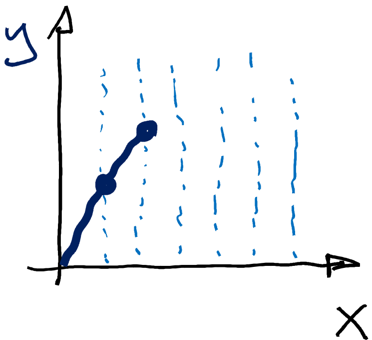
① ONE EXAMPLE :

Frequency sensors and non-linearities

Image-computable models



NO BACKGROUND



SIMILAR BACKGROUND

DIFFERENT BACKGROUND

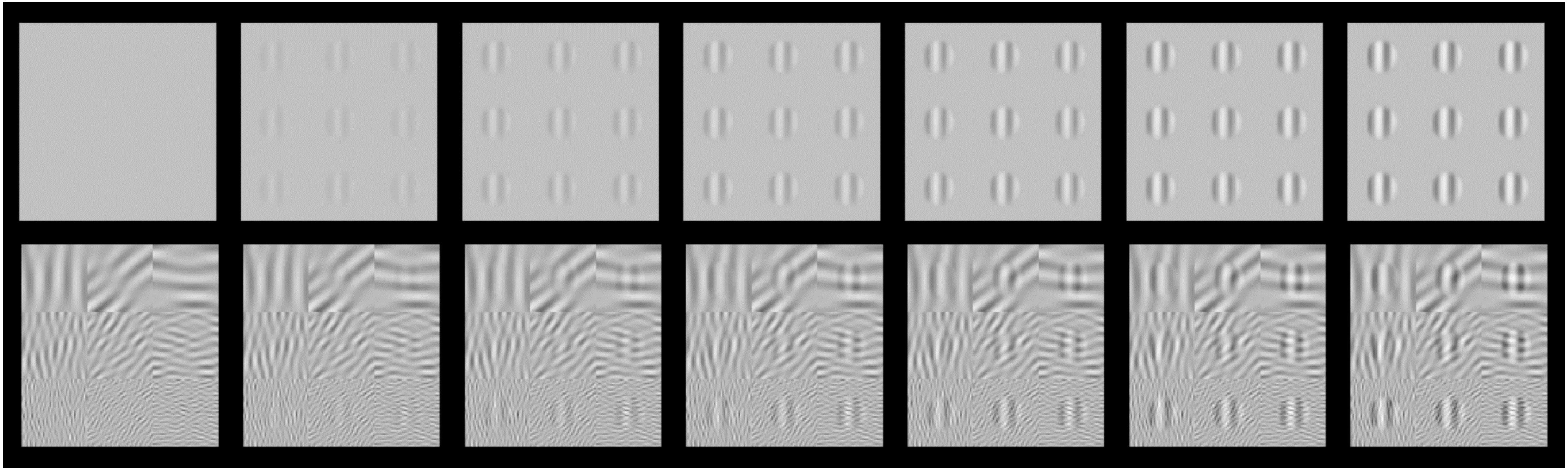
Martinez et al. Front. Neurosci 2019

Watson & Solomon JOSA 1997

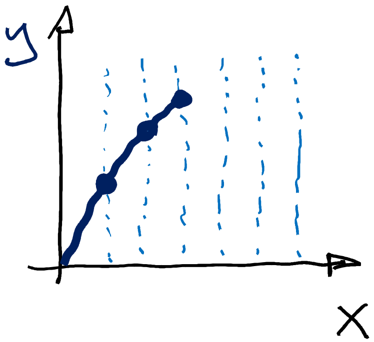
1 ONE EXAMPLE :

Frequency sensors and non-linearities

Image-computable models



NO BACKGROUND



SIMILAR BACKGROUND

DIFFERENT BACKGROUND

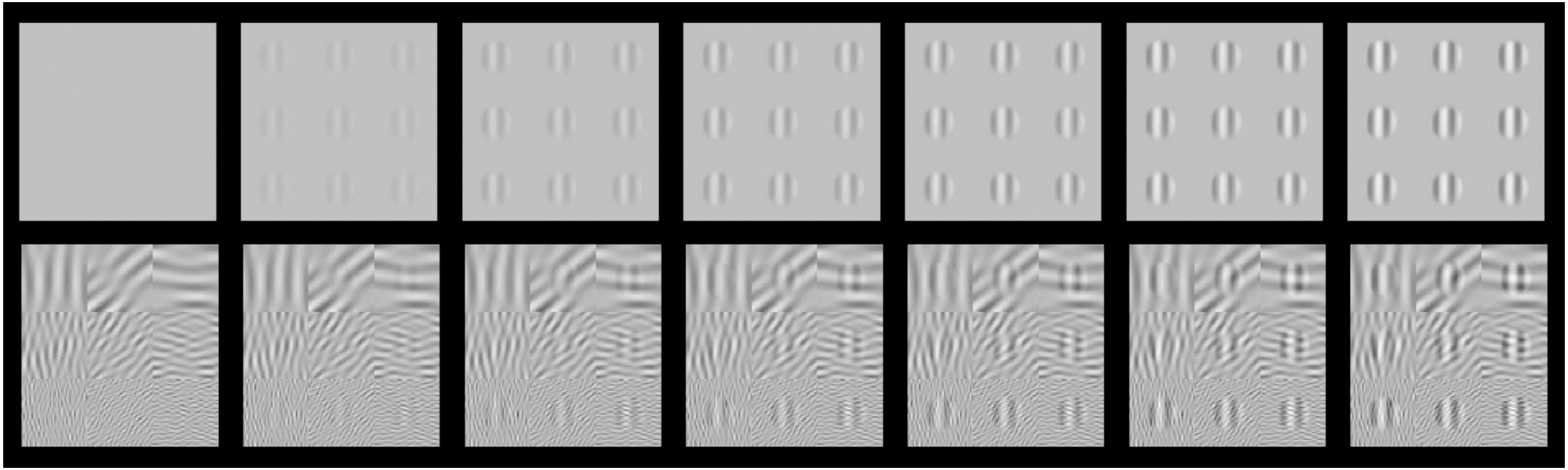
Martinez et al. Front. Neurosci 2019

Watson & Solomon JOSA 1997

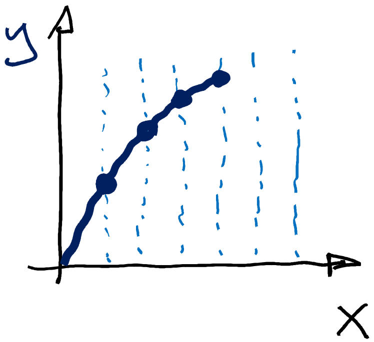
① ONE EXAMPLE :

Frequency sensors and non-linearities

Image-computable models



NO BACKGROUND



SIMILAR BACKGROUND

DIFFERENT BACKGROUND

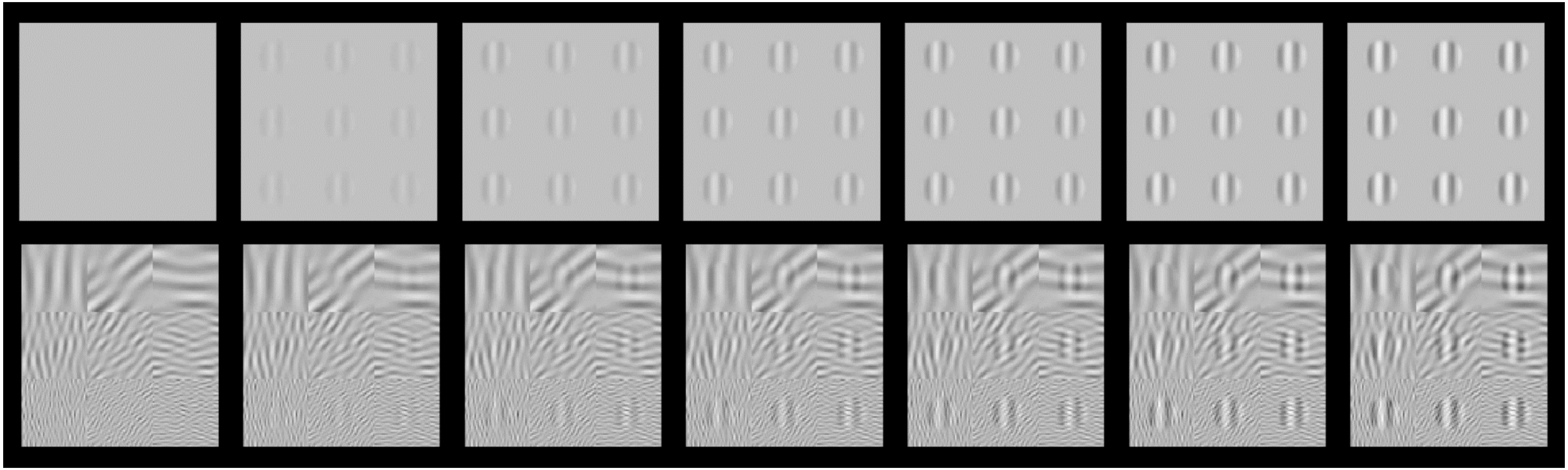
Martinez et al. Front. Neurosci 2019

Watson & Solomon JOSA 1997

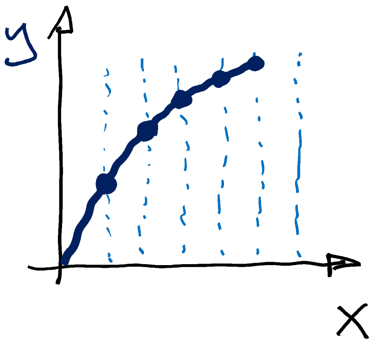
1 ONE EXAMPLE :

Frequency sensors and non-linearities

Image-computable models



NO BACKGROUND



SIMILAR BACKGROUND

DIFFERENT BACKGROUND

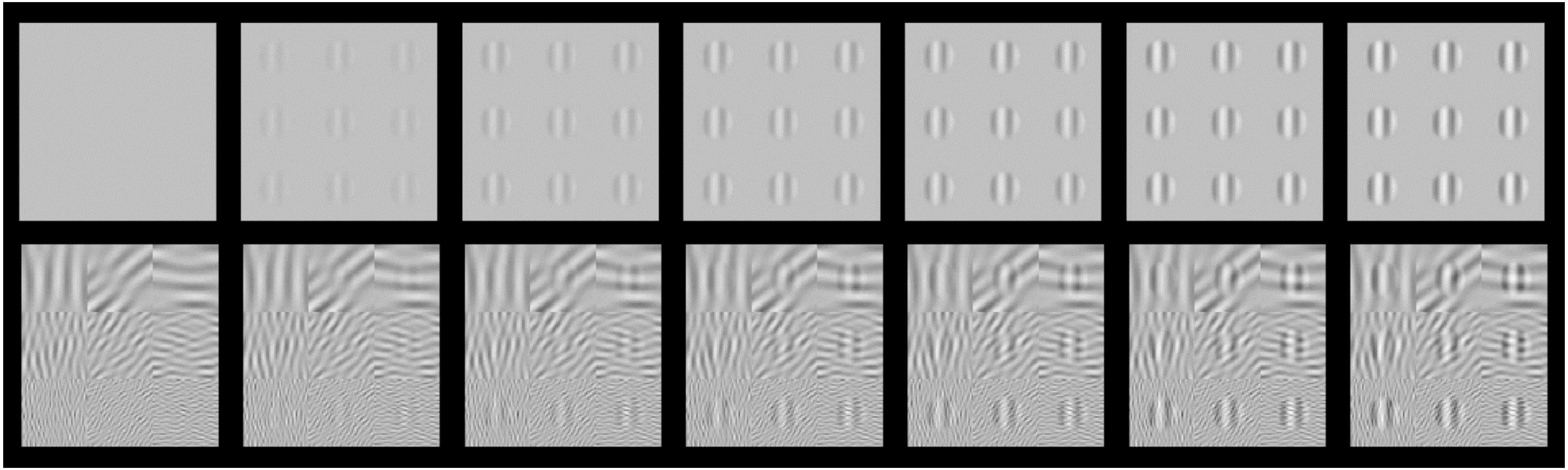
Martinez et al. Front. Neurosci 2019

Watson & Solomon JOSA 1997

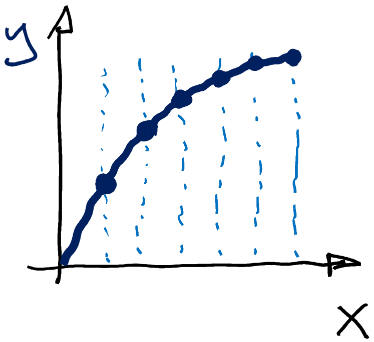
① ONE EXAMPLE:

Frequency sensors and non-linearities

Image-computable models



NO BACKGROUND



SIMILAR BACKGROUND

DIFFERENT BACKGROUND

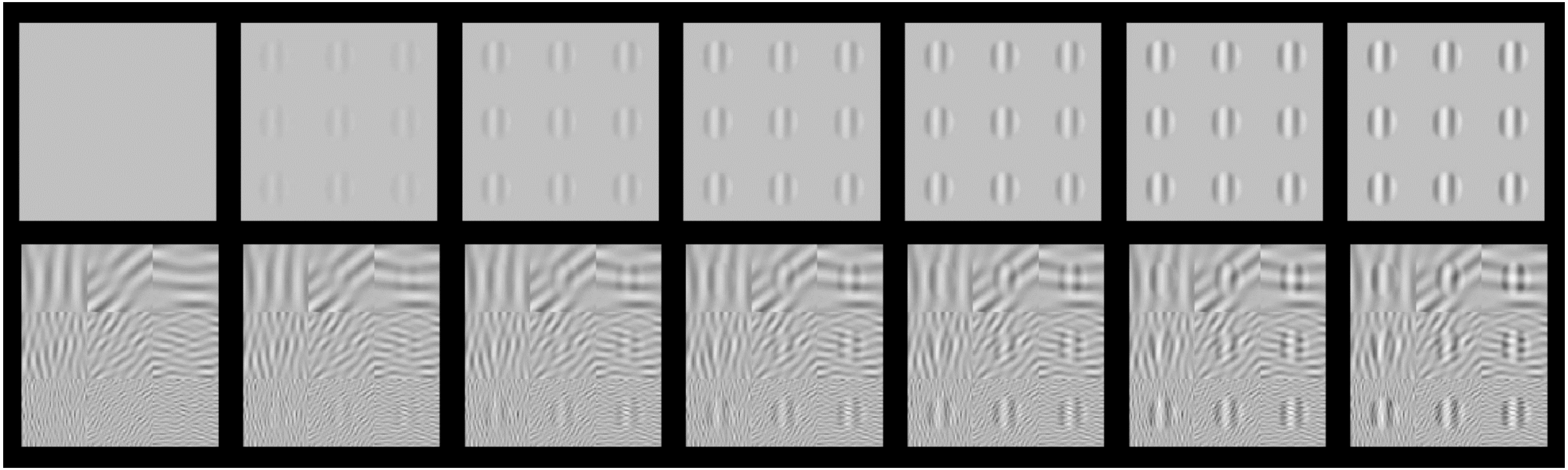
Martinez et al. Front. Neurosci 2019

Watson & Solomon JOSA 1997

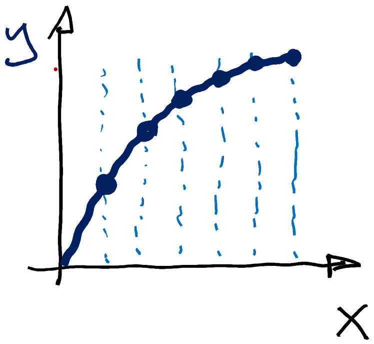
1 ONE EXAMPLE:

Frequency sensors and non-linearities

Image-computable models

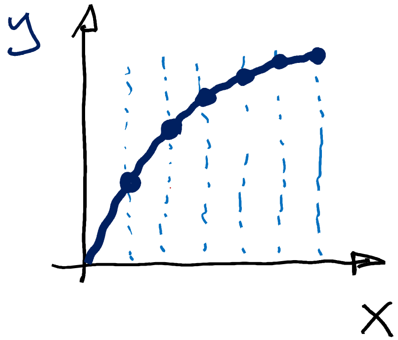


NO BACKGROUND



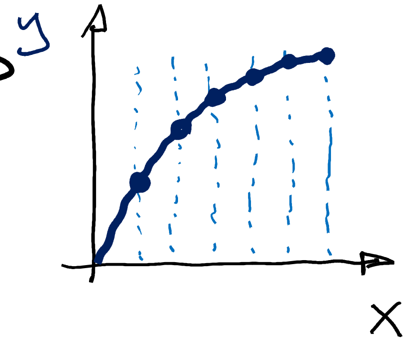
SIMILAR BACKGROUND

$f \sim f'$



DIFFERENT BACKGROUND

$f \neq f'$



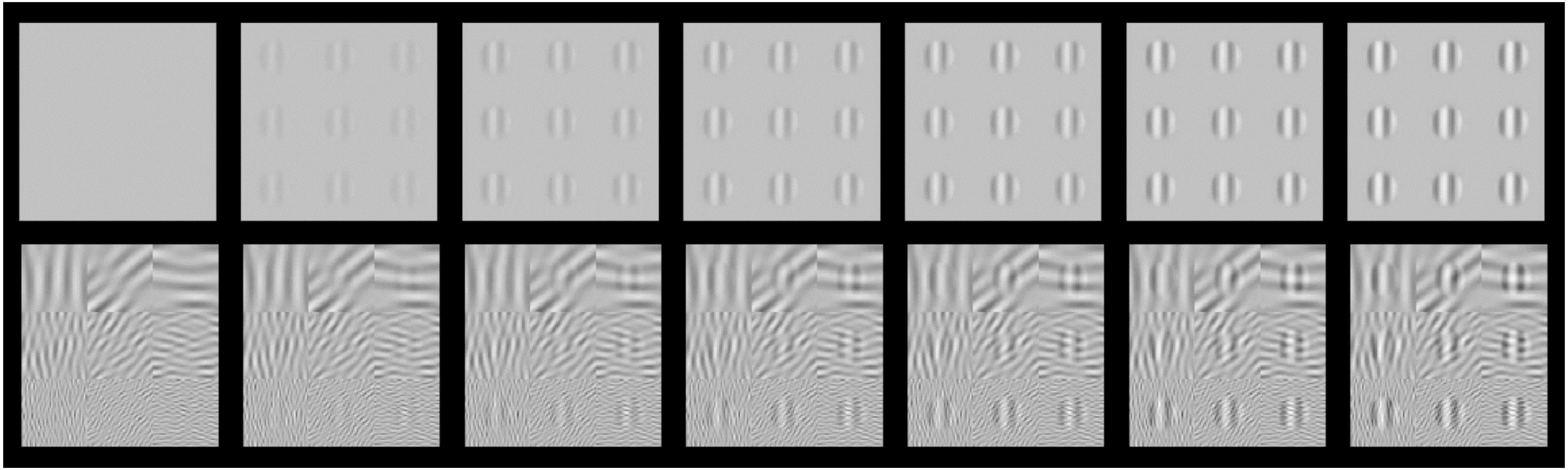
Martinez et al. Front. Neurosci 2019

Watson & Solomon JOSA 1997

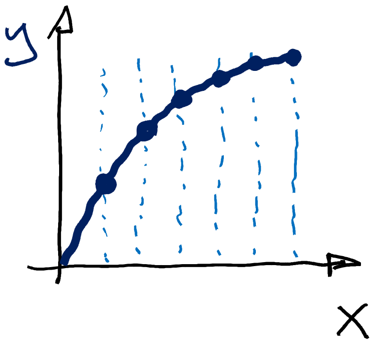
1 ONE EXAMPLE :

Frequency sensors and non-linearities

Image-computable models



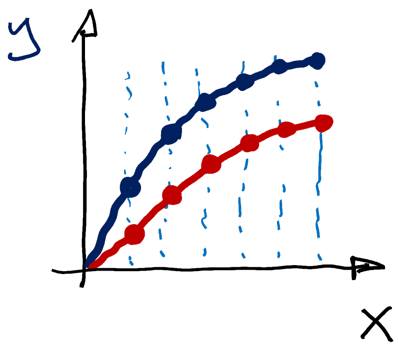
NO BACKGROUND



SIMILAR BACKGROUND

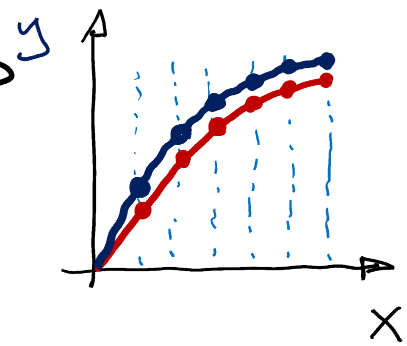
$f \sim f'$

increase c'



DIFFERENT BACKGROUND

$f \neq f'$



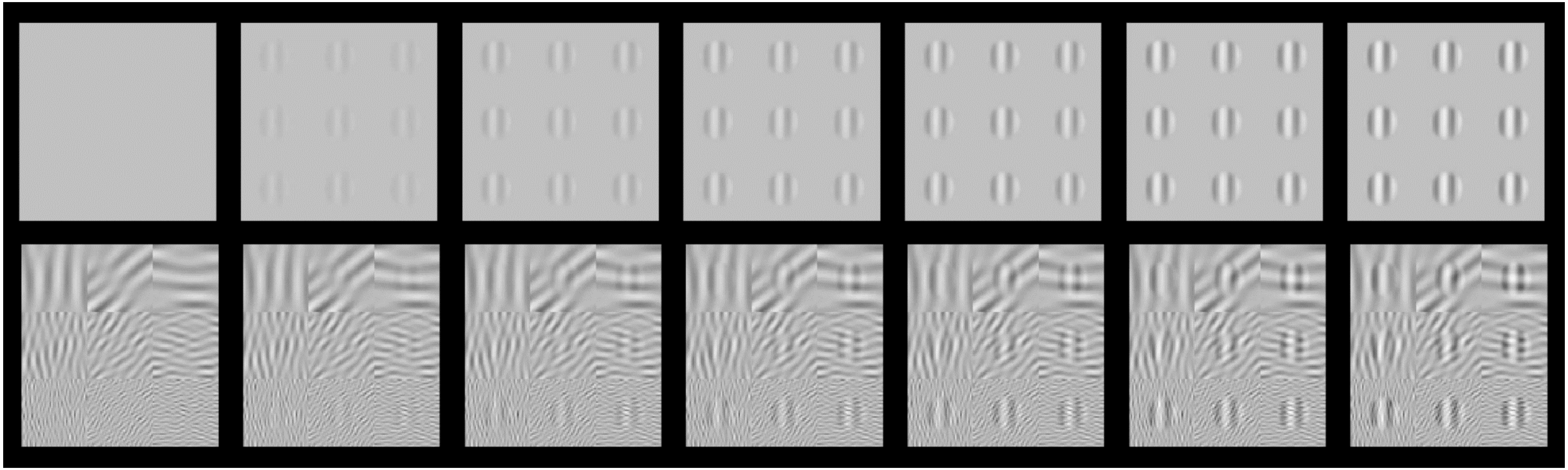
Martinez et al. Front. Neurosci 2019

Watson & Solomon JOSA 1997

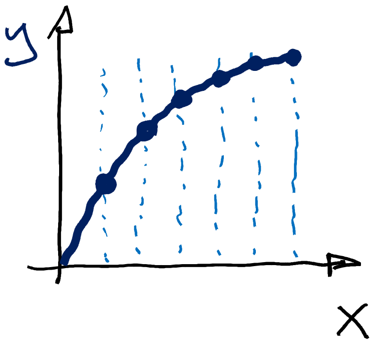
1 ONE EXAMPLE :

Frequency sensors and non-linearities

Image-computable models



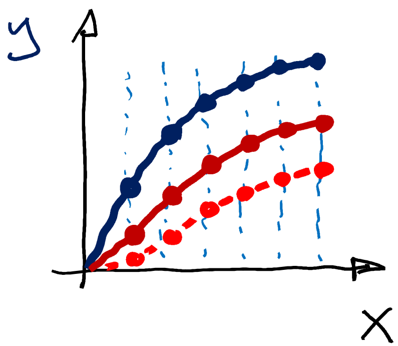
NO BACKGROUND



SIMILAR BACKGROUND

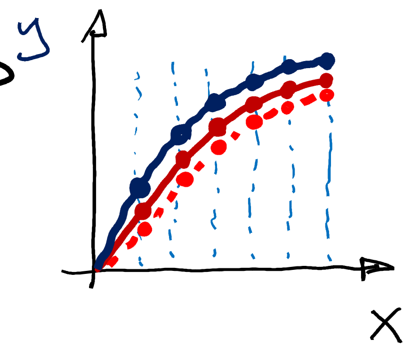
$f \sim f'$

increase c'



DIFFERENT BACKGROUND

$f \neq f'$



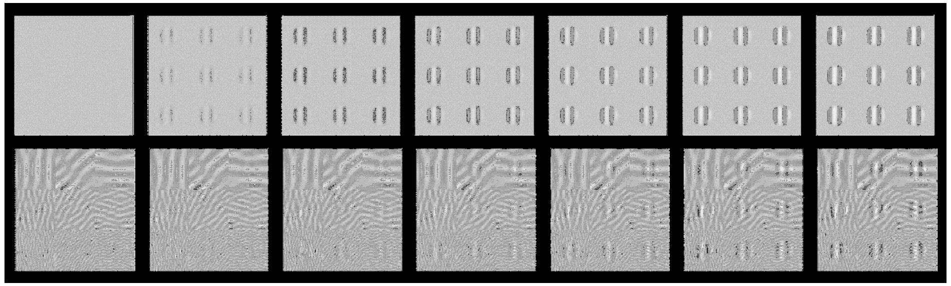
Martinez et al. Front. Neurosci 2019

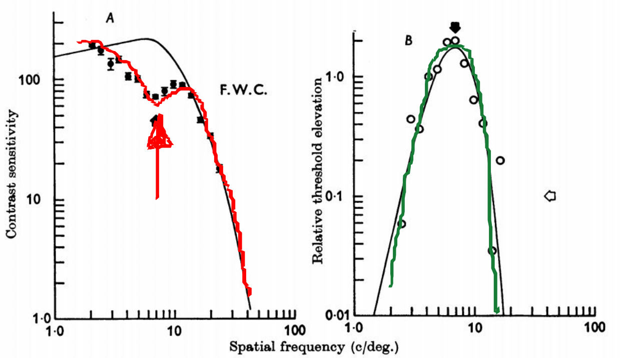
Watson & Solomon JOSA 1997

① ONE EXAMPLE :

Frequency sensors and non-linearities

Image-computable models





Text-fig. 6. The effect of adapting at 7·1 c/deg. *A*. The continuous curve from Text-fig. 5 is reproduced. The filled circles and vertical bars are the means and S.E. ($n = 6$) for re-determinations of contrast sensitivity at a number of spatial frequencies while F. W.C. was continuously adapting to a grating of 7·1 c/deg., 1·5 log. units above threshold. The exact procedure is described in the text.

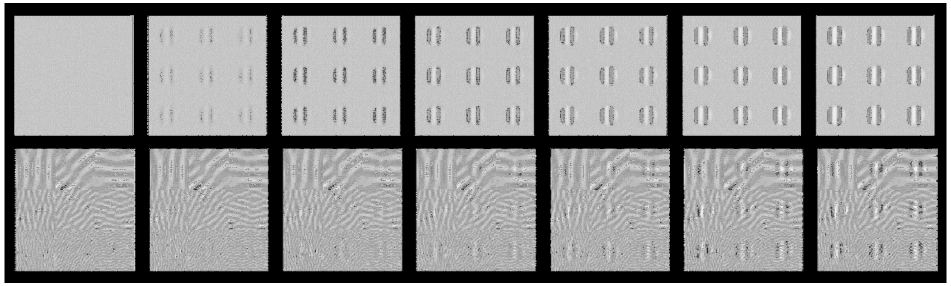
B. The depression in sensitivity due to adaptation at 7·1 c/deg. is plotted, with open circles, as relative threshold elevation against spatial frequency. The vertical difference between each point and the smooth curve in Text-fig. 6*A* is the ratio of sensitivity before and after adaptation. The relative threshold elevation is the anti-logarithm of this difference minus 1, so that no change in threshold would give a value of zero on the ordinate. The continuous curve is the function $[e^{-f^2}-e^{-(2f)^2}]^2$, fitted by eye to the data points. The filled arrows show the adapting frequency of 7·1 c/deg. The open arrow marks the value on the ordinate for a threshold elevation equivalent to $2\sqrt{2}$ times an average S.E. for determining contrast sensitivity.

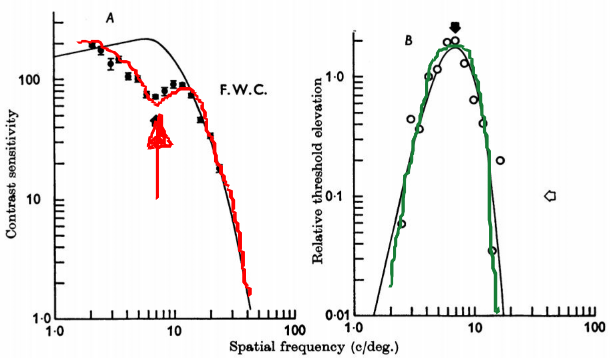
Blakemore & Campbell J. Physiol. 69

① ONE EXAMPLE:

Frequency sensors and non-linearities

Image-computable models

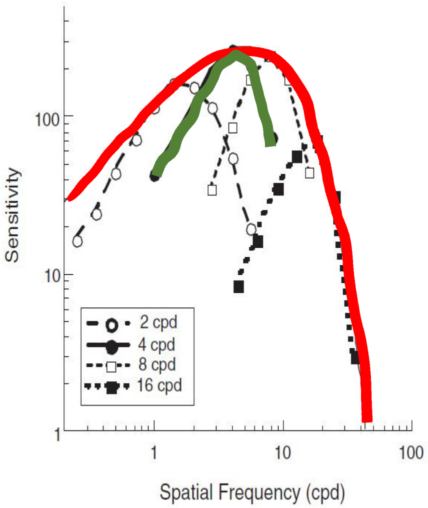


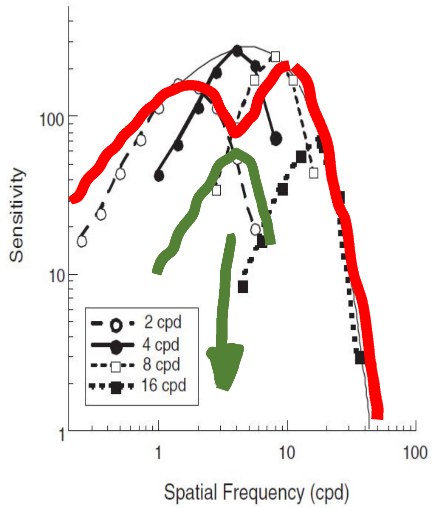


Text-fig. 6. The effect of adapting at 7·1 c/deg. *A*. The continuous curve from Text-fig. 5 is reproduced. The filled circles and vertical bars are the means and S.E. ($n = 6$) for re-determinations of contrast sensitivity at a number of spatial frequencies while F. W.C. was continuously adapting to a grating of 7·1 c/deg., 1·5 log. units above threshold. The exact procedure is described in the text.

B. The depression in sensitivity due to adaptation at 7·1 c/deg. is plotted, with open circles, as relative threshold elevation against spatial frequency. The vertical difference between each point and the smooth curve in Text-fig. 6*A* is the ratio of sensitivity before and after adaptation. The relative threshold elevation is the anti-logarithm of this difference minus 1, so that no change in threshold would give a value of zero on the ordinate. The continuous curve is the function $[e^{-f^2} - e^{-(2f)^2}]^2$, fitted by eye to the data points. The filled arrows show the adapting frequency of 7·1 c/deg. The open arrow marks the value on the ordinate for a threshold elevation equivalent to $2\sqrt{2}$ times an average S.E. for determining contrast sensitivity.

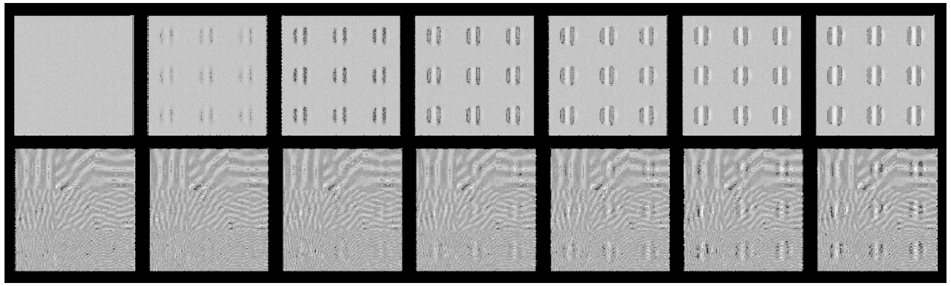
Blakemore & Campbell J. Physiol. 69



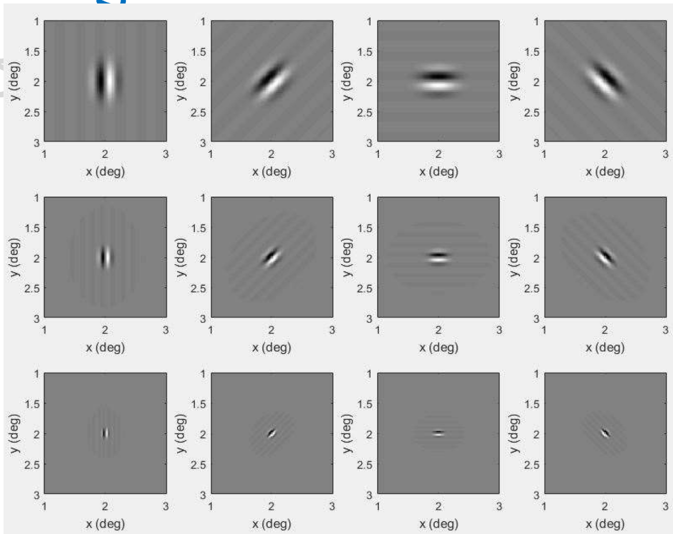


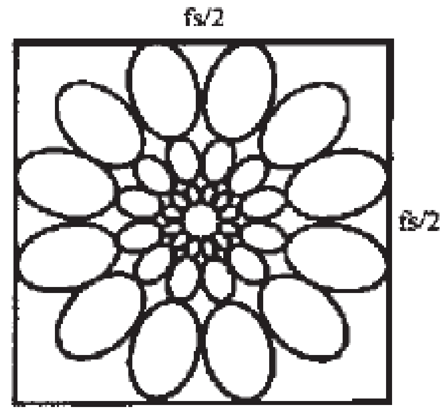
Sensitivity loss in mechanism

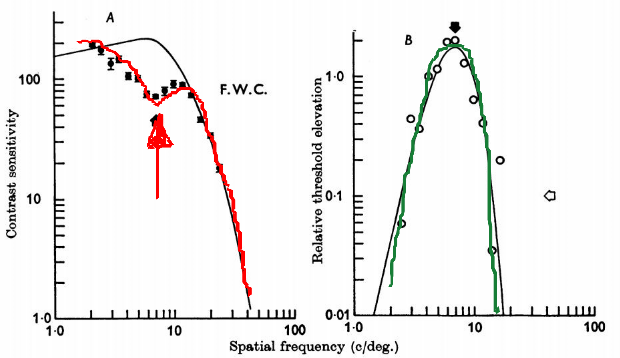
1 ONE EXAMPLE :



Frequency sensors and non-linearities



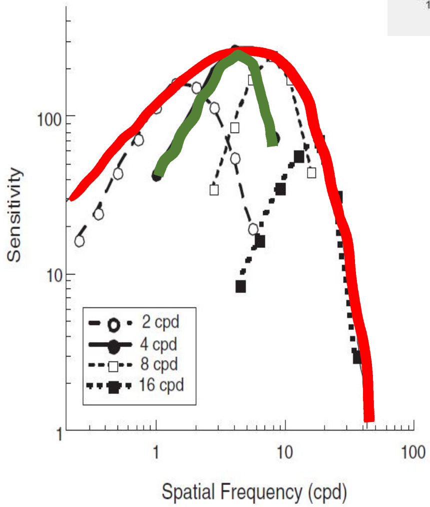


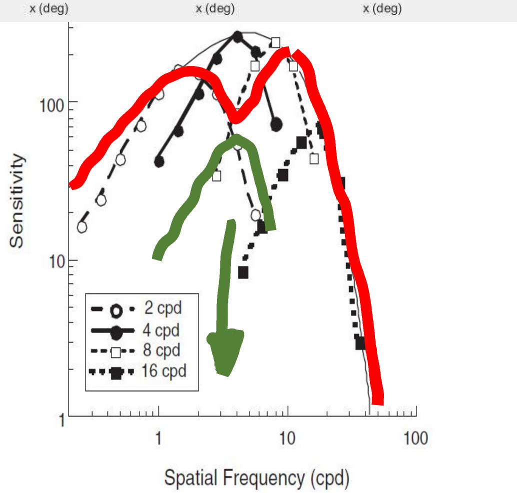


Text-fig. 6. The effect of adapting at 7·1 c/deg. *A*. The continuous curve from Text-fig. 5 is reproduced. The filled circles and vertical bars are the means and S.E. ($n = 6$) for re-determinations of contrast sensitivity at a number of spatial frequencies while F. W.C. was continuously adapting to a grating of 7·1 c/deg., 1·5 log. units above threshold. The exact procedure is described in the text.

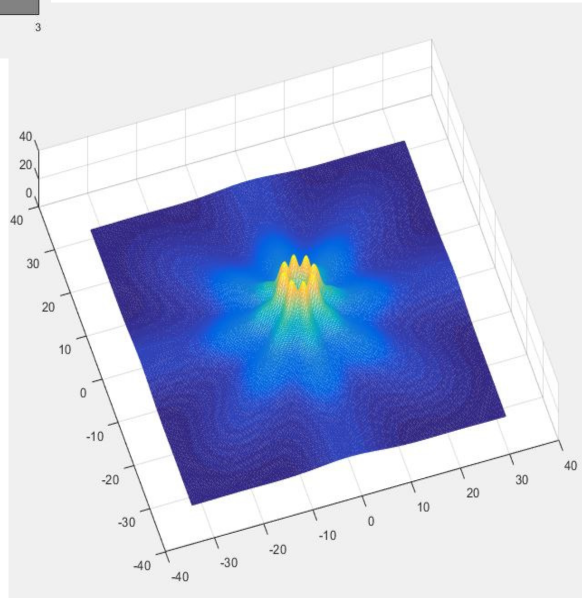
B. The depression in sensitivity due to adaptation at 7·1 c/deg. is plotted, with open circles, as relative threshold elevation against spatial frequency. The vertical difference between each point and the smooth curve in Text-fig. 6*A* is the ratio of sensitivity before and after adaptation. The relative threshold elevation is the anti-logarithm of this difference minus 1, so that no change in threshold would give a value of zero on the ordinate. The continuous curve is the function $[e^{-f^2}-e^{-(2f)^2}]^2$, fitted by eye to the data points. The filled arrows show the adapting frequency of 7·1 c/deg. The open arrow marks the value on the ordinate for a threshold elevation equivalent to $2\sqrt{2}$ times an average S.E. for determining contrast sensitivity.

Blakemore & Campbell J. Physiol. 69





Sensitivity loss in mechanism



① ONE EXAMPLE:

Frequency sensors and non-linearities

Image-computable models DIVISIVE NORMALIZATION

$x \xrightarrow{L} c \xrightarrow{N} y$

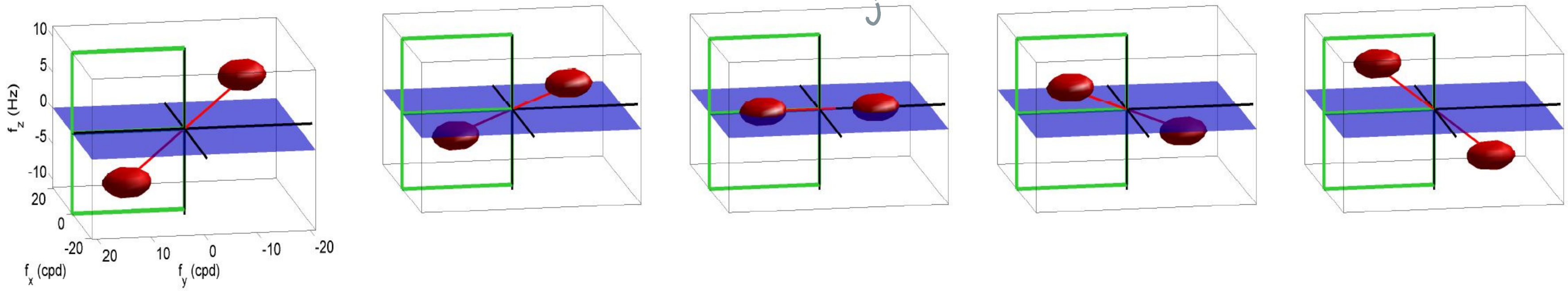
LINEAR $c = L \cdot x$

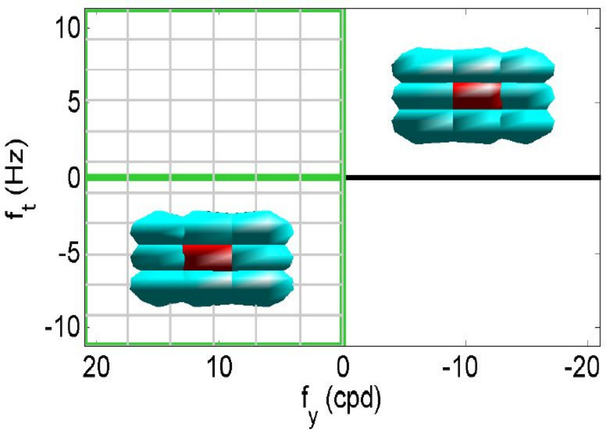
NON-LINEAR $y_i = \dfrac{c_i^{\gamma}}{b_i + \sum_j H_{ij} c_j^{\gamma}}$

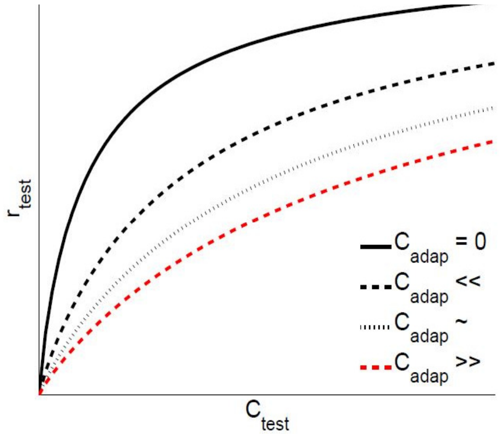
Heeger & Carandini 94

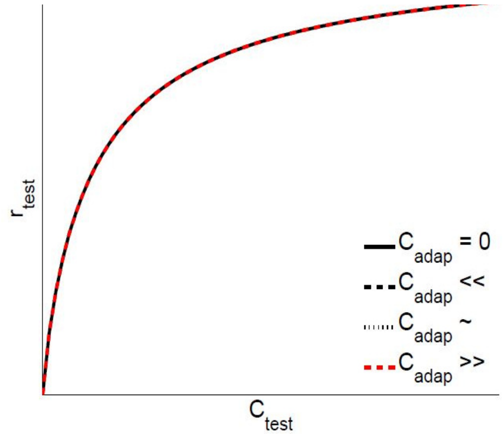
Carandini & Heeger 12

Simoncelli & Heeger 98







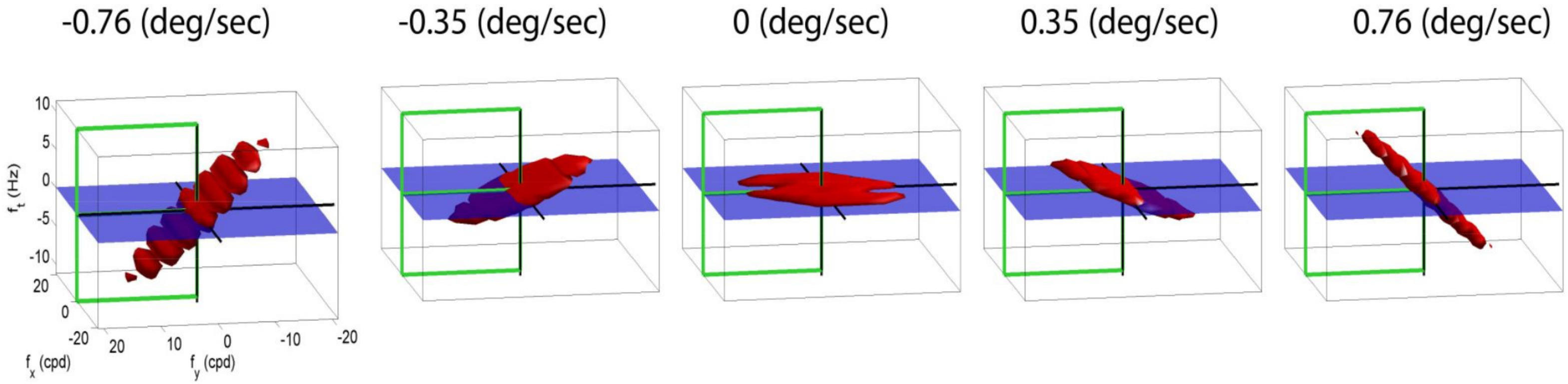


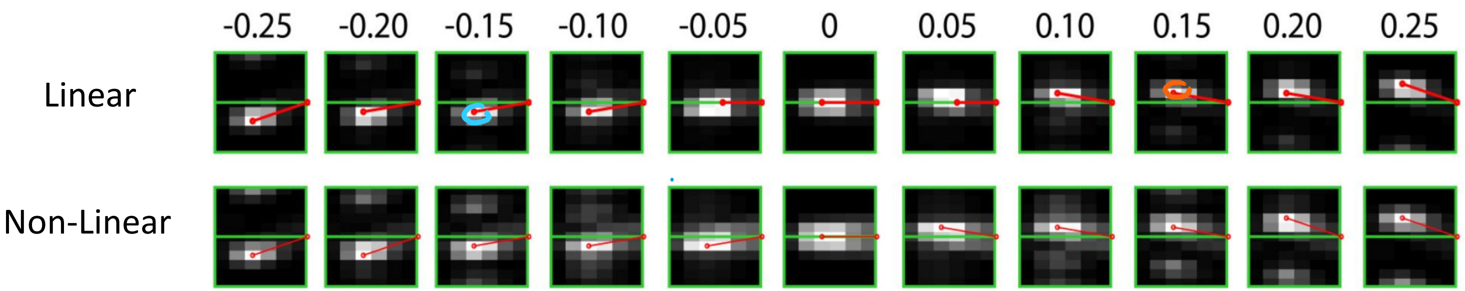
Laparra & Malo Front. Neurosc. 2015

1 ONE EXAMPLE:

Frequency sensors and non-linearities

Image-computable models DIVISIVE NORMALIZATION



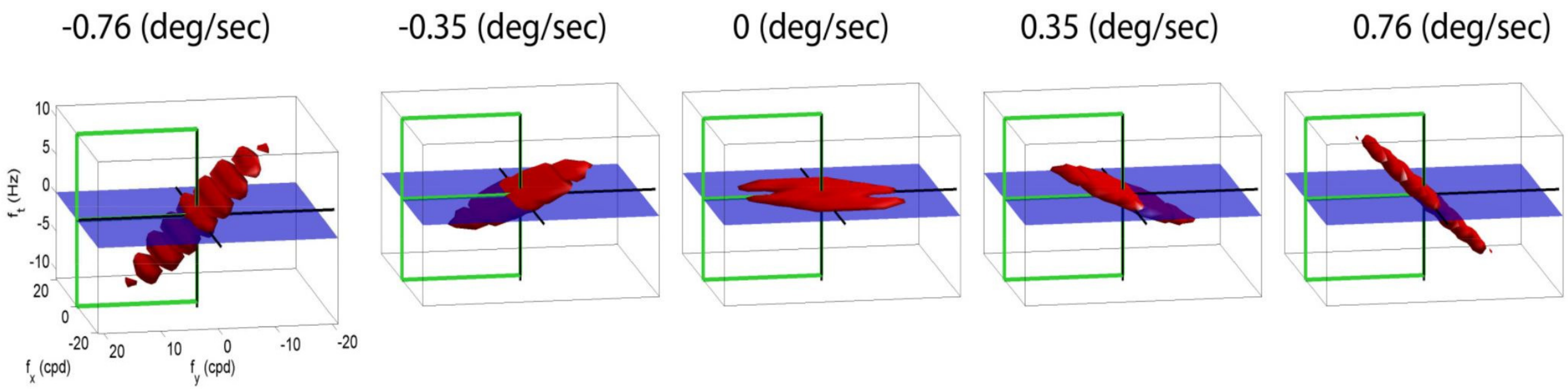


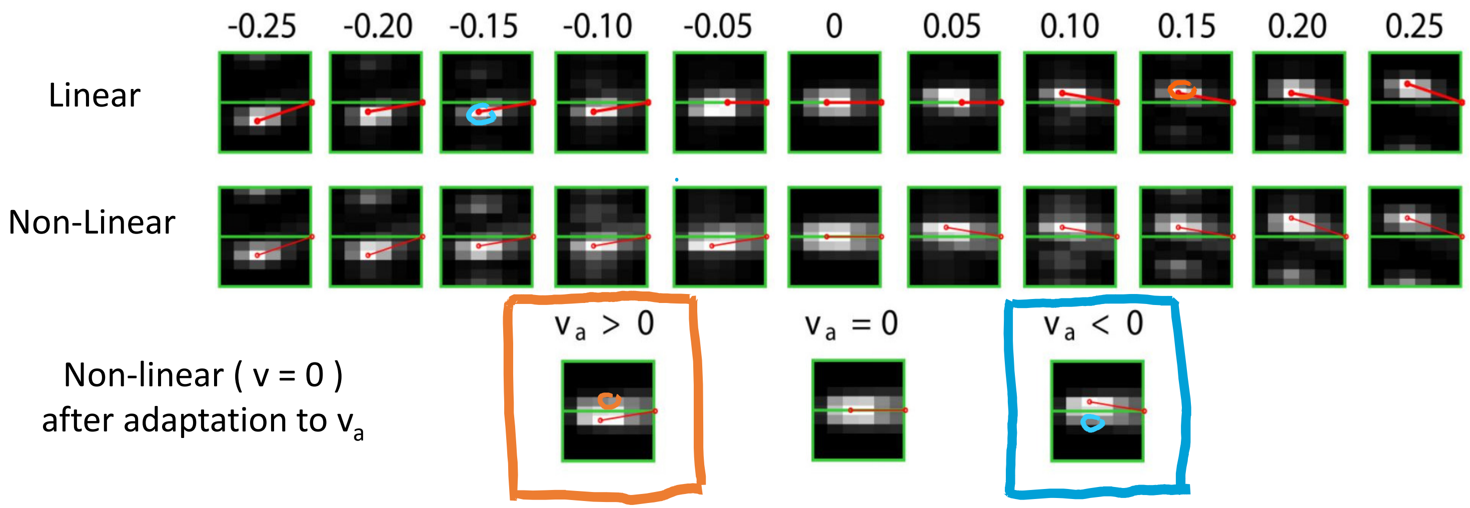
Laparra & Malo Front. Neurosc. 2015

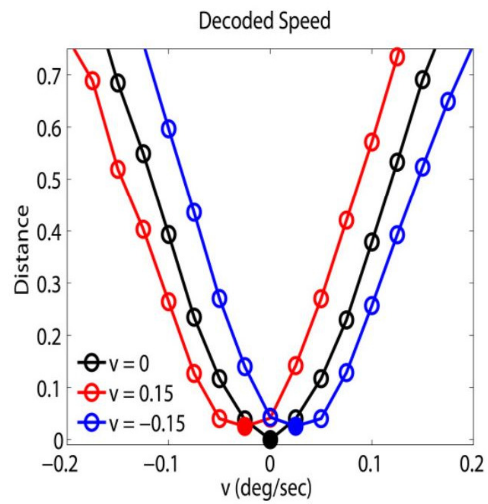
① ONE EXAMPLE:

Frequency sensors and non-linearities

Image-computable models DIVISIVE NORMALIZATION







Laparra & Malo Front. Neurosc. 2015

1 ONE EXAMPLE :

Frequency sensors and non-linearities

Image-computable models DIVISIVE NORMALIZATION

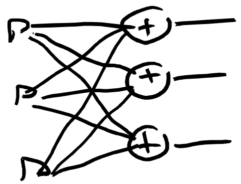
BIOL. Carandini & Heeger Nat. Rev. Neurosci 12
Schütt & Wichmann J. Vision 17

MATH. Martinez, Malo et al. PLOS ONE 18

https://isp.uv.es/code/visioncolor/vistamodels.html

$$\begin{tikzcd} x \arrow[r, "\mathcal{L}"'] \arrow[rr, bend left, "S"] & e \arrow[r, "\mathcal{N}"'] & y \end{tikzcd}$$

$\mathcal{L}$ $\quad e = W \cdot x$



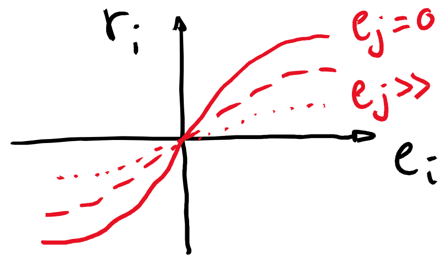
e = linear response

b = semisaturation

H = interaction kernel

K = constant → dyn. range

$\mathcal{N}$ $\quad y = K \cdot \dfrac{e}{b + H \cdot e}$



Masking and adaptation

$$\nabla_x S \sim \left[I - D_{r(x)} H \right] \cdot D_e \cdot W \Rightarrow M = \nabla_x S^T \nabla_x S$$

NON DIAGONAL!
INPUT DEPENDENT!

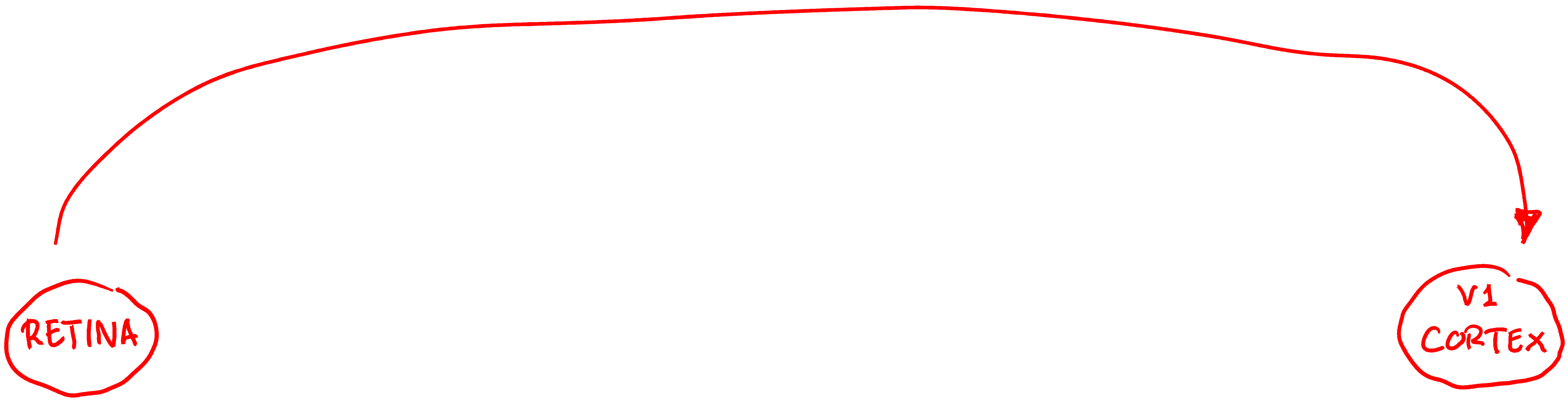
1 ONE EXAMPLE :

Frequency sensors and non-linearities

Image-computable models

BIOL. Carandini & Heeger Nat. Rev. Neurosci 12
Schütt & Wichmann J. Vision 17

MATH. Martinez, Malo et al. PLOS ONE 18
https://isp.uv.es/code/visioncolor/vistamodels.html



Gomez et al. J. Neurophysiol. 2020 (Wilson-Cowan)
Malo J. Math. Neurosci. 2020 (Divisive Normaliz.)

1 ONE EXAMPLE:

Frequency sensors and non-linearities

Image-computable models

BIOL. Carandini & Heeger Nat. Rev. Neurosci 12
Schütt & Wichmann J. Vision 17

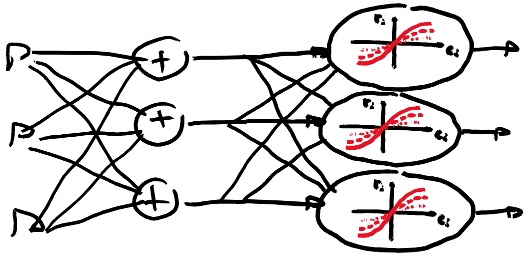
MATH. Martinez, Malo et al. PLOS ONE 18

https://isp.uv.es/code/visioncolor/vistamodels.html

$\mathcal{L}$ $\mathcal{N}$

Radiation LMS Von Kries

RETINA



V1 CORTEX

Chromatic Adaptation

Gomez et al. J. Neurophysiol. 2020 (Wilson-Cowan)

Malo J. Math. Neurosci. 2020 (Divisive Normaliz.)

1 ONE EXAMPLE :

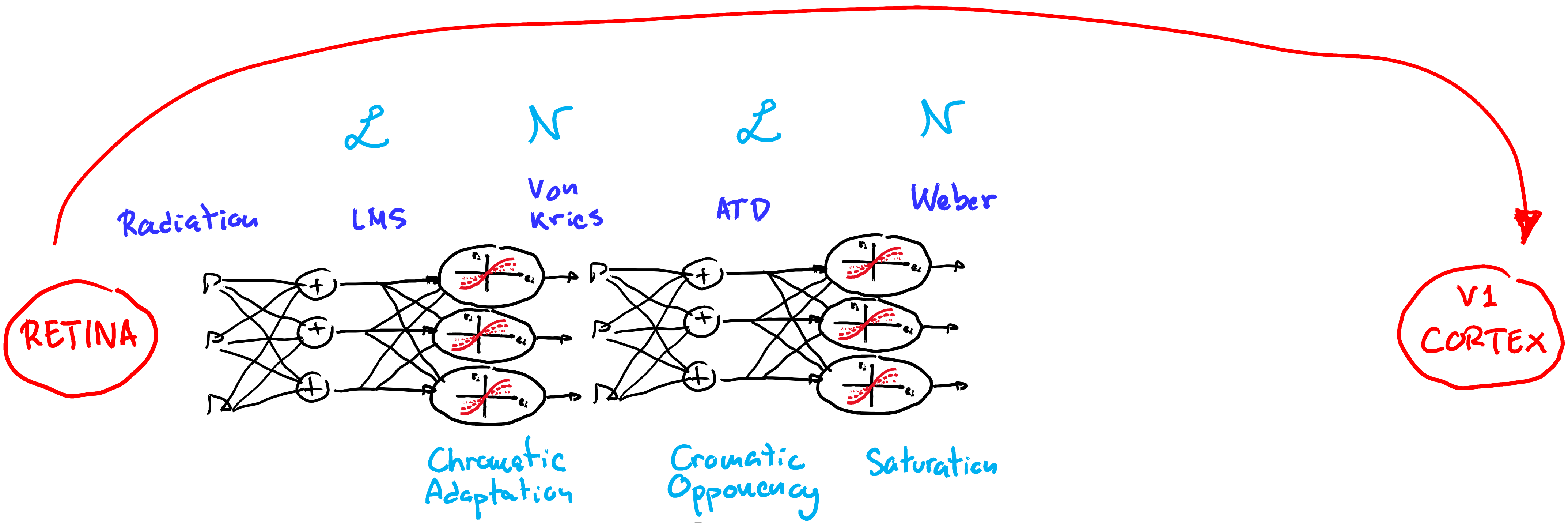
Frequency sensors and non-linearities

Image-computable models

BIOL. Carandini & Heeger Nat. Rev. Neurosci 12
Schütt & Wichmann J.Vision 17

MATH. Martinez, Malo et al. PLOS ONE 18

https://isp.uv.es/code/visioncolor/vistamodels.html



Gomez et al. J. Neurophysiol. 2020 (Wilson-Cowan)
Malo J. Math. Neurosci. 2020 (Divisive Normaliz.)

1 ONE EXAMPLE :

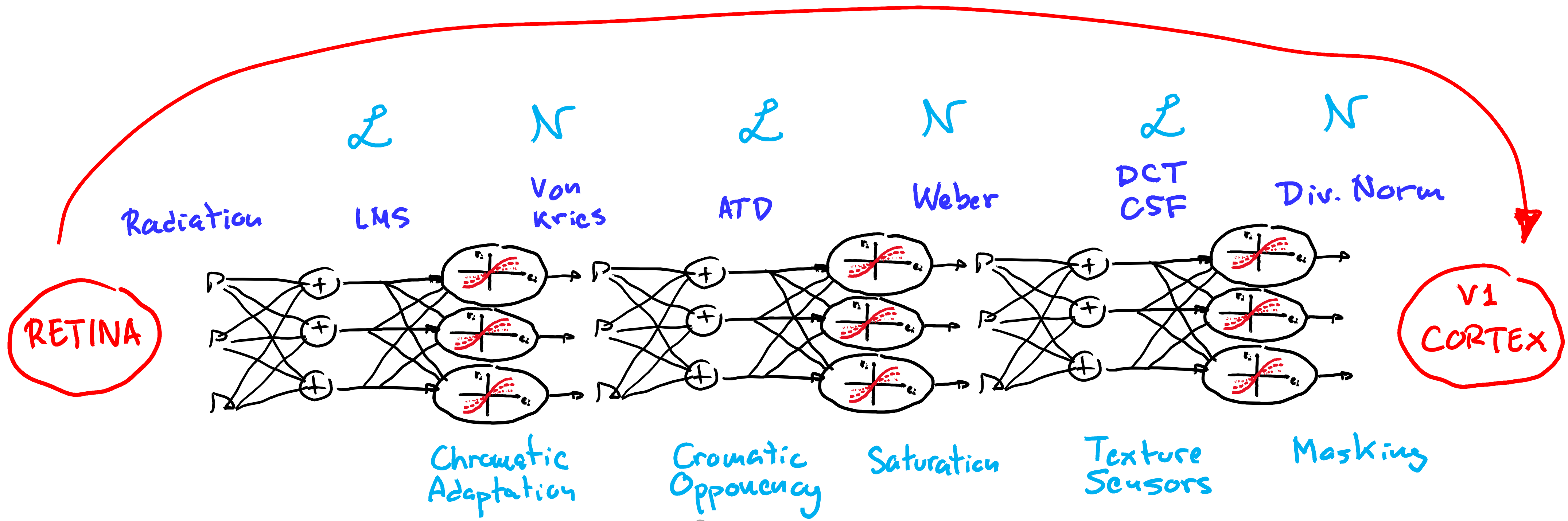
Frequency sensors and non-linearities

Image-computable models

BIOL. Carandini & Heeger Nat. Rev. Neurosci 12
Schütt & Wichmann J. Vision 17

MATH. Martinez, Malo et al. PLOS ONE 18

https://isp.uv.es/code/visioncolor/vistamodels.html



Gomez et al. J. Neurophysiol. 2020 (Wilson-Cowan)
Malo J. Math. Neurosci. 2020 (Divisive Normaliz.)

1 ONE EXAMPLE:

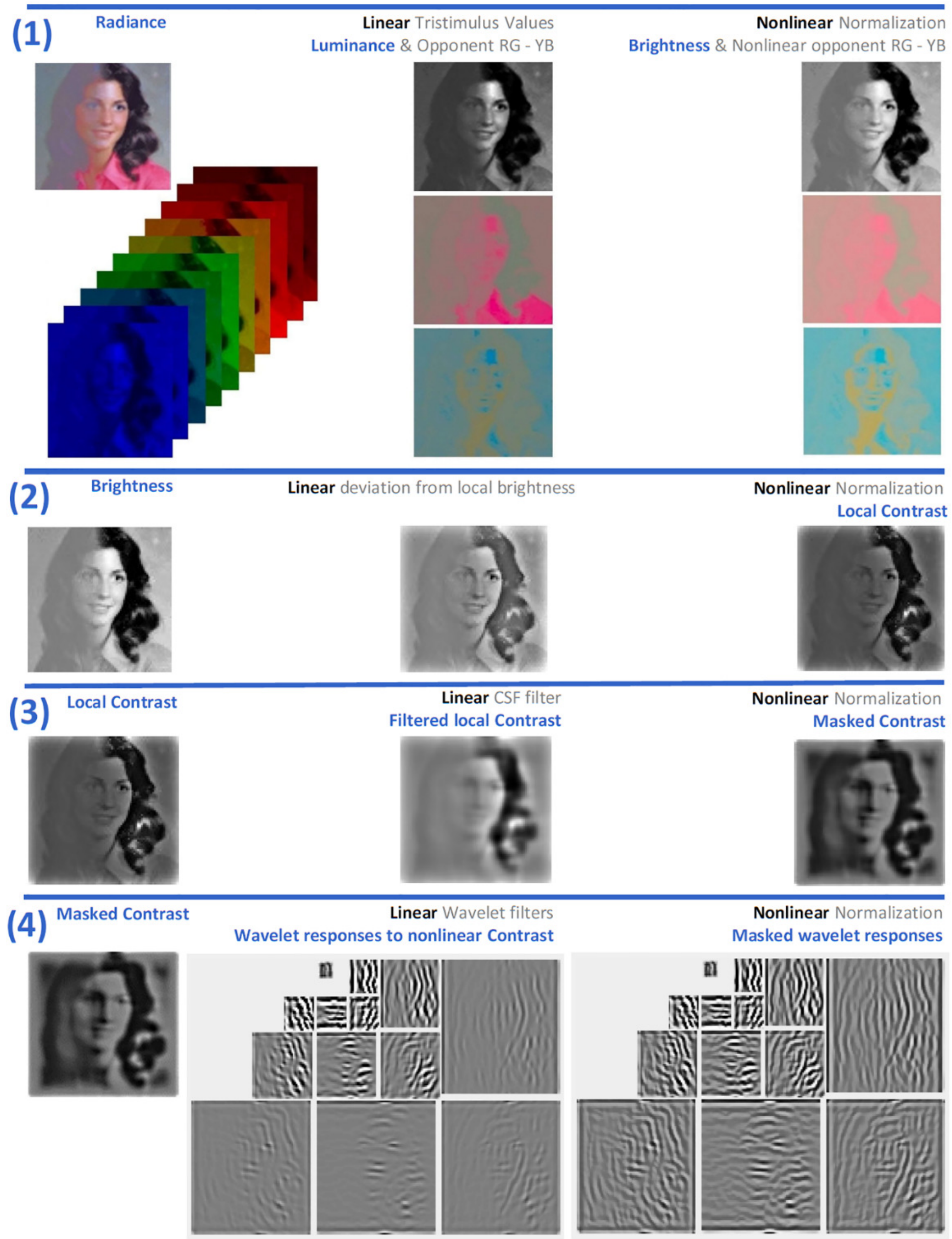
$$\begin{tikzcd} \text{Radiance} \arrow[r, "S^{(1)}"'] \arrow[rrrr, bend left, "S(x^0,\Theta)"] & \text{Brightness} \arrow[r, "S^{(2)}"'] & \text{Local Contrast} \arrow[r, "S^{(3)}"'] & \text{CSF filter+Spat.Masking} \arrow[r, "S^{(4)}"'] & \text{Wavelet filters+Freq.Masking} \end{tikzcd}$$

$\mathcal{L}$ $\mathcal{N}$ $\quad$ $\mathcal{L}$ $\mathcal{N}$ $\quad$ $\mathcal{L}$ $\mathcal{N}$ $\quad$ $\mathcal{L}$ $\mathcal{N}$

https://isp.uv.es/code/visioncolor/vistamodels.html

Martinez, Malo et al. PLOS ONE 18

- Derivatives { · Metrics
  · New Psychophysics
- Inverse Decoding. Stimuli



① ONE EXAMPLE:

WHY?

$$S(x^0, \Theta)$$

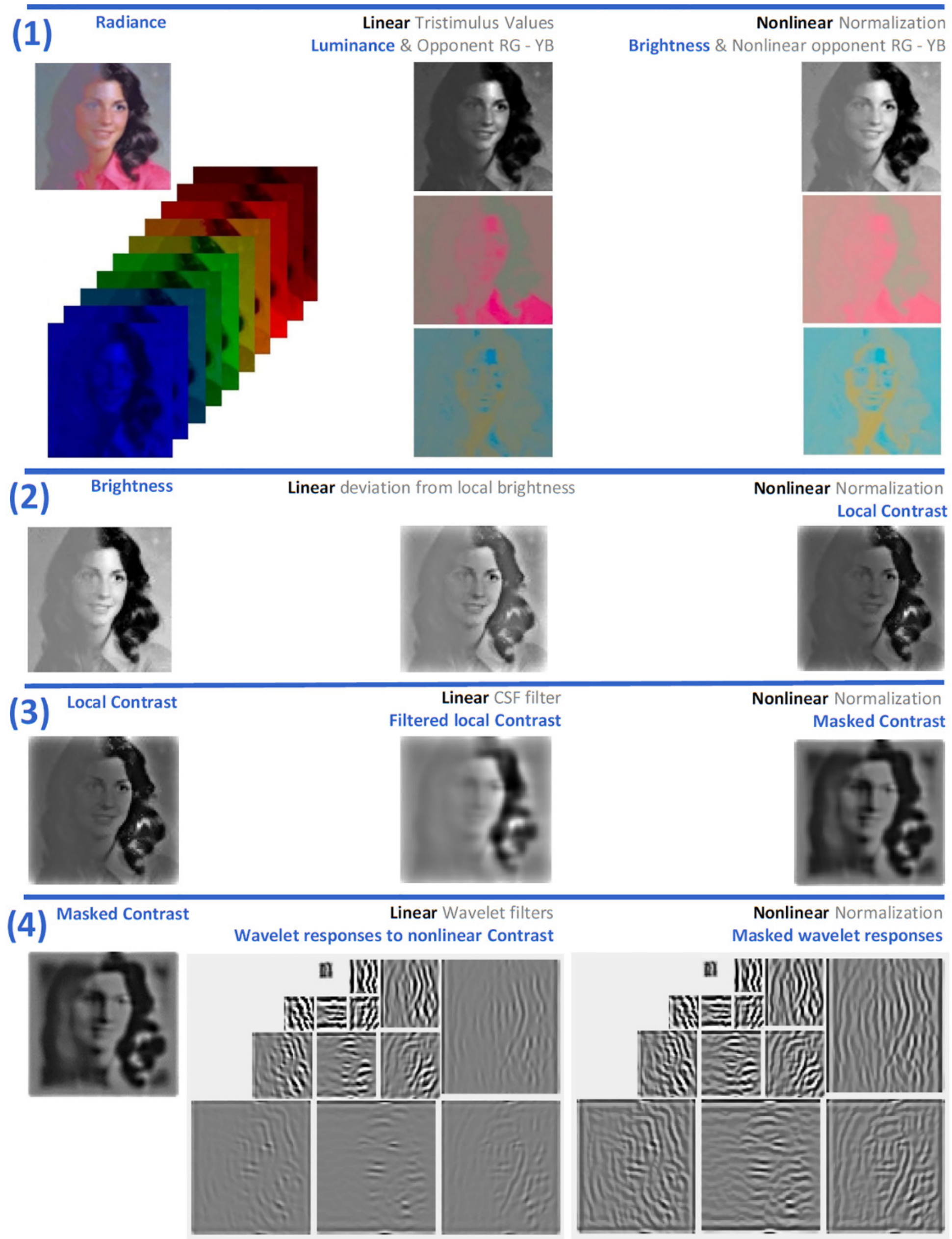
Radiance $\xrightarrow{S^{(1)}}$ Brightness $\xrightarrow{S^{(2)}}$ Local Contrast $\xrightarrow{S^{(3)}}$ CSF filter+Spat.Masking $\xrightarrow{S^{(4)}}$ Wavelet filters+Freq.Masking

$\mathcal{L}$ $\mathcal{N}$ $\quad$ $\mathcal{L}$ $\mathcal{N}$ $\quad$ $\mathcal{L}$ $\mathcal{N}$ $\quad$ $\mathcal{L}$ $\mathcal{N}$

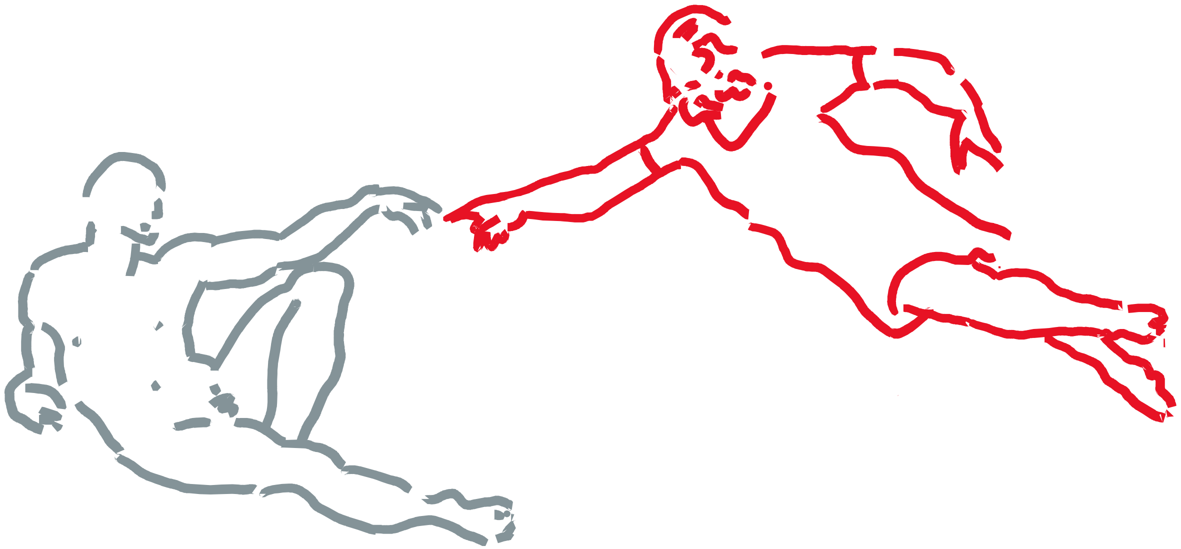
https://isp.uv.es/code/visioncolor/vistamodels.html

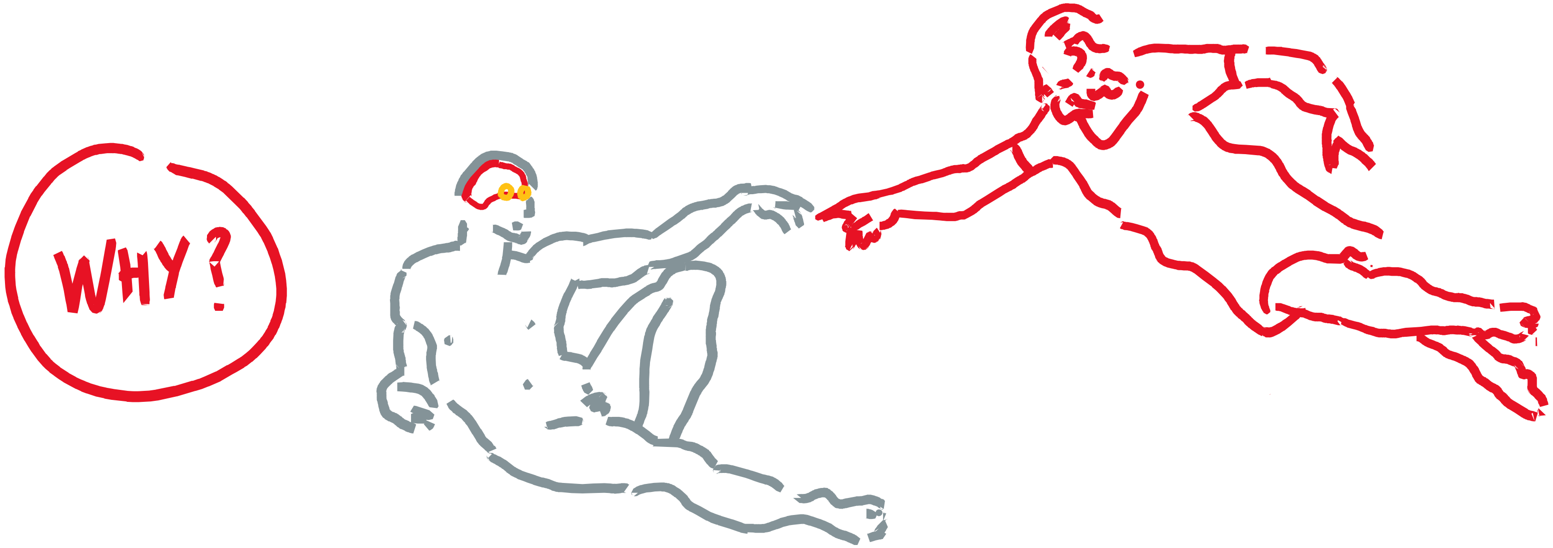
Martinez, Malo et al. PLOS ONE 18

- Derivatives { · Metrics
  · New Psychophysics
- Inverse Decoding - Stimuli

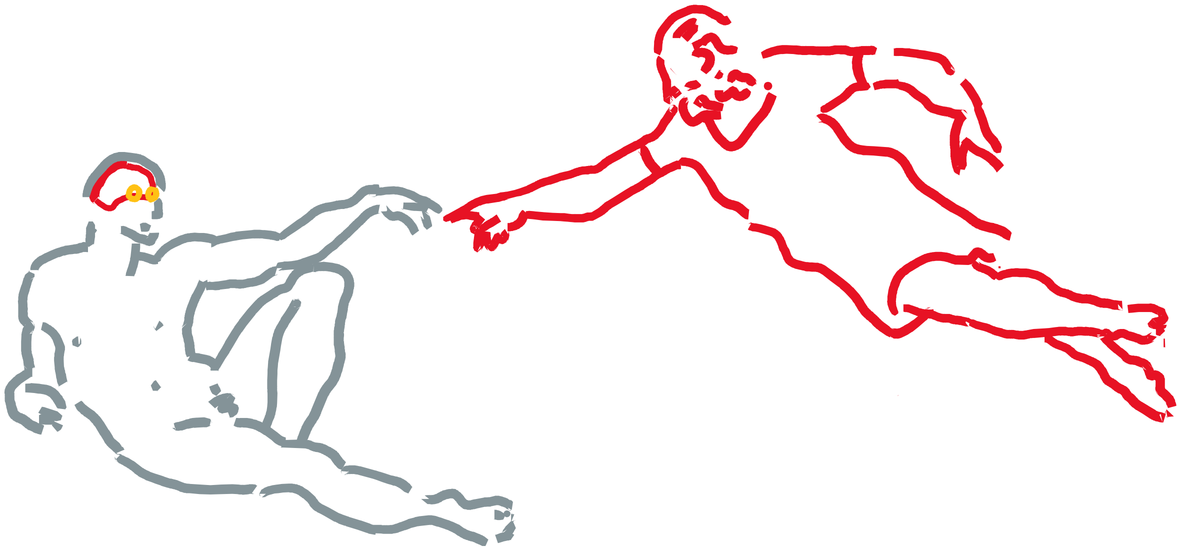


WHY?

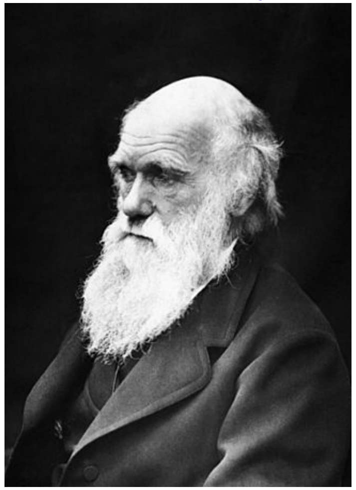



WHY?

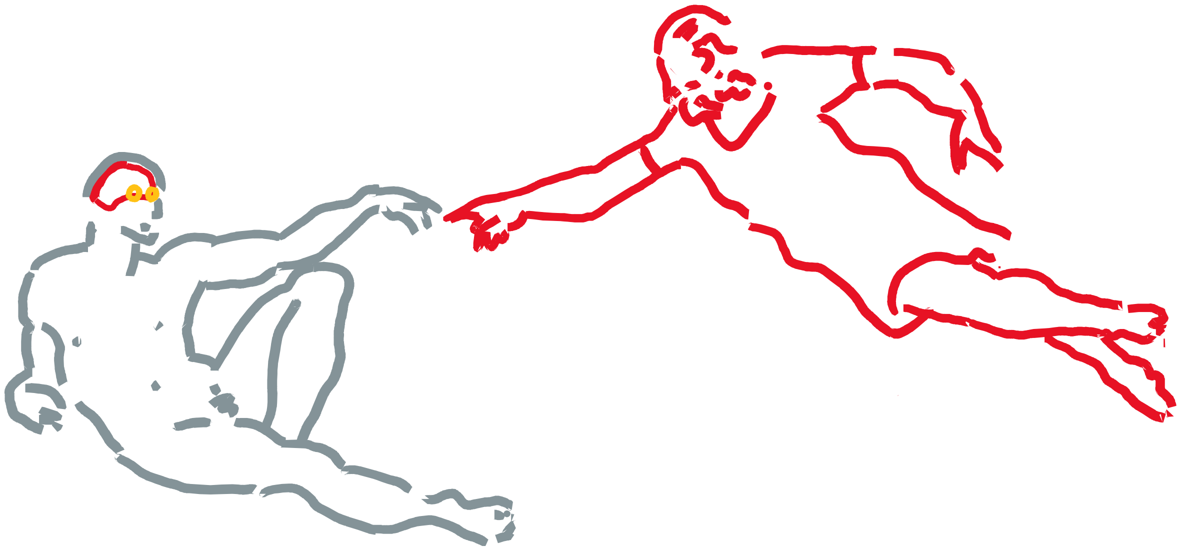
WHY?



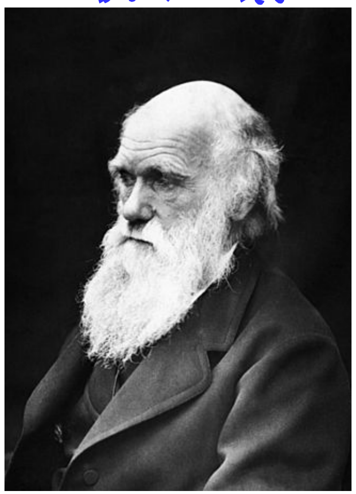
DARWIN

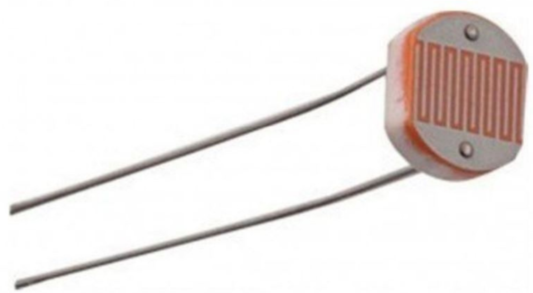


WHY?



DARWIN





SENSOR FOTORESISTENCIA LDR GL-5528

Fotoresistencia que permite medir niveles de luz.

0,25 €

0,21 € (IVA no incluido)

Fracciona tu pago desde 50,00 €

SEQURA

Estado: Nuevo

Fabricante: tiendatec

Referencia: GL5528

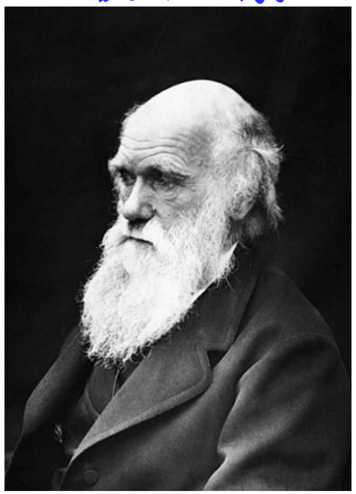
EAN: 8472496014380

Disponible, recíbelo el lunes 4

!!

WHAT WOULD YOU DO ?

DARWIN



SENSOR FOTORESISTENCIA LDR GL-5528

Fotoresistencia que permite medir niveles de luz.

0,25 €

0,21 € (IVA no incluido)

Fracciona tu pago desde 50,00 €

SEQURA

Estado: Nuevo

Fabricante: tiendatec

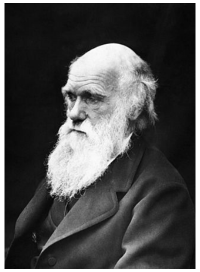
Referencia: GL5528

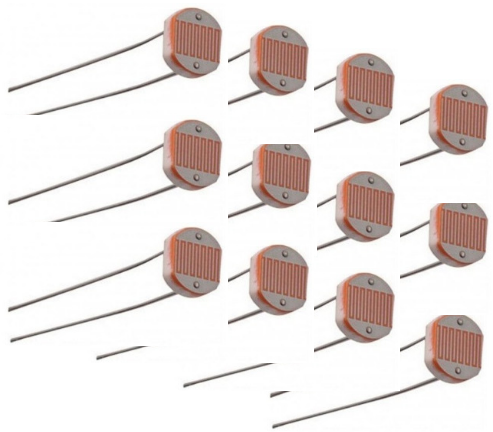
EAN: 8472496014380

Disponible, recíbelo el lunes 4

!! WHAT WOULD YOU DO ?

DARWIN

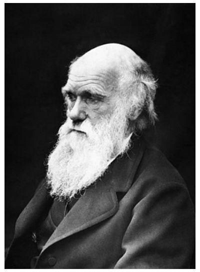


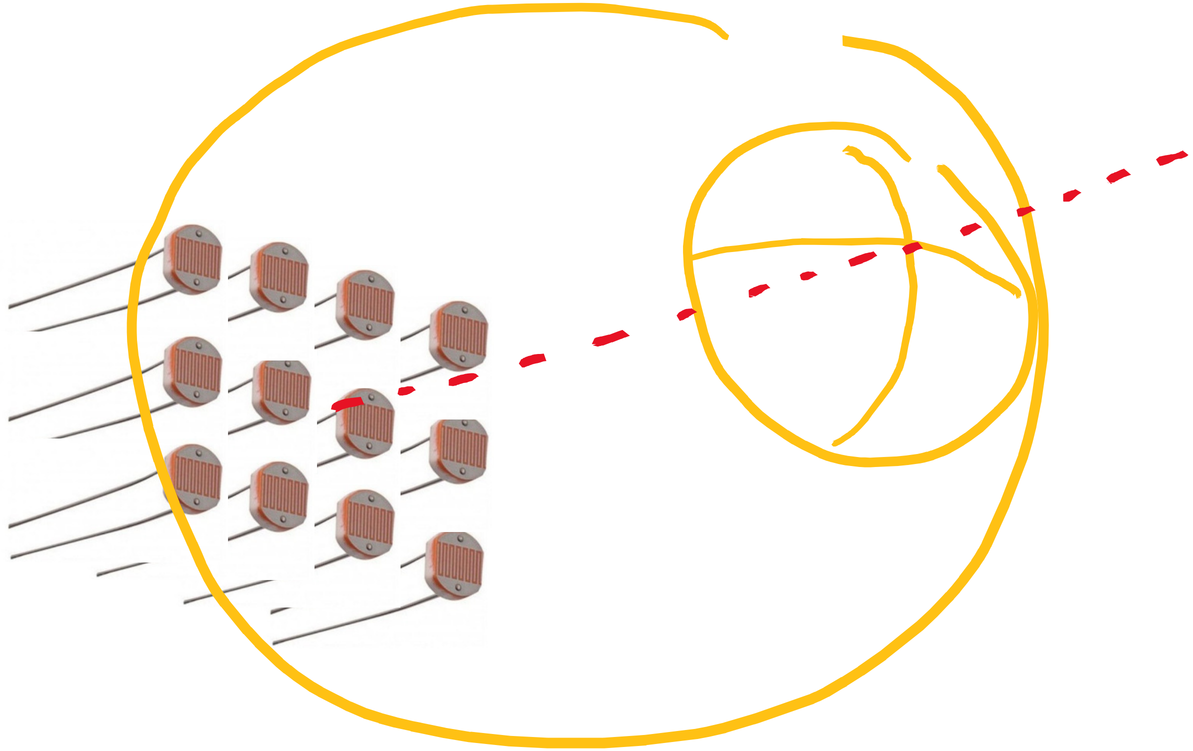


WHAT WOULD YOU DO ?

!!

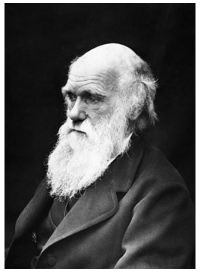
DARWIN

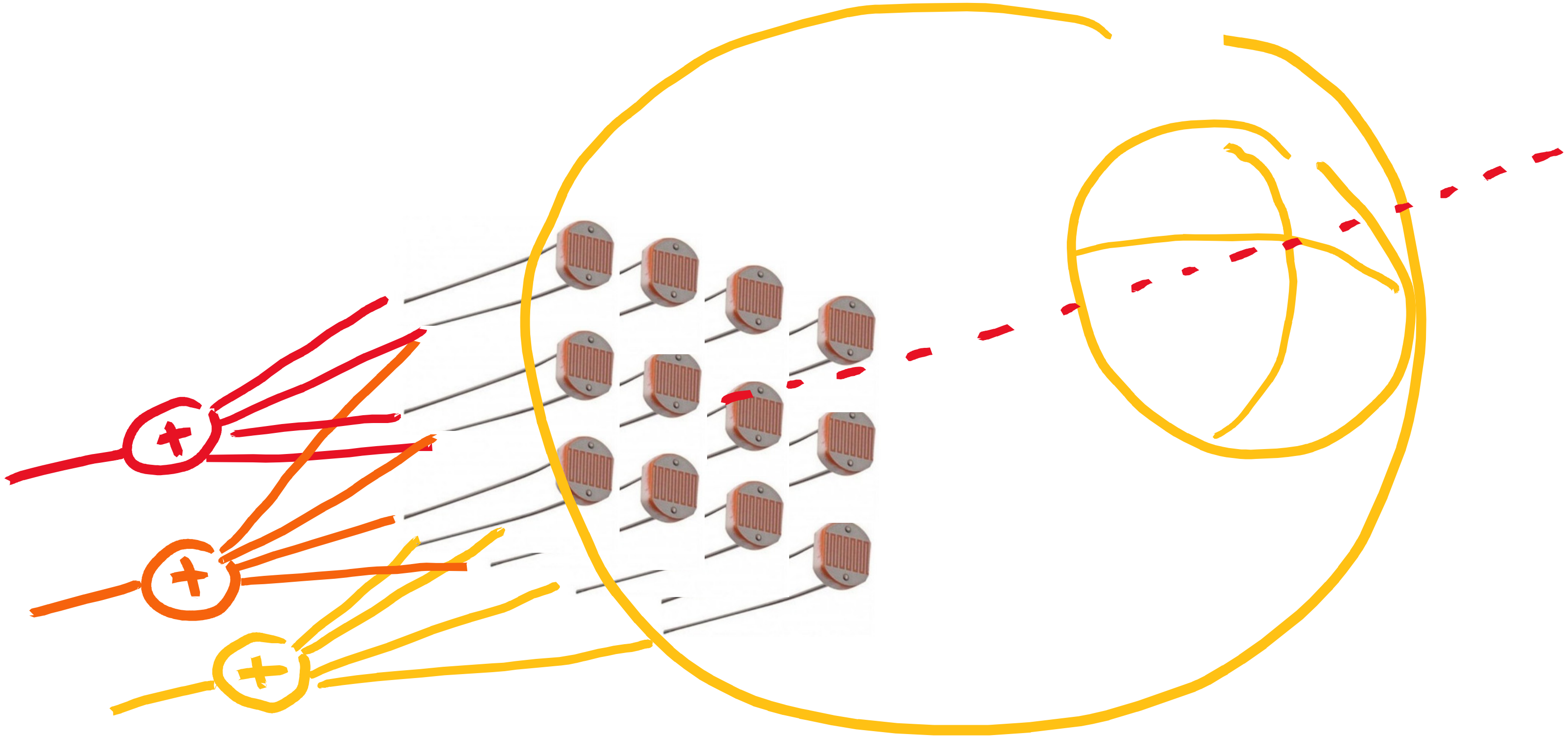




WHAT WOULD YOU DO ?

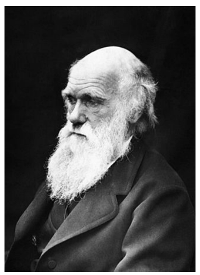
DARWIN

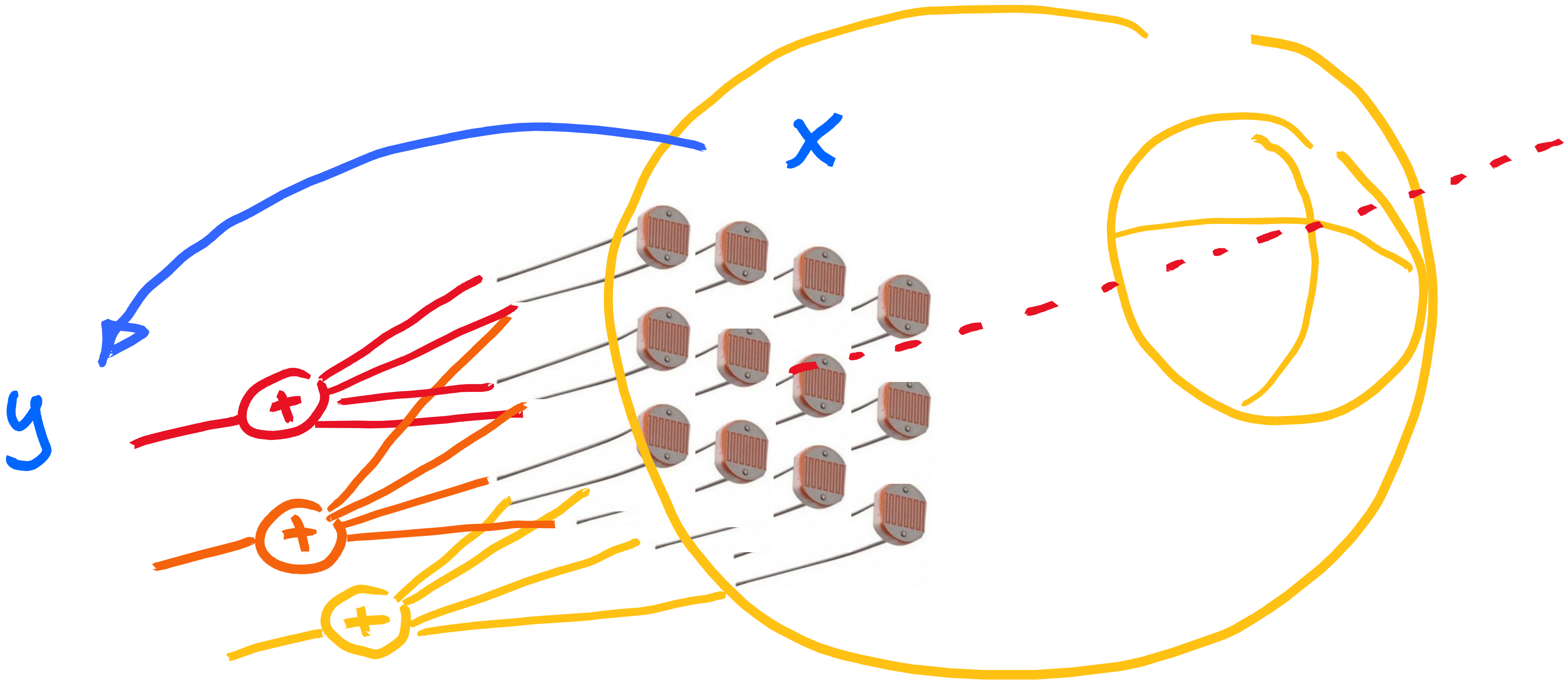




WHAT WOULD YOU DO ?

DARWIN

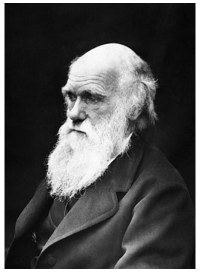


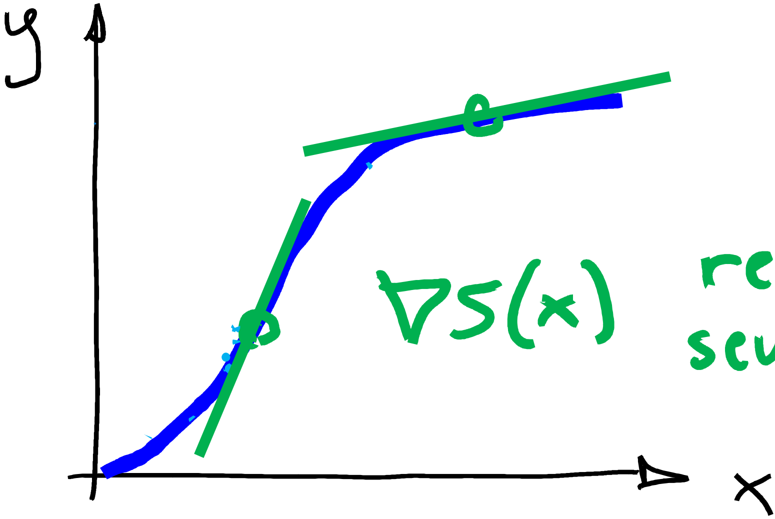


WHAT WOULD YOU DO?

$$x \xrightarrow{S} y = S(x) + n$$

DARWIN



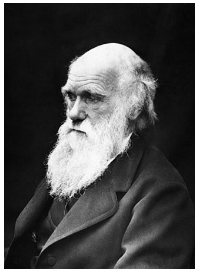


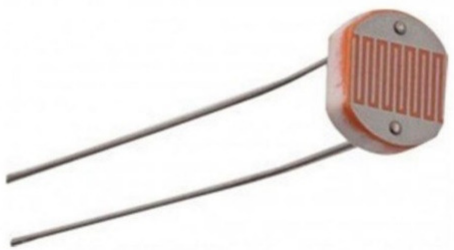
$\nabla S(x)$ response sensitivity

WHAT WOULD YOU DO ?

$$x \xrightarrow{S} y = S(x) + n$$

DARWIN





SENSOR FOTORESISTENCIA LDR GL-5528

Fotoresistencia que permite medir niveles de luz.

0,25 €

0,21 € (IVA no incluido)

Fracciona tu pago desde 50,00 €

SEQURA

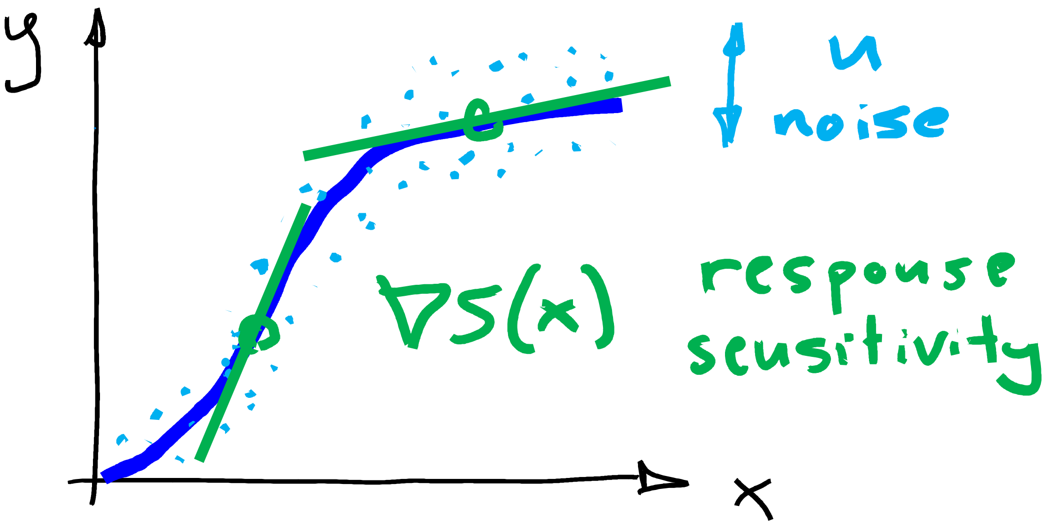
Estado: Nuevo

Fabricante: tiendatec

Referencia: GL5528

EAN: 8472496014380

Disponible, recíbelo el lunes 4

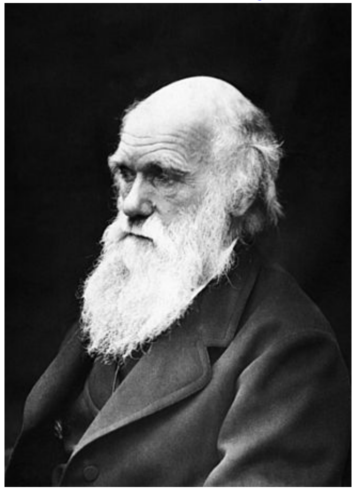


2 COMPUTATIONAL TOOLS :

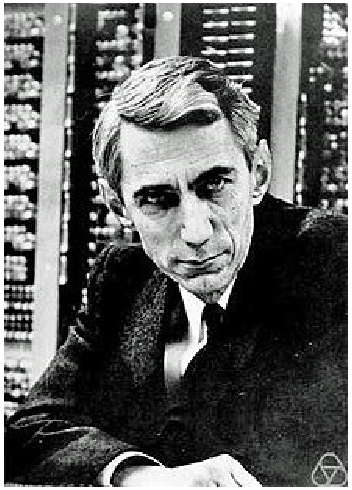
Information theory
Uniformization & Gaussianization

- Function determines structure!!
- One example: Information transmission
  - Entropy maximization
  - Redundancy reduction
- Original tools :
  - UNIFORMIZATION
  - GAUSSIANIZATION

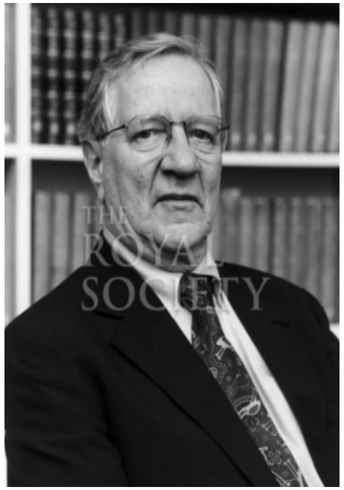
DARWIN



SHANNON



BARLOW

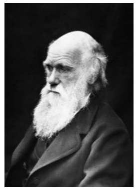


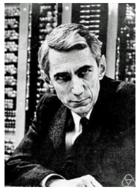
2 COMPUTATIONAL TOOLS : Information theory

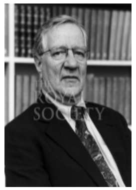
$$x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \xrightarrow{S_\theta} y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$$

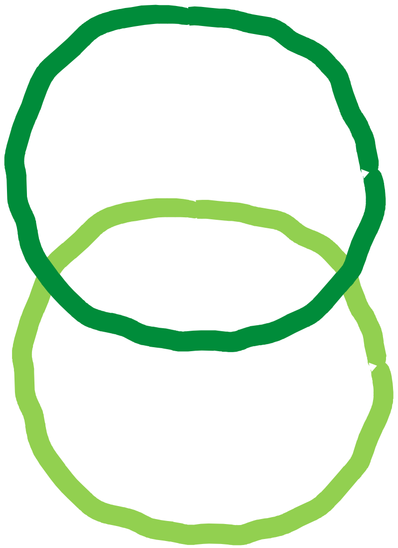
information $\propto \log\left(\frac{1}{P(x)}\right)$

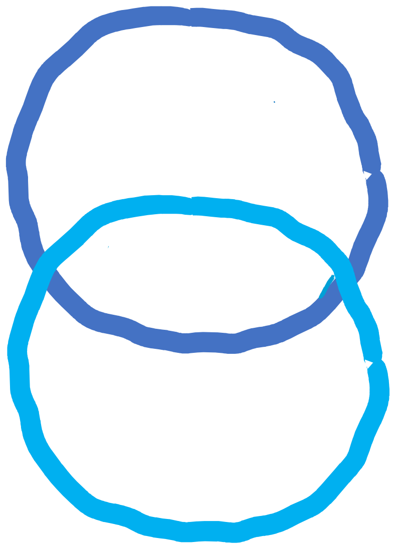
Entropy = $\langle$inform$\rangle$



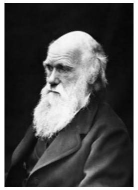


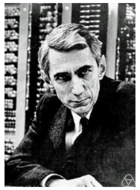


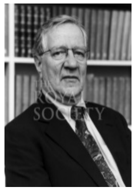




2 COMPUTATIONAL TOOLS : Information theory





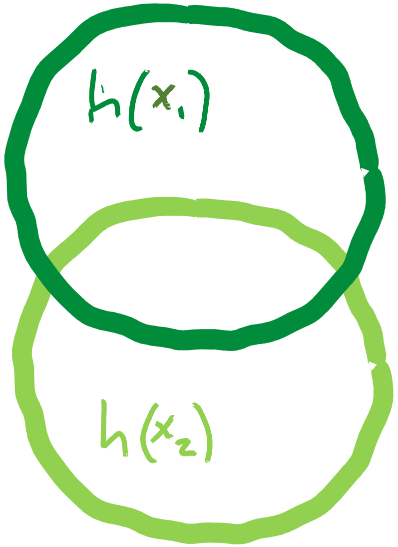


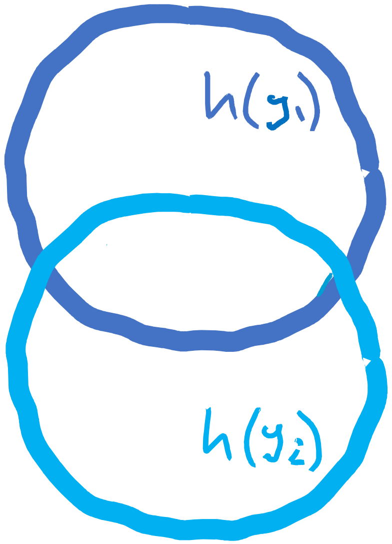
$$x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \xrightarrow{S_\theta} y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$$

information $\propto \log\left(\frac{1}{p(x)}\right)$

Entropy $= \langle \text{inform} \rangle$

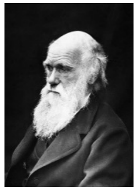
$$h(x) = -\int p(x) \log(p(x))\, dx$$

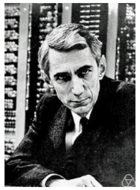


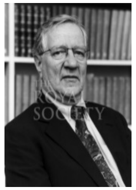


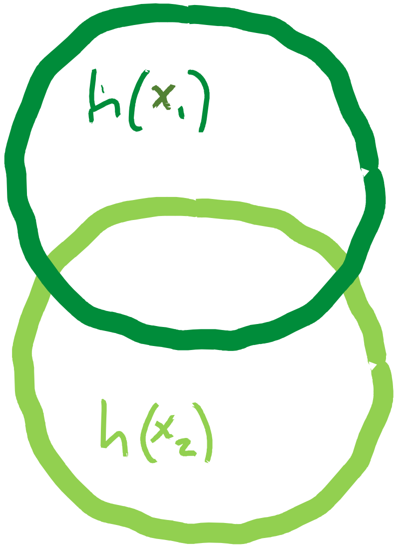
2 COMPUTATIONAL TOOLS : Information theory

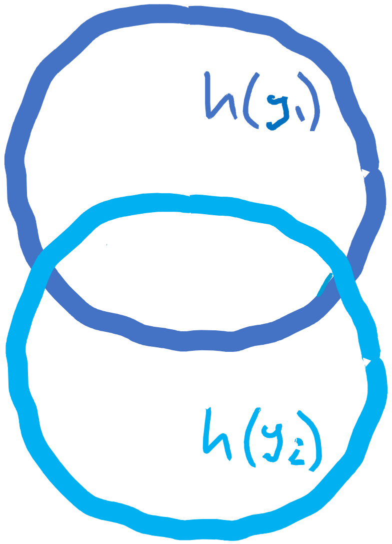
$$x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \xrightarrow{S_\theta} y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$$











$$\text{information} \propto \log\left(\frac{1}{p(x)}\right)$$

$$\text{Entropy} = \langle \text{inform} \rangle$$

$$h(x) = -\int p(x) \log(p(x))\, dx$$

$$h(y_1) + h(y_2) > h([y_1, y_2])$$

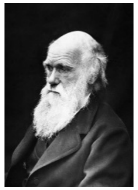
② COMPUTATIONAL TOOLS : Information theory

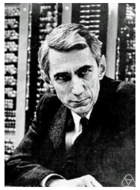
$$x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \xrightarrow{S_\theta} y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$$

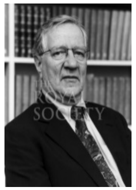
information $\propto \log\left(\frac{1}{p(x)}\right)$

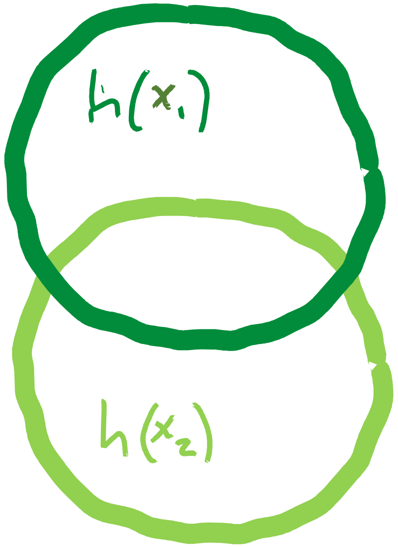
Entropy = $\langle$inform$\rangle$

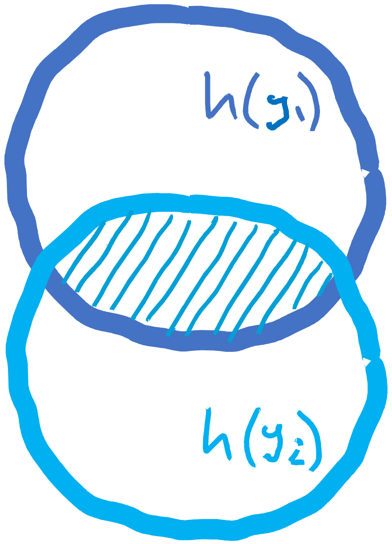
$$h(x) = -\int p(x) \log(p(x))\, dx$$









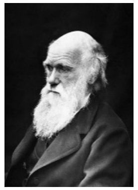


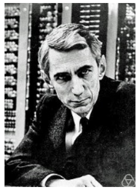
$$h(y_1) + h(y_2) > h([y_1, y_2])$$

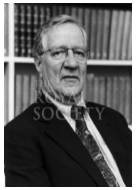
/// TOTAL CORRELATION ≡ Redundancy within a vector $T(y) = \sum_i h(y_i) - h(y)$

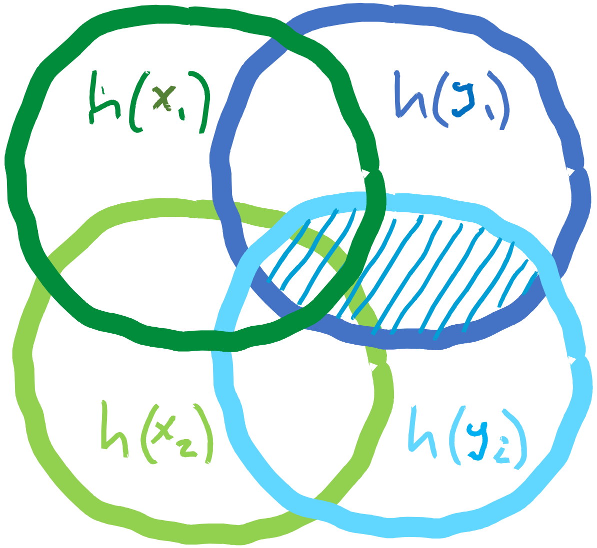
② COMPUTATIONAL TOOLS : Information theory

$$x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \xrightarrow{S_\theta} y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$$









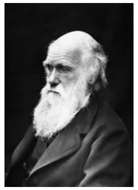
$$\text{information} \propto \log\left(\frac{1}{p(x)}\right)$$

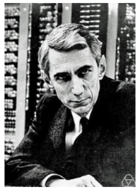
$$\text{Entropy} = \langle \text{inform} \rangle$$

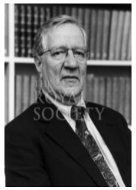
$$h(x) = -\int p(x) \log(p(x))\, dx$$

/// TOTAL CORRELATION ≡ Redundancy within a vector $T(y) = \sum_i h(y_i) - h(y)$

2 COMPUTATIONAL TOOLS : Information theory





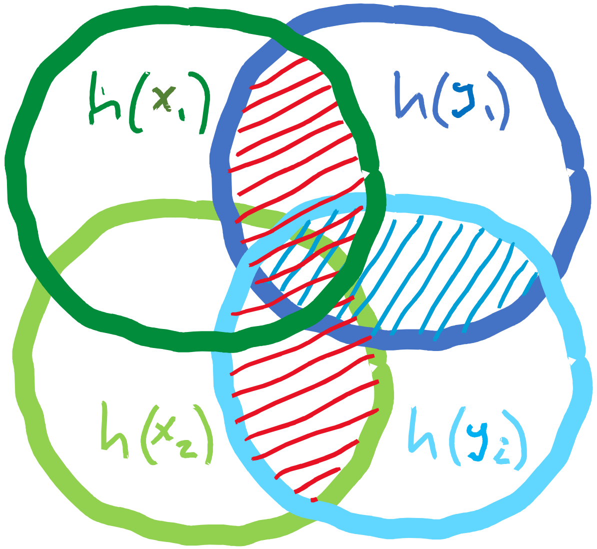


$$x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \xrightarrow{S_\theta} y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$$

$$\text{information} \propto \log\left(\frac{1}{p(x)}\right)$$

$$\text{Entropy} = \langle \text{inform} \rangle$$

$$h(x) = -\int p(x) \log(p(x))\, dx$$

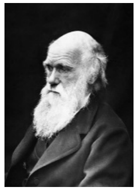


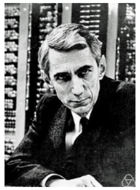
/// TOTAL CORRELATION ≡ Redundancy within a vector $T(y) = \sum_i h(y_i) - h(y)$

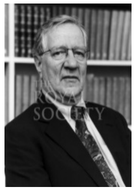
(2) COMPUTATIONAL TOOLS : Information theory

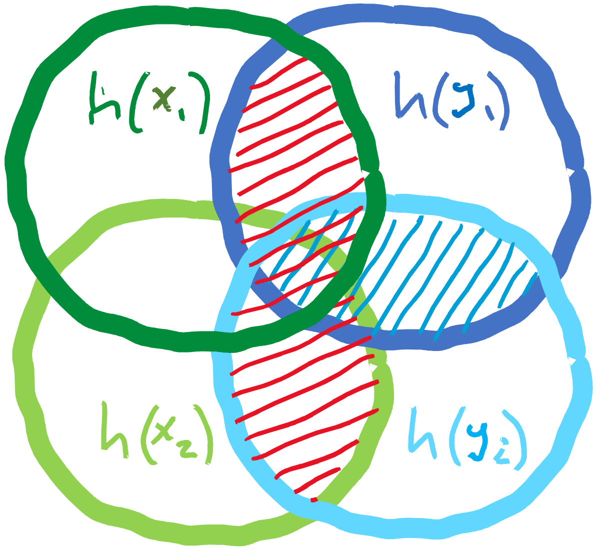
$$x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \xrightarrow{S_\theta} y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$$

$\text{information} \propto \log\left(\frac{1}{p(x)}\right)$









$\text{Entropy} = \langle \text{inform} \rangle$

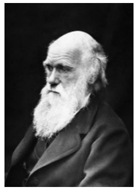
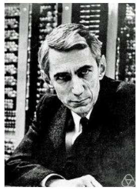
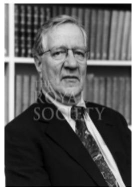
$h(x) = -\int p(x) \log(p(x))\, dx$

TOTAL CORRELATION ≡ Redundancy within a vector $T(y) = \sum_i h(y_i) - h(y)$

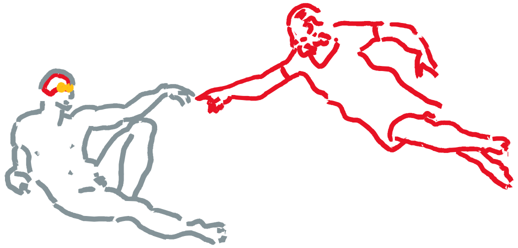
MUTUAL INFORMATION ≡ Info shared by two vectors $I(x,y) = h(x) + h(y) - h([x,y])$

(2) COMPUTATIONAL TOOLS : Information theory

The "design" problem:

$$x \xrightarrow{S} y = S(x) + n$$



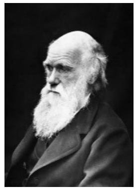
please maximize $I(x, y)$!

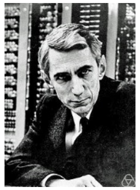
J. Malo (2020) **J. Math. Neuroscience** Spatio-chromatic information available from different neural layers

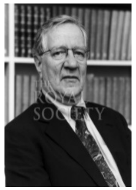
(2) COMPUTATIONAL TOOLS : Information theory

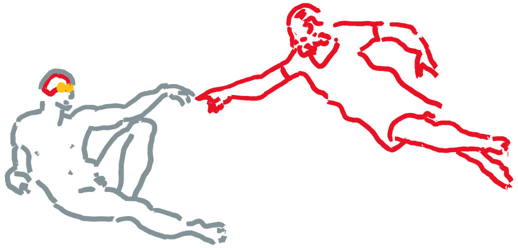
The "design" problem:

$$x \xrightarrow{S} y = S(x) + n$$









please maximize $I(x,y)$!

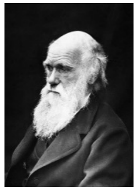
$$I(x,y) = h(x) + E_x\left[\log \|\nabla S\|\right] - \left(h(n) - E_n\left[D_{KL}\left(p(S(x)) \,|\, p(S(x)+n)\right)\right]\right)$$

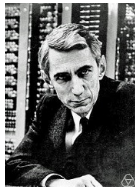
J. Malo (2020) **J. Math. Neuroscience** Spatio-chromatic information available from different neural layers

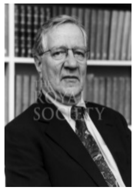
(2) COMPUTATIONAL TOOLS : Information theory

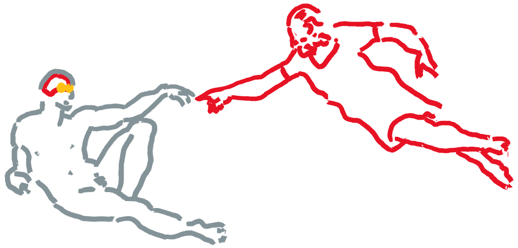
The "design" problem:

$$x \xrightarrow{S} y = S(x) + n$$









please maximize $I(x,y)$!

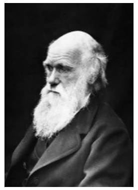
$$I(x,y) = h(x) + E_x\left[\log \|\nabla S\|\right] - \left(h(n) - E_n\left[D_{KL}\left(p(S(x)) \,|\, p(S(x)+n)\right)\right]\right)$$

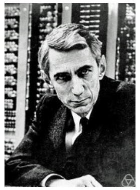
$$I(x,y) = \underbrace{\sum_i h(y_i)}_{(1)} - \underbrace{T(y)}_{(2)} - h(n) \Rightarrow \begin{cases} (1)\ \text{MAXIMIZE ENTROPY} \\ (2)\ \text{MINIMIZE REDUNDANCY} \end{cases}$$

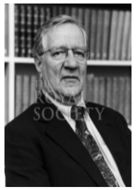
J. Malo (2020) **J. Math. Neuroscience** Spatio-chromatic information available from different neural layers

② COMPUTATIONAL TOOLS : Information theory

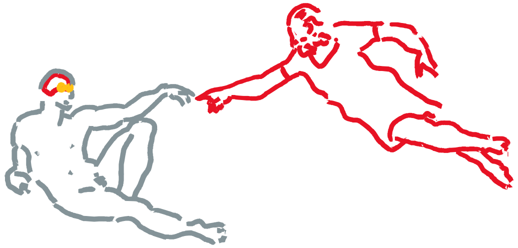
The "design" problem:







$$x \xrightarrow{S} y = S(x) + n$$



please maximize $I(x,y)$!

$$I(x,y) = h(x) + E_x\left[\log \|\nabla S\|\right] - \left(h(n) - E_n\left[D_{KL}\left(p(S(x)) \,|\, p(S(x)+n)\right)\right]\right)$$

$$I(x,y) = \underbrace{\sum_i h(y_i)}_{(1)} - \underbrace{T(y)}_{(2)} - h(n) \Rightarrow \begin{cases} (1) \text{ MAXIMIZE ENTROPY} \\ (2) \text{ MINIMIZE REDUNDANCY} \end{cases} \Rightarrow \begin{array}{l} \text{UNIFORMIZATION} \\ \text{GAUSSIANIZATION} \end{array}$$

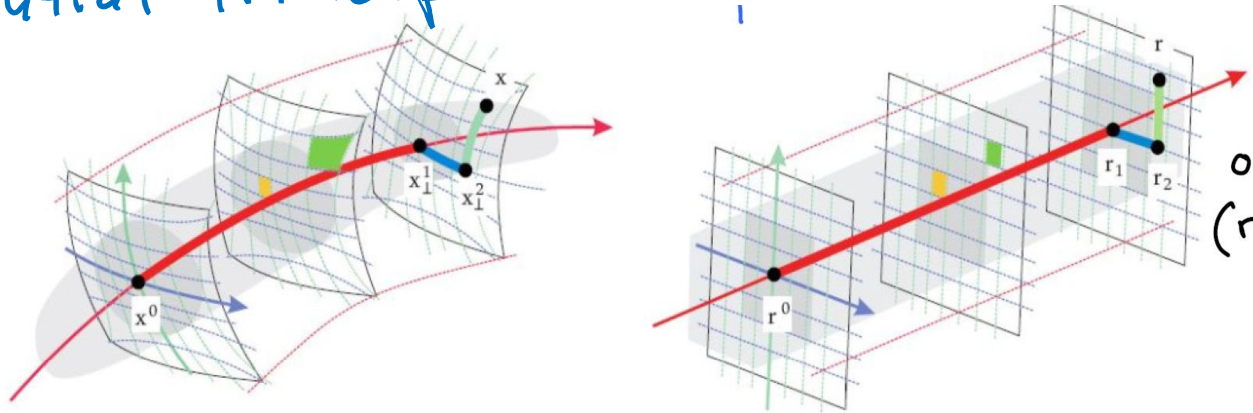
J. Malo (2020) **J. Math. Neuroscience** Spatio-chromatic information available from different neural layers

② COMPUTATIONAL TOOLS : Information theory

UNIFORMIZATION : GAUSSIANIZATION

Sequential Principal Curves Analysis (SPCA)

INPUT (signal) x



OUTPUT (response) u

* INFOMAX

* ERROR MINIMIZATION

γ

$$y = S(\mathrm{x}) = C \cdot \int_{\mathrm{x}^o}^{\mathrm{x}} \nabla \; (\mathrm{x}') \cdot d\mathrm{x}' = C \cdot \int_{\mathrm{x}^o}^{\mathrm{x}} D(\mathrm{x}') \cdot \nabla U(\mathrm{x}') \cdot d\mathrm{x}'$$

$$! \;\; _i = C_{ii} \cdot \int_{\mathrm{x}_\perp^{i-1}}^{\mathrm{x}_\perp^{i}} D(\mathrm{x}') \cdot \nabla U(\mathrm{x}') \cdot d\mathrm{x}' = C_{ii} \int_0^{u_{i\perp}^{i}} p_{u_i}(u_i')^{\gamma} du_i'$$

J. Malo & J. Gutiérrez (2006) V1-nonlinearities emerge from Local-to-Global ICA
Network: Comp. Neur. Syst. Vol. 17, 85–102

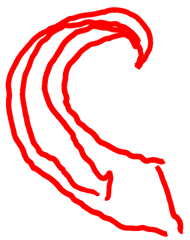
V. Laparra, J. Malo et al. (2012) Nonlinearities and adaptation in color vision from Sequential Principal Curves Analysis
Neural Comput. Vol. 24, 2751–2788. doi: 10.1162/NECO_a_00342

V. Laparra & J. Malo (2015) Visual aftereffects and sensory nonlinearities from a single statistical framework
Front. Hum. Neurosci., https://doi.org/10.3389/fnhum.2015.00557

2 COMPUTATIONAL TOOLS : Information theory

UNIFORMIZATION : GAUSSIANIZATION

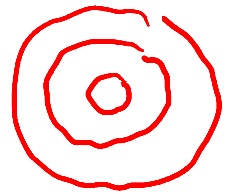
$x^{(0)}$



ANY PDF

$p(x^{(0)})$

$x^{(N)}$



GAUSSIAN PDF

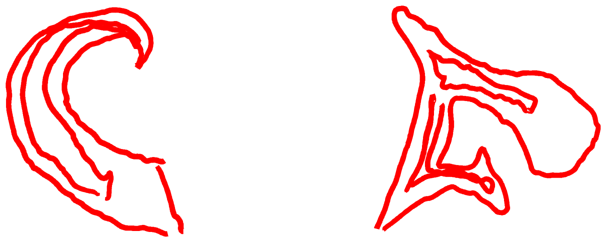
$p(x^{(N)}) = \mathcal{N}(x^{(N)}, 0, I)$

$$x^{(n+1)} = \underbrace{R^{(n)}}_{\text{Rotation}} \cdot \underbrace{\Psi^{(n)}(x^{(n)})}_{\text{Marginal Gaussianization}}$$

2 COMPUTATIONAL TOOLS : Information theory

: GAUSSIANIZATION

$x^{(0)} \longrightarrow x^{(1)}$



ANY PDF

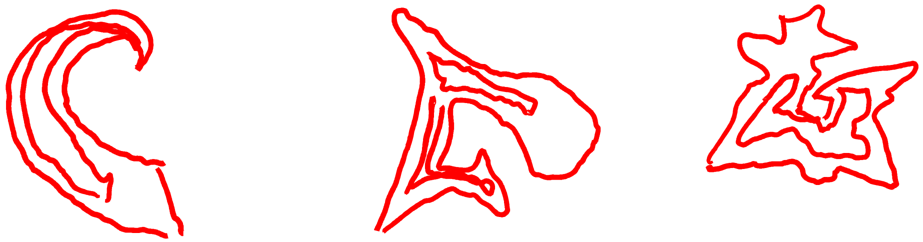
$p(x^{(0)})$

$$x^{(n+1)} = \underbrace{R^{(n)}}_{\text{Rotation}} \cdot \underbrace{\Psi^{(n)}(x^{(n)})}_{\text{Marginal Gaussianization}}$$

(2) COMPUTATIONAL TOOLS : Information theory

: GAUSSIANIZATION

$$x^{(0)} \longrightarrow x^{(1)} \longrightarrow x^{(2)}$$



ANY PDF

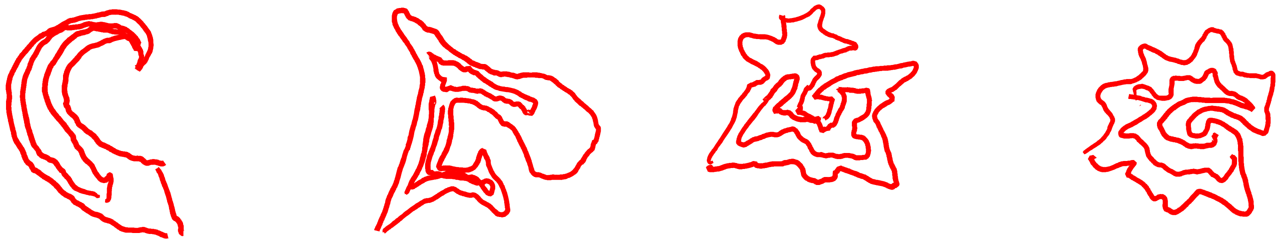
$P(x^{(0)})$

$$x^{(n+1)} = \underbrace{R^{(n)}}_{\text{Rotation}} \cdot \underbrace{\Psi^{(n)}(x^{(n)})}_{\text{Marginal Gaussianization}}$$

② COMPUTATIONAL TOOLS : Information theory

: GAUSSIANIZATION

$$x^{(0)} \to x^{(1)} \to x^{(2)} \to x^{(3)}$$



ANY PDF

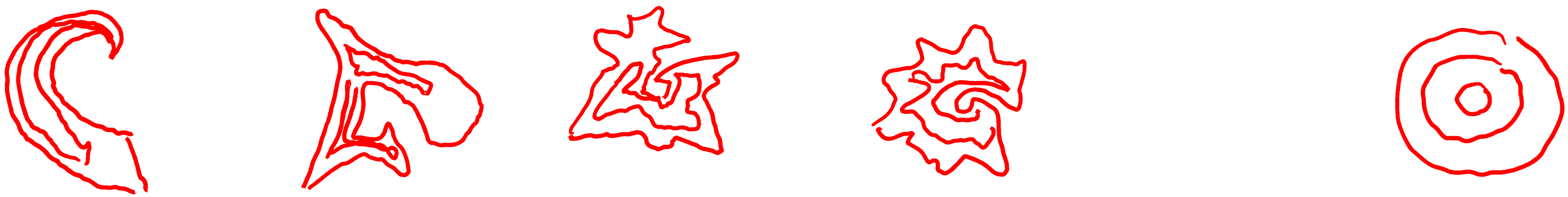
$p(x^{(0)})$

$$x^{(n+1)} = \underbrace{R^{(n)}}_{\text{Rotation}} \cdot \underbrace{\Psi^{(n)}(x^{(n)})}_{\text{Marginal Gaussianization}}$$

2 COMPUTATIONAL TOOLS : Information theory

: GAUSSIANIZATION

$$x^{(0)} \longrightarrow x^{(1)} \longrightarrow x^{(2)} \longrightarrow x^{(3)} \cdots\cdots \longrightarrow x^{(N)}$$



ANY PDF

$p(x^{(0)})$

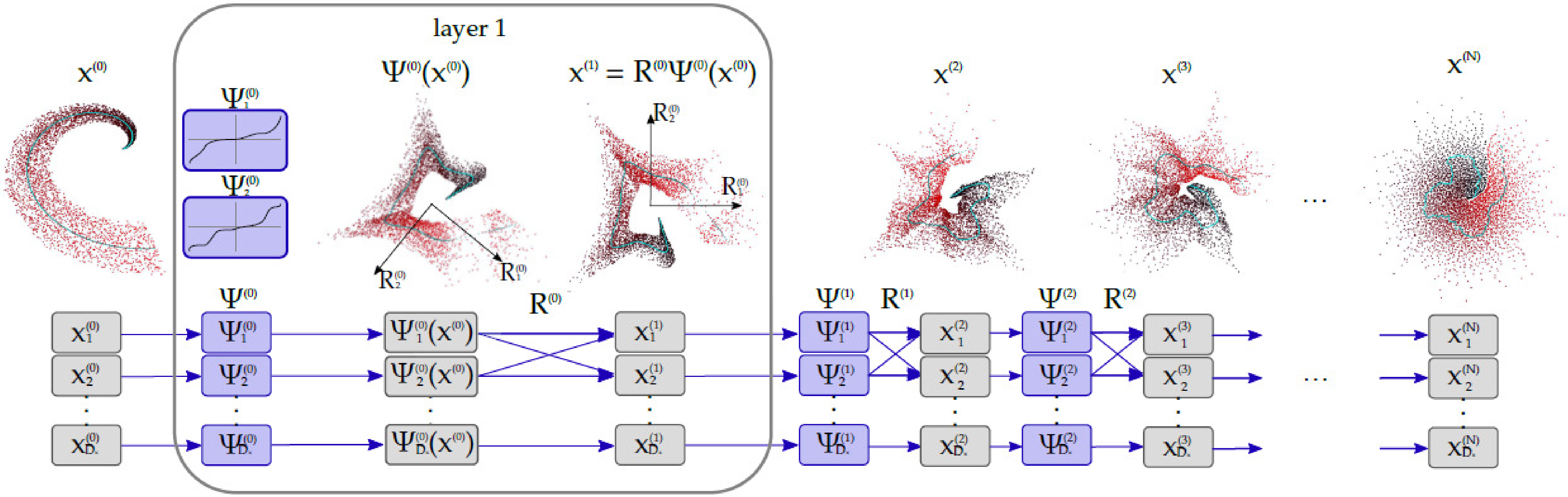
GAUSSIAN PDF

$p(x^{(N)}) = \mathcal{N}(x^{(n)}, 0, I)$

$$x^{(n+1)} = \underbrace{R^{(n)}}_{\text{Rotation}} \cdot \underbrace{\Psi^{(n)}(x^{(n)})}_{\text{Marginal Gaussianization}}$$

② COMPUTATIONAL TOOLS : Information theory

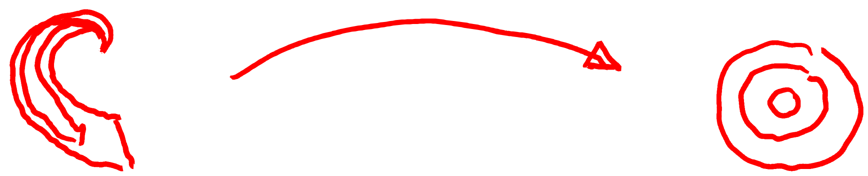
: GAUSSIANIZATION



$$x^{(n+1)} = \underbrace{R^{(n)}}_{\text{Rotation}} \cdot \underbrace{\Psi^{(n)}(x^{(n)})}_{\text{Marginal Gaussianization}}$$

② COMPUTATIONAL TOOLS : Information theory : GAUSSIANIZATION

$x^{(0)}$ $x^{(N)}$



ANY PDF $p(x^{(0)})$

GAUSSIAN PDF $p(x^{(N)}) = \mathcal{N}(x^{(N)}, 0, I)$

In ANY differentiable transform $\Rightarrow$ In ANY Gaussianization $T(x') = 0$

$$\begin{aligned}\Delta T(\mathrm{x}, \mathrm{x}') &= T(\mathrm{x}) - T(\mathrm{x}') \\ &= \sum_{i=1}^{D_x} H(x_i) - \sum_{i=1}^{D_{x'}} H(x'_i) + \frac{1}{2} \underbrace{\mathbb{E}_x\left(\log |\nabla G_x(\mathrm{x})^\top \cdot \nabla G_x(\mathrm{x})|\right)}\end{aligned}$$

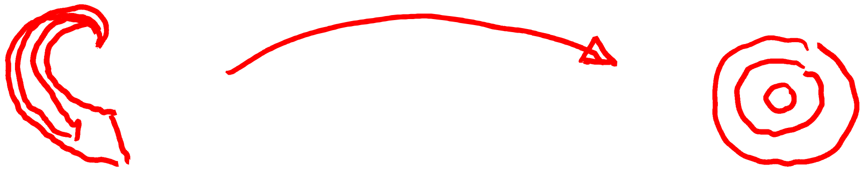
$$T(\mathrm{x}) = \sum_{i=1}^{D_x} H(x_i) - \frac{D_x}{2}\log(2\pi e) + \underbrace{\mathbb{E}_x\left(\log |\nabla G_x(\mathrm{x})|\right)}$$

IN RBIG $\equiv$ ONLY UNIVARIATE OPERATIONS

$$\boxed{\tilde{T}(\mathrm{x}) = \sum_{n=0}^{N-1} \Delta \tilde{T}^{(n)} = \frac{(N-1)D_x}{2}\log(2\pi e) - \sum_{n=1}^{N}\sum_{i=1}^{D_x} \tilde{H}(x_i^{(n)})}$$

② COMPUTATIONAL TOOLS : Information theory : GAUSSIANIZATION

$x^{(0)}$ $\qquad\qquad\qquad$ $x^{(N)}$



ANY PDF $\qquad\qquad$ GAUSSIAN PDF

$p(x^{(0)})$ $\qquad\qquad$ $p(x^{(N)}) = \mathcal{N}(x^{(N)}, 0, I)$

In ANY differentiable transform $\Rightarrow$ In ANY Gaussianization $T(x')=0$

$$\begin{aligned}\Delta T(\mathrm{x},\mathrm{x}') &= T(\mathrm{x}) - T(\mathrm{x}') \\ &= \sum_{i=1}^{D_x} H(x_i) - \sum_{i=1}^{D_{x'}} H(x'_i) + \frac{1}{2}\underbrace{\mathbb{E}_x\left(\log|\nabla G_x(\mathrm{x})^\top \cdot \nabla G_x(\mathrm{x})|\right)}\end{aligned}$$

$$T(\mathrm{x}) = \sum_{i=1}^{D_x} H(x_i) - \frac{D_x}{2}\log(2\pi e) + \underbrace{\mathbb{E}_x\left(\log|\nabla G_x(\mathrm{x})|\right)}$$

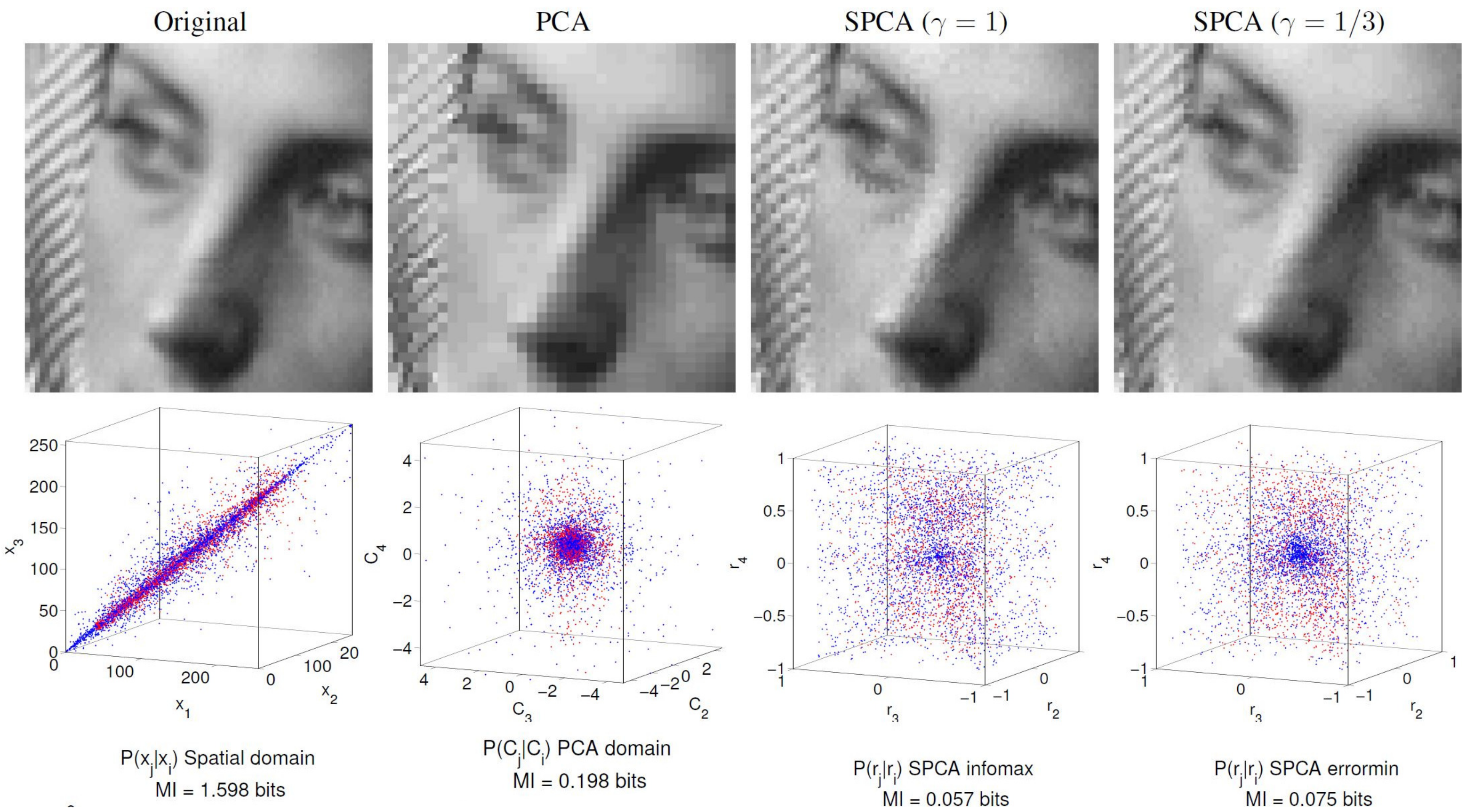
IN RBIG ≡ ONLY UNIVARIATE OPERATIONS

$$\boxed{\tilde{T}(\mathrm{x}) = \sum_{n=0}^{N-1} \Delta \tilde{T}^{(n)} = \frac{(N-1)D_x}{2}\log(2\pi e) - \sum_{n=1}^{N}\sum_{i=1}^{D_x} \tilde{H}(x_i^{(n)})}$$

GOOD TO ESTIMATE INFORM. THEORY MEASURES!

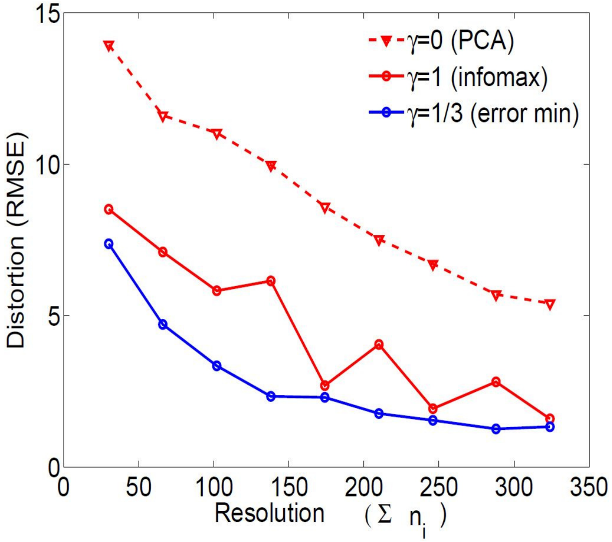
② COMPUTATIONAL TOOLS : Information theory

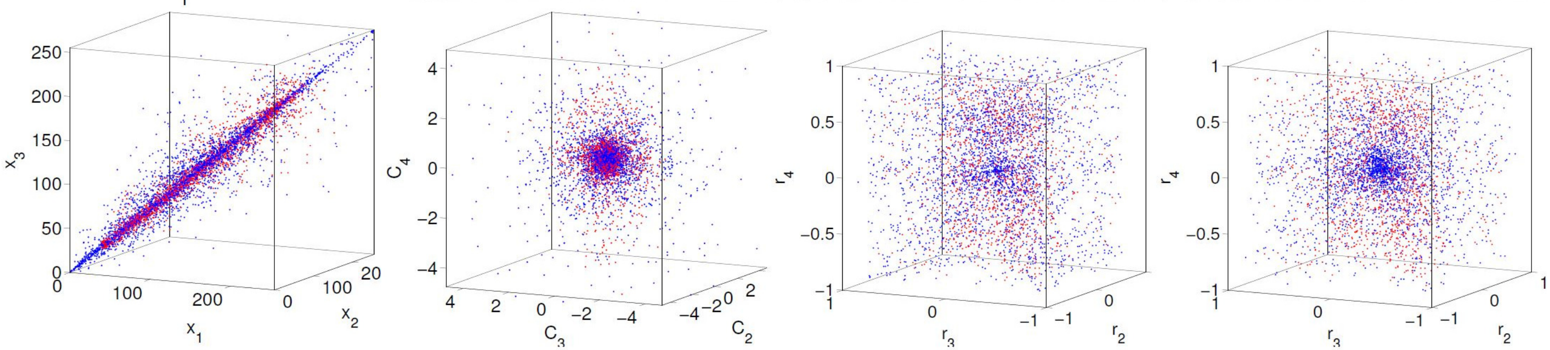
UNIFORMIZATION : GAUSSIANIZATION



② COMPUTATIONAL TOOLS : Information theory

UNIFORMIZATION : GAUSSIANIZATION





P(x_j|x_i) Spatial domain
MI = 1.598 bits

P(C_j|C_i) PCA domain
MI = 0.198 bits

P(r_j|r_i) SPCA infomax
MI = 0.057 bits

P(r_j|r_i) SPCA errormin
MI = 0.075 bits

② COMPUTATIONAL TOOLS : Information theory

UNIFORMIZATION : GAUSSIANIZATION

Total Correlation

$\tilde{T}(\mathbf{x})$

		D_x	RBIG	kNN	KDP	expF	vME	Ens
Gaussian	-	3	0.87	0.94	76.65	0.63	4.27	4.03
		10	0.97	23.48	>100	0.27	31.72	34.83
		50	1.45	45.77	>100	0.52	>100	54.74
		100	1.55	52.78	>100	0.41	>100	59.94
Rotated	-	3	1.70	1.80	82.90	16.80	1.90	9.40
		10	8.30	27.20	>100	11.00	24.20	38.70
		50	7.70	51.10	>100	15.10	>100	59.40
		100	7.50	57.80	>100	15.50	>100	64.50
Student	$\nu = 3$	3	7.01	13.55	>100	94.03	>100	66.59
		10	32.93	16.73	>100	67.32	>100	15.27
		50	18.18	12.02	>100	29.44	>100	24.65
		100	12.71	17.41	>100	21.12	>100	28.63
	$\nu = 5$	3	26.61	52.76	>100	89.74	81.85	133.12
		10	23.94	19.74	>100	49.60	>100	12.31
		50	10.10	16.87	>100	20.29	>100	32.14
		100	7.10	22.53	>100	15.39	>100	34.96
	$\nu = 20$	3	88.27	>100	>100	48.56	>100	>100
		10	3.05	11.86	>100	10.51	>100	19.93
		50	3.07	33.17	>100	4.54	>100	52.62
		100	1.31	35.56	>100	3.43	>100	49.46

Differential Entropy

$\tilde{H}(\mathbf{x})$

		D_x	RBIG	kNN	KDP	expF	vME	Ens
Gaussian		3	[illegible]	1.70	112.30	1.10	8.80	12.00
		10	1.20	27.90	179.80	0.10	34.70	40.30
		50	0.90	32.20	107.40	0.10	108.40	38.10
		100	0.70	30.70	89.60	0.10	94.20	34.60
Rotated		3	4.00	6.20	171.40	36.80	3.70	30.10
		10	12.20	38.50	241.80	17.90	31.80	53.90
		50	6.80	44.60	136.60	13.50	87.90	51.60
		100	5.50	42.50	110.70	11.50	94.30	47.20
Student	$\nu = 3$	3	0.75	0.76	34.93	11.90	3.13	1.99
		10	2.82	1.44	137.19	15.55	53.19	1.77
		50	6.03	3.47	195.30	22.11	175.94	7.09
		100	6.61	8.57	228.62	24.44	166.09	13.72
	$\nu = 5$	3	0.51	0.70	24.75	3.42	1.38	1.94
		10	1.12	1.25	96.73	5.52	59.29	1.21
		50	2.82	4.84	146.63	9.59	202.48	8.89
		100	3.11	10.67	184.02	11.36	195.17	16.24
	$\nu = 20$	3	0.54	0.71	19.19	0.76	1.32	1.56
		10	0.48	0.84	69.83	1.56	46.62	0.37
		50	0.95	6.66	107.74	3.31	219.86	11.13
		100	0.68	13.45	138.98	4.20	214.41	19.35

Kullback-Leibler Div.

$D_{\mathrm{KL}}(\mathbf{y}|\mathbf{x})$

		dim	RBIG	kNN	expF	vME
Gaussian, different means	$\mu_2 = 0.2$	3	13.49	16.90	10.28	92.28
		10	19.50	22.47	3.27	>1000
		50	32.04	40.85	13.34	>1000
		100	47.40	41.24	28.14	>1000
	$\mu_2 = 0.4$	3	3.55	7.98	5.57	24.22
		10	5.25	22.93	2.02	604.17
		50	8.67	40.91	3.40	>1000
		100	4.83	43.49	8.70	>1000
	$\mu_2 = 0.6$	3	3.75	6.39	3.89	12.23
		10	2.81	24.72	1.85	213.72
		50	13.83	43.11	1.83	897.65
		100	42.42	46.00	5.11	686.96
Gaussians, different covs.	$\mu_2 = 0.5$	3	24.93	27.30	4.89	63.90
		10	18.80	103.65	2.64	>1000
		50	23.62	173.62	8.42	>1000
		100	32.56	200.33	17.59	>1000
	$\mu_2 = 0.75$	3	21.04	24.77	3.72	36.64
		10	10.44	96.85	1.86	605.00
		50	10.07	159.16	5.70	>1000
		100	13.66	179.67	11.40	>1000
	$\mu_2 = 0.9$	3	17.12	25.95	3.40	26.15
		10	6.77	94.42	1.60	448.87
		50	3.40	152.46	4.81	>1000
		100	5.96	170.28	9.43	>1000
Gaussian vs. Student	$\nu = 2$	3	5.08	29.58	793.53	5.78
		10	32.72	83.51	1278.91	596.63
		50	59.37	468.33	2783.43	>1000
		100	42.11	1024.30	4330.18	>1000
	$\nu = 4$	3	17.09	95.02	148.08	22.52
		10	42.84	157.63	219.26	963.37
		50	60.53	584.46	547.48	>1000
		100	41.71	1214.61	962.45	>1000
	$\nu = 7$	3	8.34	271.61	35.78	59.69
		10	38.78	307.82	49.77	>1000
		50	48.80	713.36	145.15	>1000
		100	26.01	1399.34	278.93	>1000
Student vs. Student	$\nu = 2$	3	9.08	13.87	3442.45	>1000
		10	20.57	57.60	7462.58	346.61
		50	85.14	405.47	19991.36	>1000
		100	242.80	939.24	35064.60	>1000
	$\nu = 4$	3	9.51	47.03	1502.19	48.89
		10	36.33	139.12	2561.86	>1000
		50	37.29	656.95	7997.12	>1000
		100	60.52	1441.18	13033.03	>1000
	$\nu = 7$	3	13.13	126.41	589.47	128.84
		10	23.13	301.97	1070.70	>1000
		50	28.34	976.95	3689.57	>1000
		100	145.88	2046.95	6370.43	>10000

Szabo JMLR 2014
- KNN
- Partition trees
- Exp. Family
- Von Mises
- Ensemble

https://isp.uv.es/RBIG4IT.htm

Mutual Information

$\tilde{I}(\mathbf{x},\mathbf{y})$

		D_x	RBIG	kNN	KDP	expF	vME	Ens
Gaussians		3	10.60	26.00	149.10	9.20	13.20	48.50
		10	9.60	76.30	102.60	23.70	311.00	91.00
		50	6.80	104.70	100.70	39.50	68.00	105.50
		100	11.70	107.20	>1000	42.60	77.40	106.10
Student vs. Student	$\nu = 3$	3	35.72	95.32	>1000	63.73	>1000	86.58
		10	22.26	2.66	118.51	18.14	>1000	66.77
		50	1.51	88.38	104.50	36.10	810.02	105.83
		100	15.34	98.66	>1000	65.71	789.55	105.34
	$\nu = 5$	3	18.51	118.04	>1000	56.49	>1000	96.41
		10	3.07	24.83	113.89	9.39	>1000	101.26
		50	10.91	102.89	105.08	25.17	849.12	117.30
		100	24.43	105.41	101.10	42.57	805.44	110.58
	$\nu = 20$	3	73.63	194.16	>1000	14.63	>1000	15.36
		10	40.02	108.82	110.68	29.69	>1000	208.20
		50	29.98	149.53	102.93	36.30	946.93	154.88
		100	37.21	128.27	101.44	43.77	844.41	127.67

2 COMPUTATIONAL TOOLS : Information theory

* Information theory gives intuition
  - Criterion for $S(x)$ $(\nabla S(x))$
  - Principles
    - Entropy maxim.
    - Redundancy reduct.

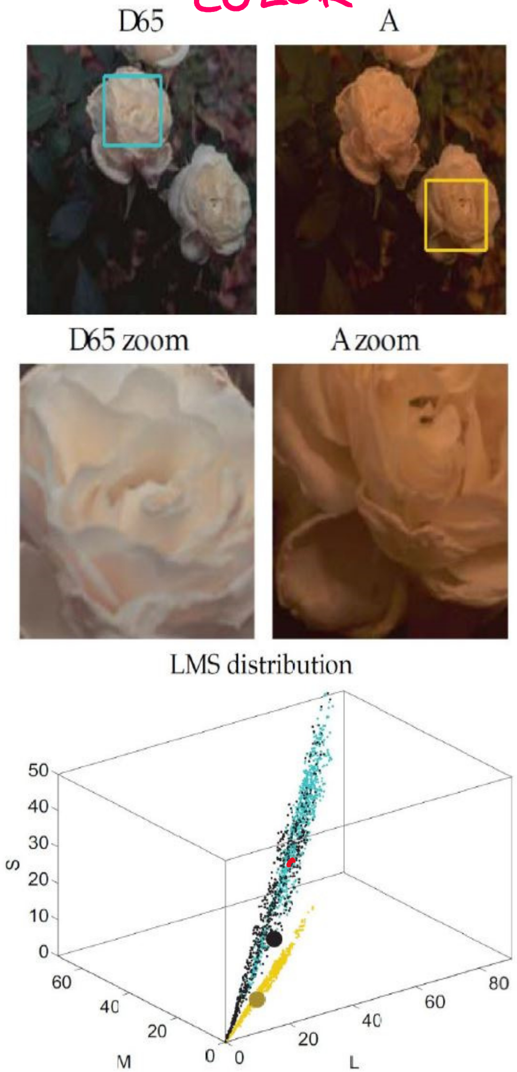
* We have tools
  - SPCA
  - RBIG https://isp.uv.es/RBIG4IT.htm

(3) NATURAL SCENES ARE NOT ARBITRARY: Color / images / movies

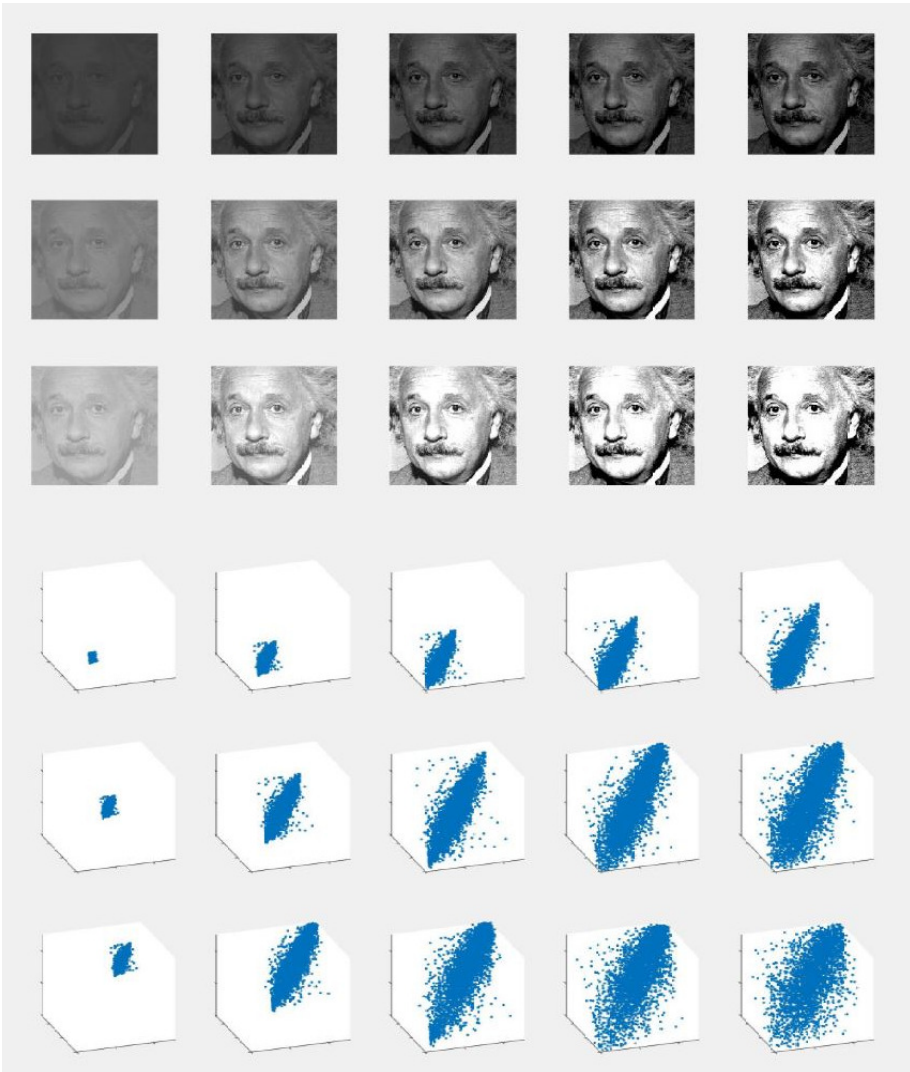
* The manifold view (rediscovered in 90's)
* Non uniformity
* Smoothness → Redundancy
* Non Gaussianity

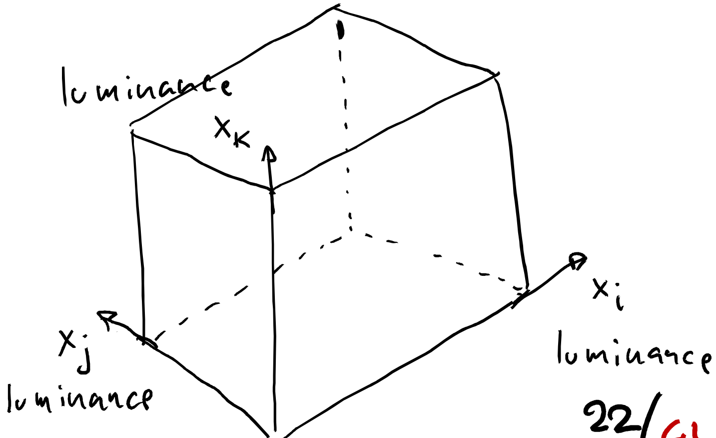
(3) NATURAL SCENES ARE NOT ARBITRARY: Color / images / movies

COLOR



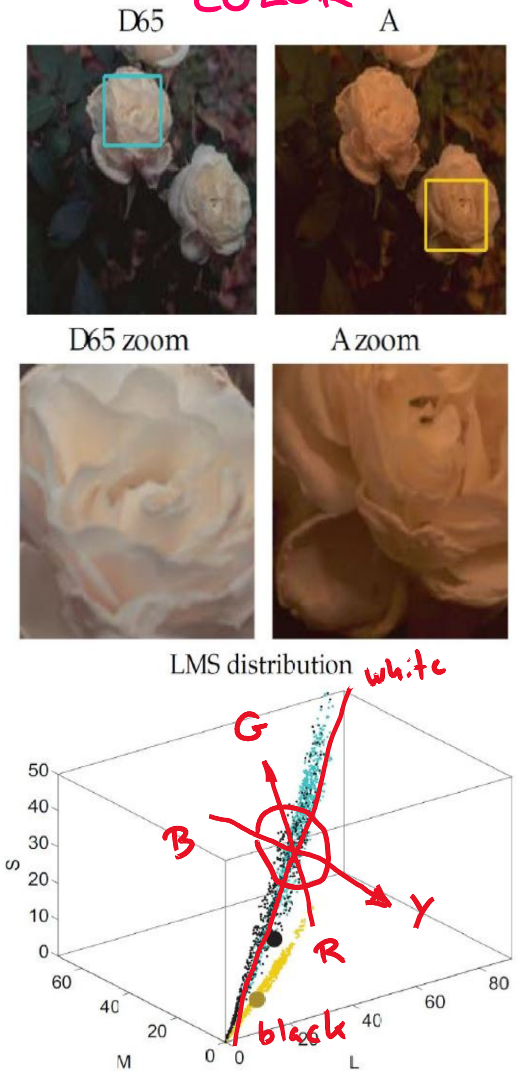
SPATIAL TEXTURE



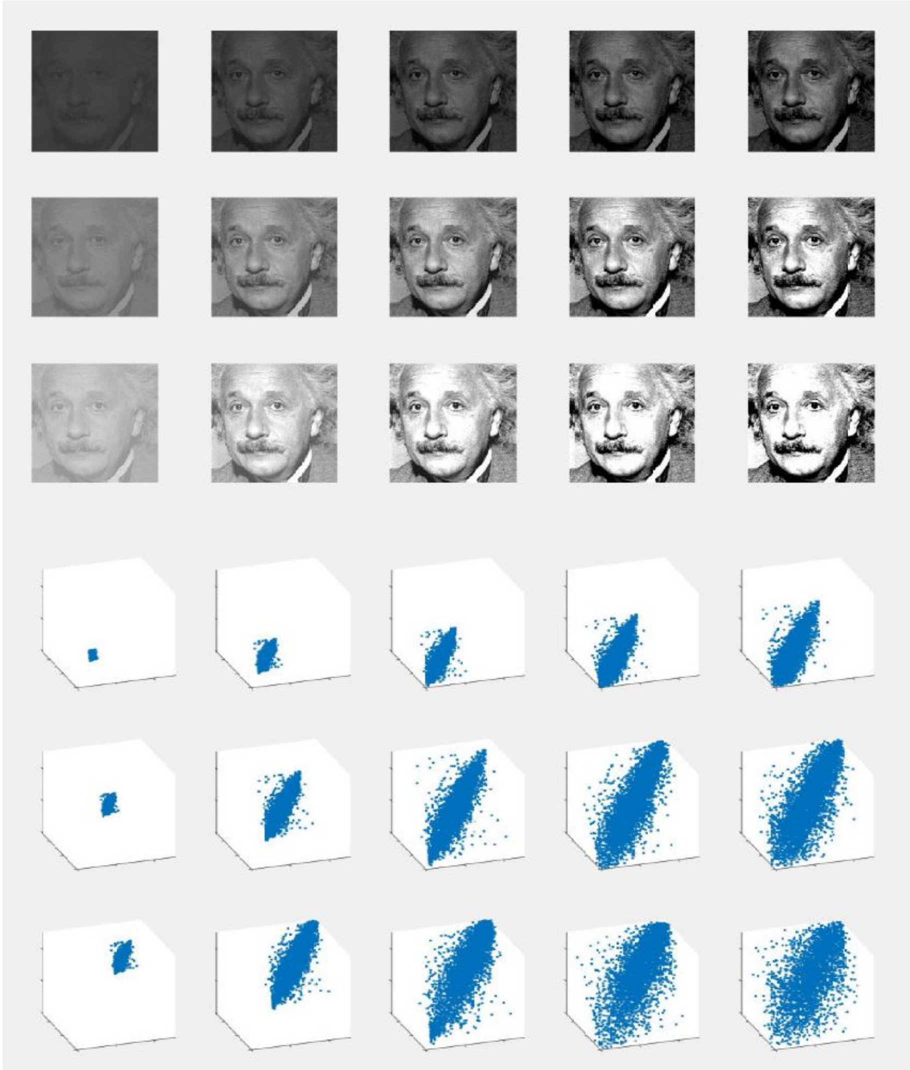


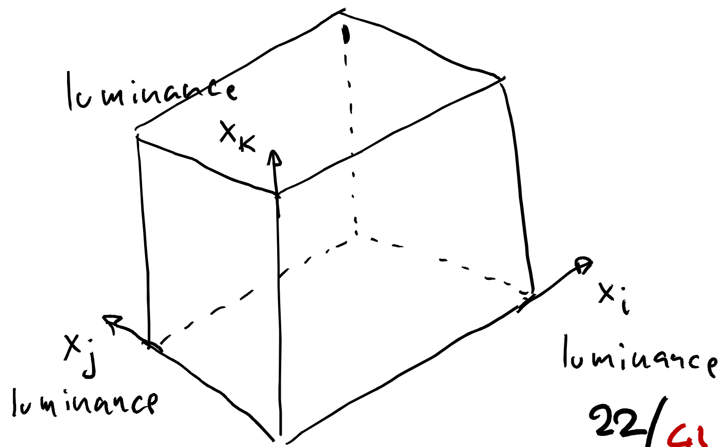
(3) NATURAL SCENES ARE NOT ARBITRARY: Color / images / movies

COLOR



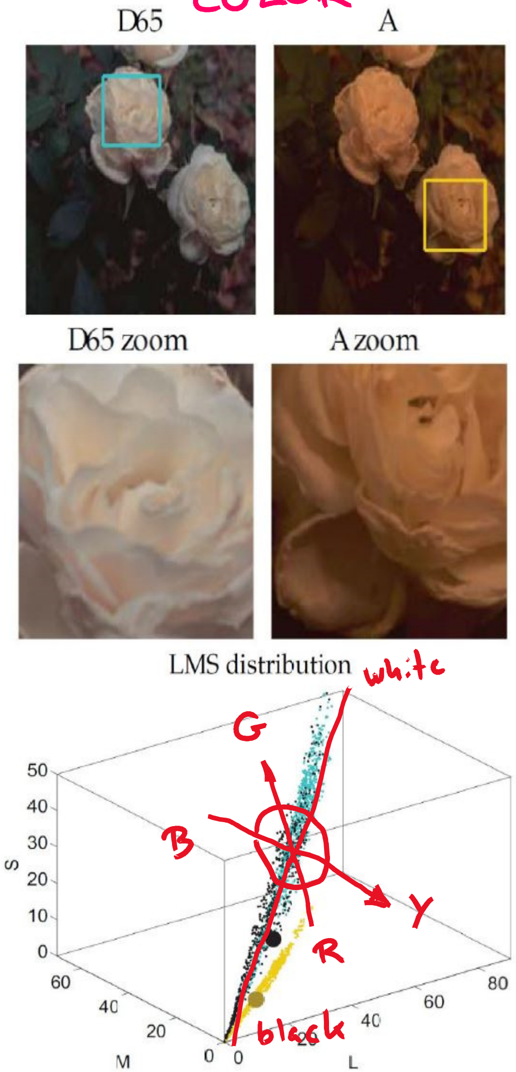
SPATIAL TEXTURE



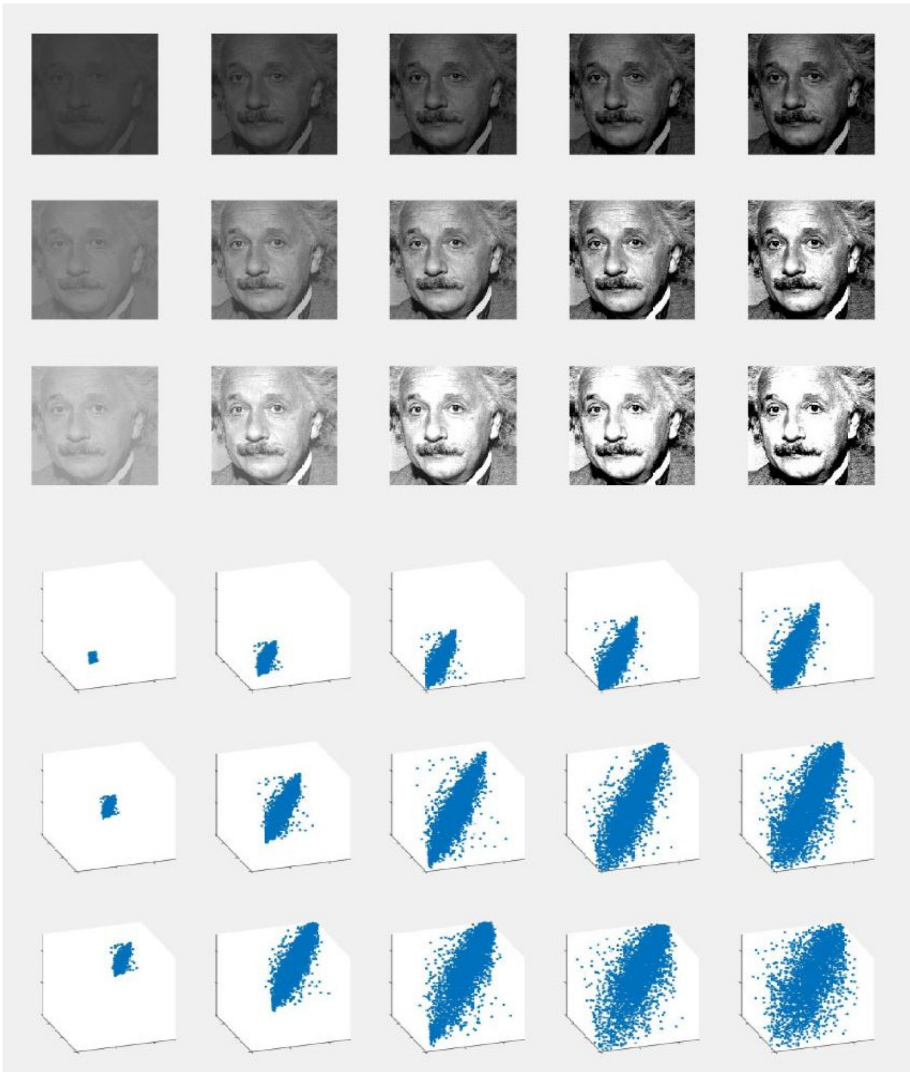


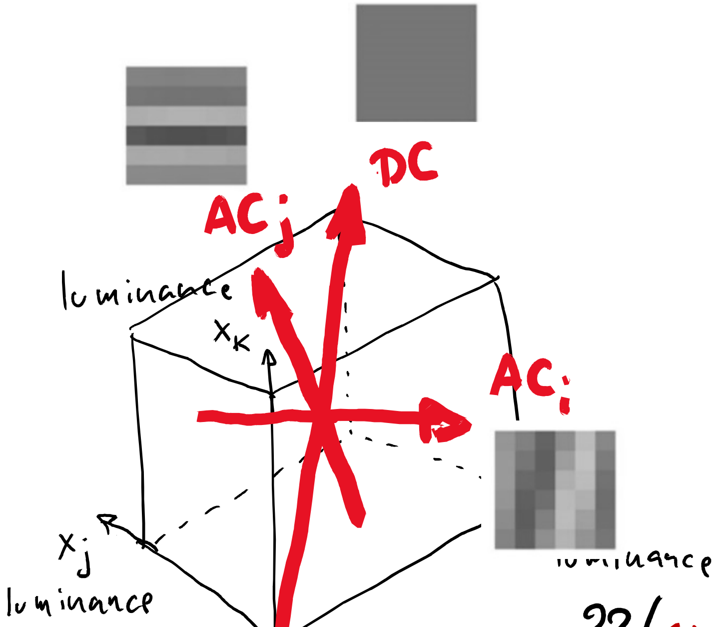
③ NATURAL SCENES ARE NOT ARBITRARY: Color / images / movies

COLOR

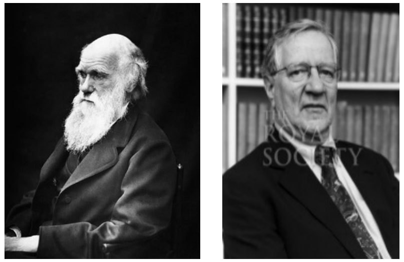


SPATIAL TEXTURE

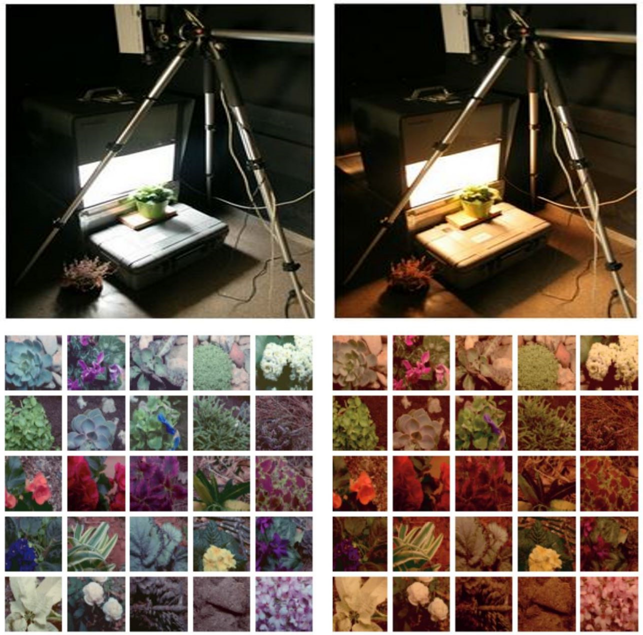


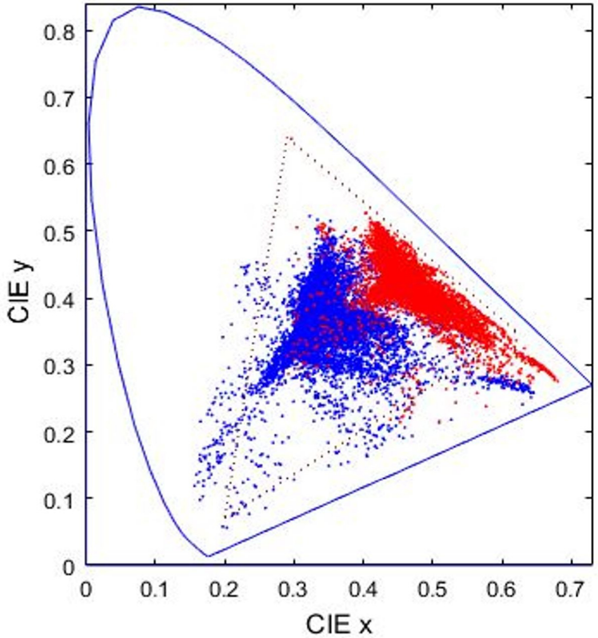


(3) NATURAL SCENES ARE NOT ARBITRARY: Color / images / movies



Natural Environment

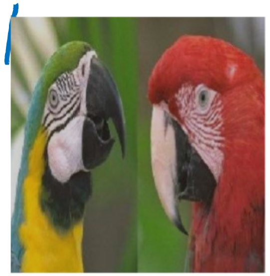




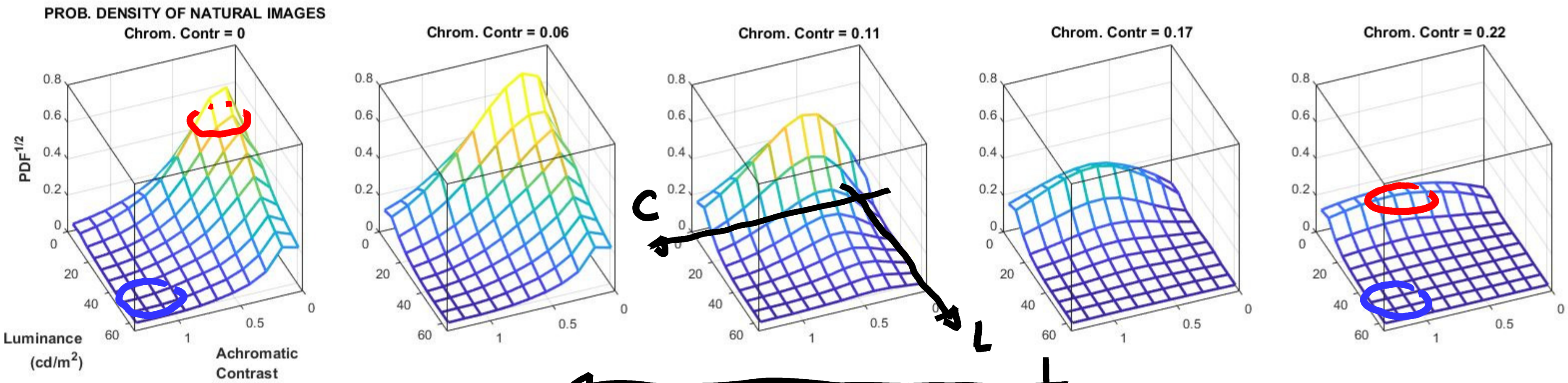
Laparra & Malo **Neural Comp.** 2012, Gutmann & Malo **PLOS** 2014 https://isp.uv.es/data_color.htm

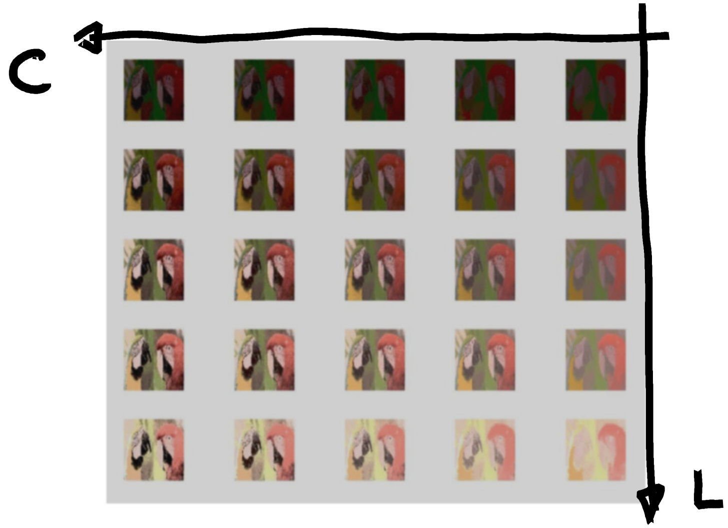
Gómez & Malo **J. Neurophysiol.** 2020 https://isp.uv.es/code/visioncolor/infoWilsonCowan.html

(3) NATURAL SCENES ARE NOT ARBITRARY: Color / images /



Probability Density Function of Natural Images





(3) NATURAL SCENES ARE NOT ARBITRARY: Color / images / movies

... movies

Laparra & Malo Front. Neurosci. 2015

Li, Gomez, Bertalmio & Malo Submit. JoV. 2021

... non-gaussianity

Olshausen & Simoncelli Ann. Rev. Neurosci. 01

In summary:

* Images have interesting regularities ⇒ will determine $S(x)$
* Images are mostly achromatic & low contrast
* Images are mostly lowpass (spatio-temporally)
* Images are not Gaussian

https://isp.uv.es/data_color.htm

https://isp.uv.es/after_effects

4 THE TWO SIDES OF THE EFFICIENT CODING HYPOTHESIS: Stats $\rightleftarrows$ Biology

The conventional way

Stats $\longrightarrow$ Biology

The alternative way

Stats $\longleftarrow$ Biology

4 THE TWO SIDES OF THE EFFICIENT CODING HYPOTHESIS: Stats $\rightleftarrows$ Biology

The conventional way

Stats $\longrightarrow$ Biology

REDUNDANCY REDUCTION / INFOMAX

- Linear analysis
  - PCA — Opponent Channels, Freq. Sens. f.
  - ICA — V1-like receptive fields.

Gutmann, Laparra, Hyvarinen & Malo **PLOS** 2014

- Non-linearities — SPCA
  - Color
  - Texture
  - Motion

Laparra & Malo **Front. Neurosci.** 2015

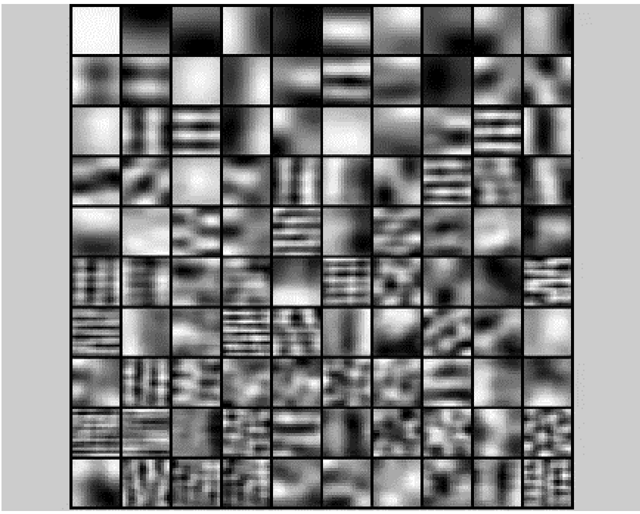
ERROR MINIMIZATION — CSFs in Autoencoders

Gomez-Villa et al. **Vision Res.** 2020

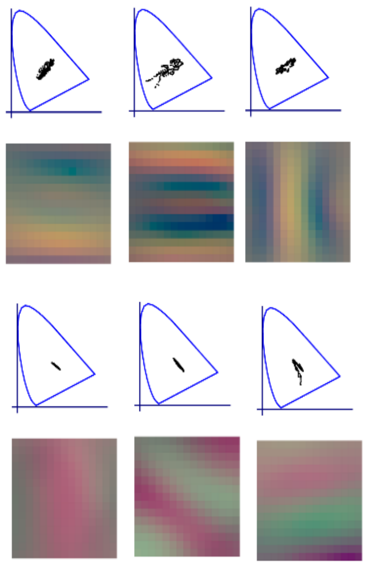
CLASSIFICATION — V1-like representation

4 THE TWO SIDES OF THE EFFICIENT CODING HYPOTHESIS:

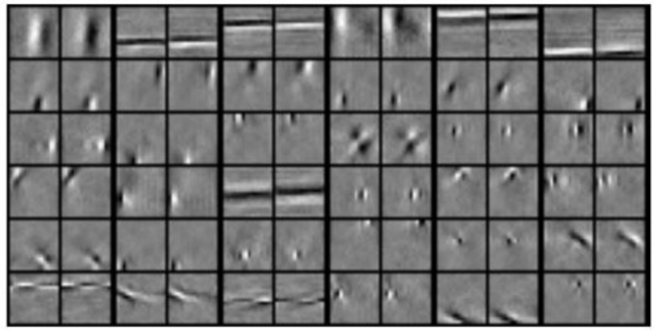
Stats, ⇄ Biology



- Linear analysis { PCA, ICA }

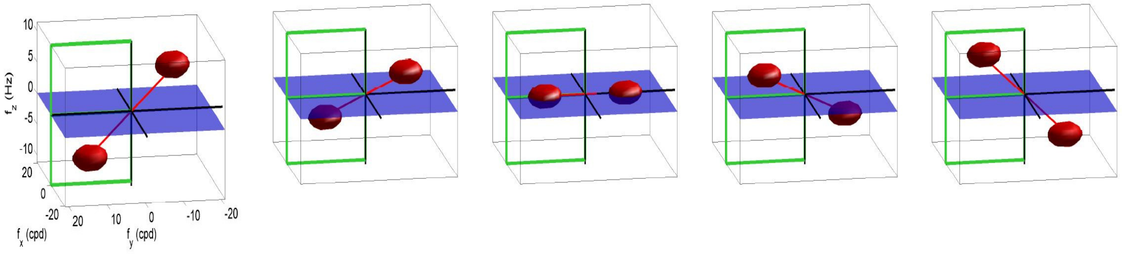


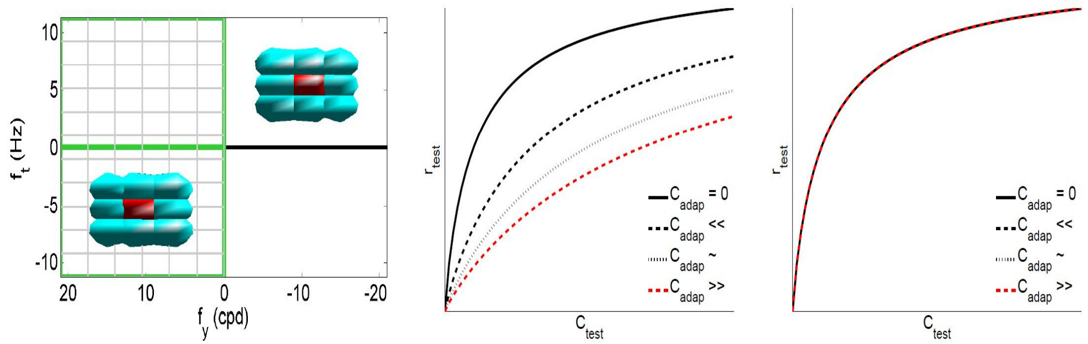
Receptive fields in phase-quadrature via Complex ICA
[LNCS 11]

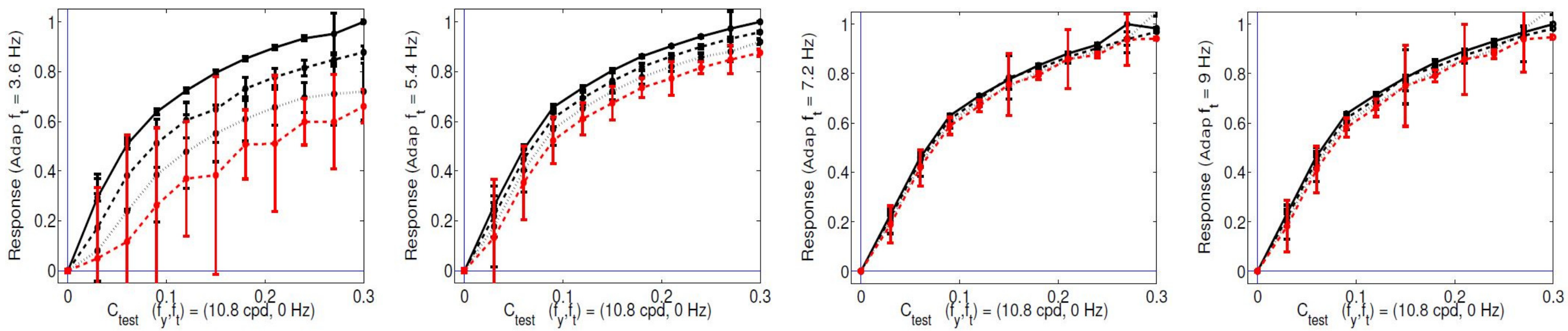


Spatio-chromatic receptive fields via Higher-Order CCA
[PLoS 14]

4 THE TWO SIDES OF THE EFFICIENT CODING HYPOTHESIS: Stats, $\rightleftarrows$ Biology





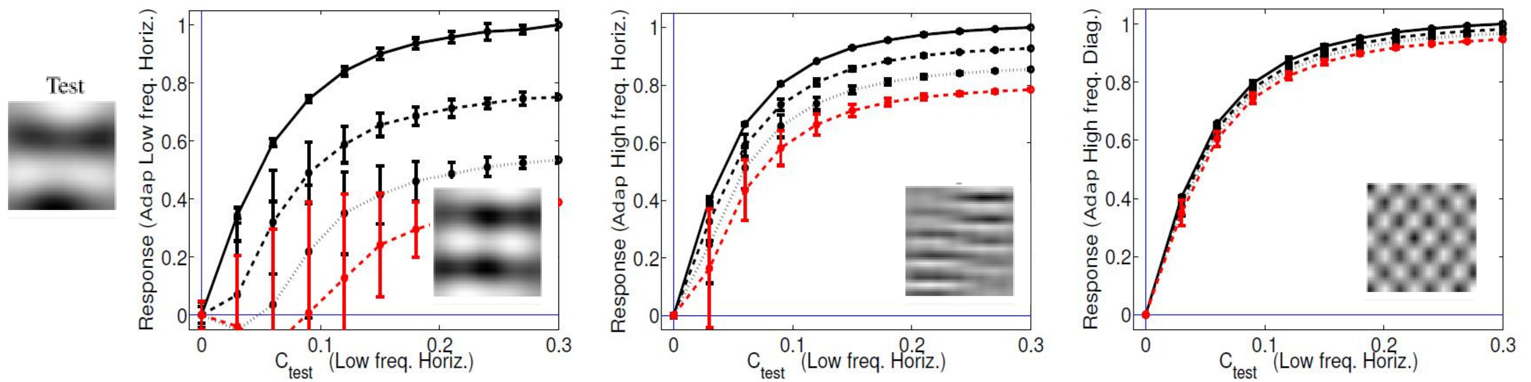


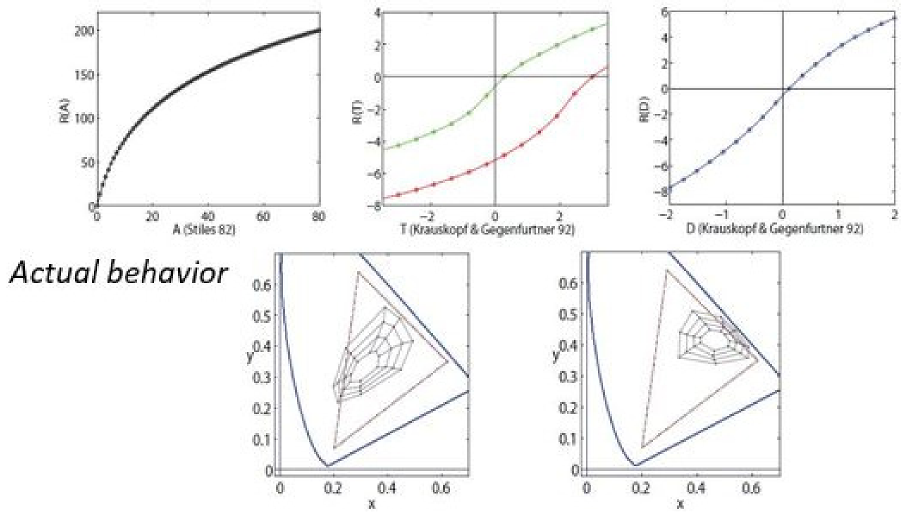
4 THE TWO SIDES OF THE EFFICIENT CODING HYPOTHESIS:

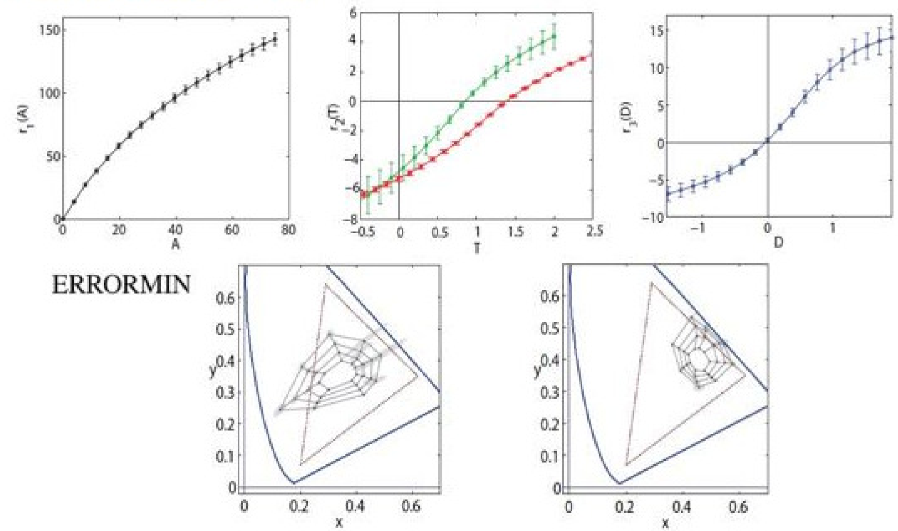
Stats, ⇄ Biology

- Non. linearities

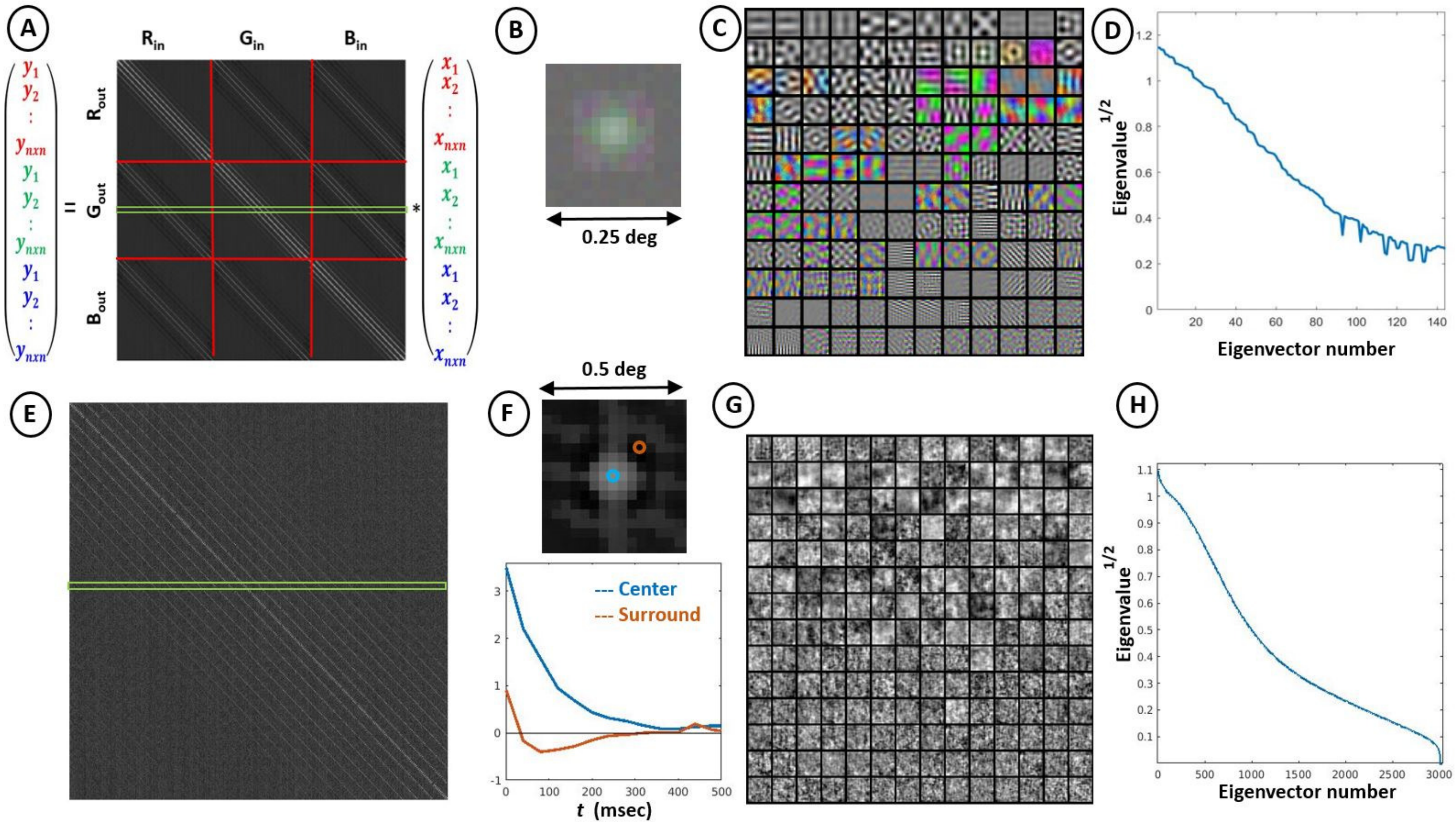
SPCA
- Color
- Texture
- Motion

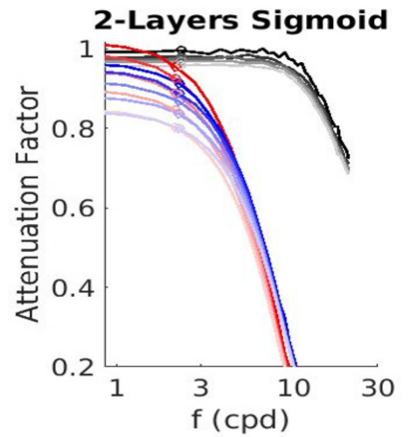


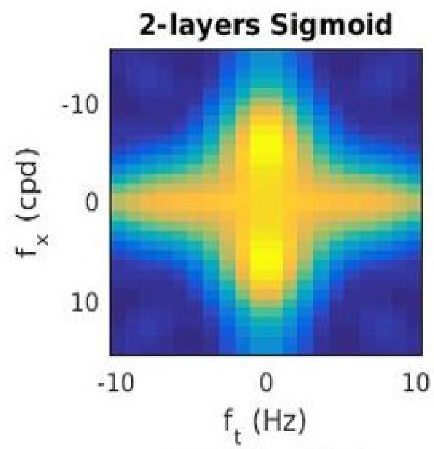




4 THE TWO SIDES OF THE EFFICIENT CODING HYPOTHESIS: Stats ⇄ Biology







· Error minimization   CSFs in Autoencoders

4 THE TWO SIDES OF THE EFFICIENT CODING HYPOTHESIS: Stats, $\rightleftarrows$ Biology

REVIEW

doi:10.1038/nature14539

NATURE 2015

Deep learning

Yann LeCun[1,2], Yoshua Bengio[3] & Geoffrey Hinton[4,5]

Deep learning allows computational models that are composed of multiple processing layers to learn representations of data with multiple levels of abstraction. These methods have dramatically improved the state-of-the-art in speech recognition, visual object recognition, object detection and many other domains such as drug discovery and genomics. Deep learning discovers intricate structure in large data sets by using the backpropagation algorithm to indicate how a machine should change its internal parameters that are used to compute the representation in each layer from the representation in the previous layer. Deep convolutional nets have brought about breakthroughs in processing images, video, speech and audio, whereas recurrent nets have shone light on sequential data such as text and speech.

Classification performance →

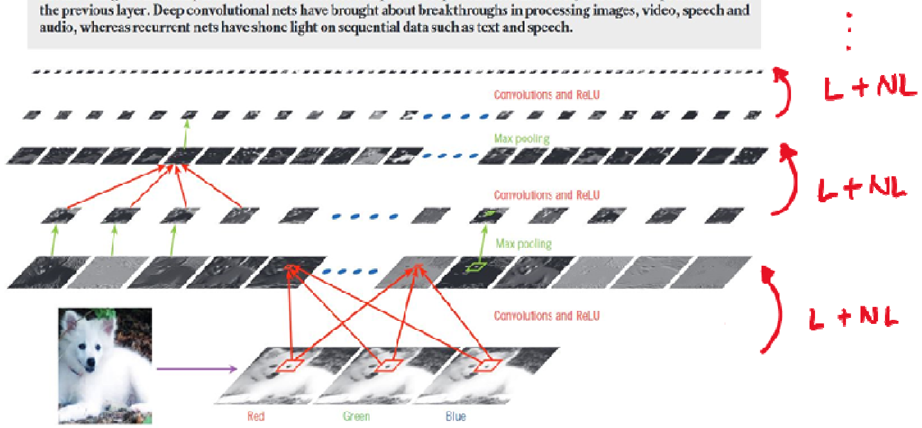
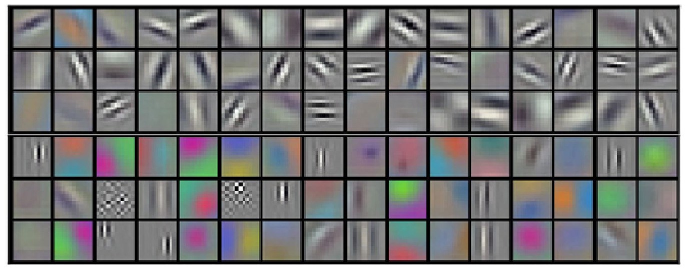


Figure 2 | Inside a convolutional network.

NIPS 2012



④ THE TWO SIDES OF THE EFFICIENT CODING HYPOTHESIS: Stats, $\rightarrow$ Biology $\leftarrow$

The alternative way

Stats $\leftarrow$ Biology

Malo & Simoncelli **IEEE Trans. Im. Proc.** 2006
Malo & Laparra **Neural Comp.** 2010

Redundancy reduction in psychophysical models } · Analytic
· RBIG

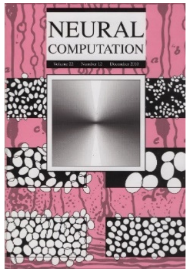
Gómez-Villa et al. **J. Neurophysiol.** 2020

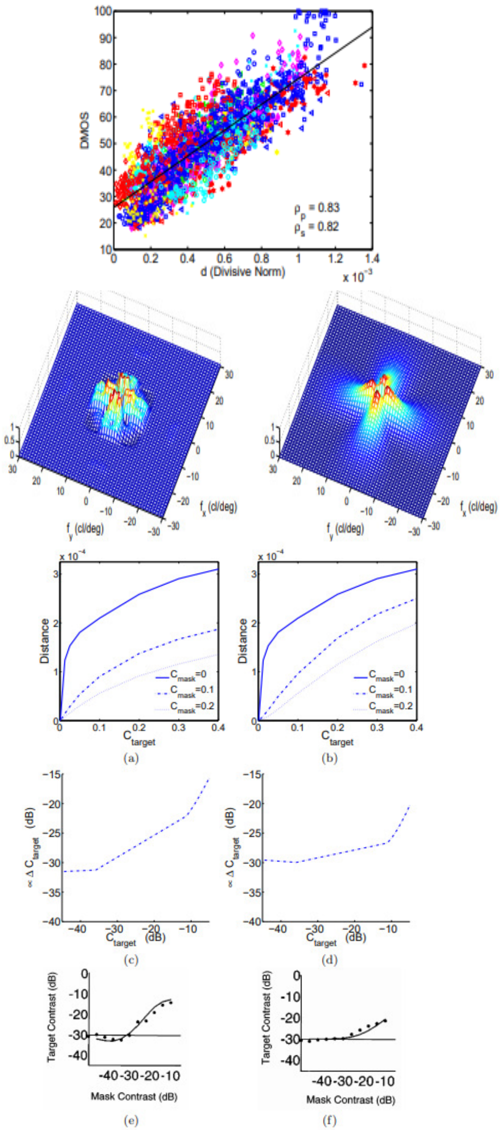
Information transmission in psychophysical models RBIG

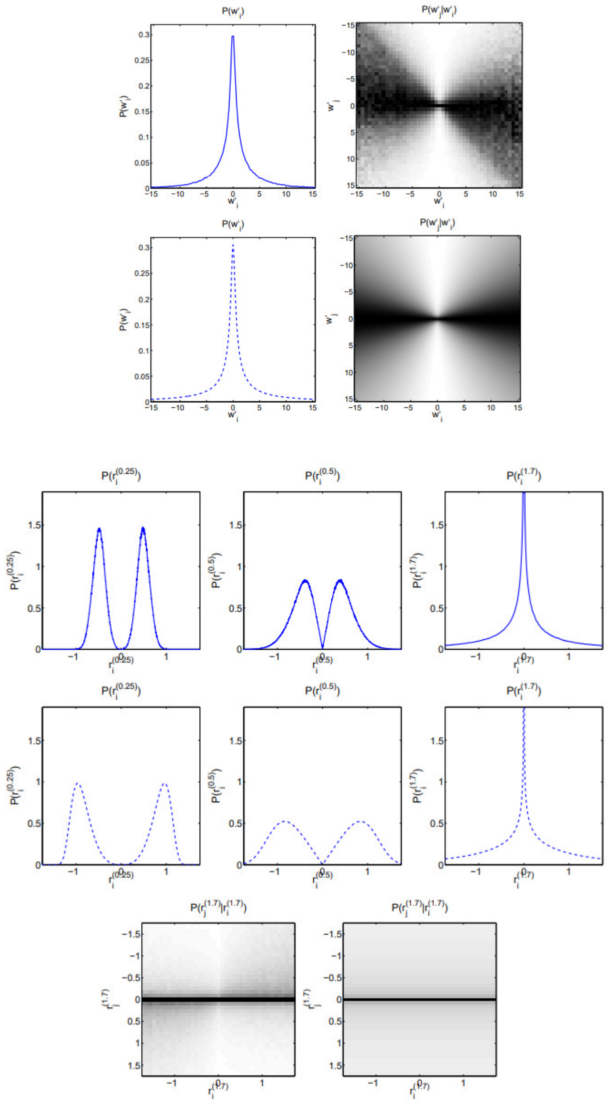
Malo **J. Math. Neurosci.** 2020

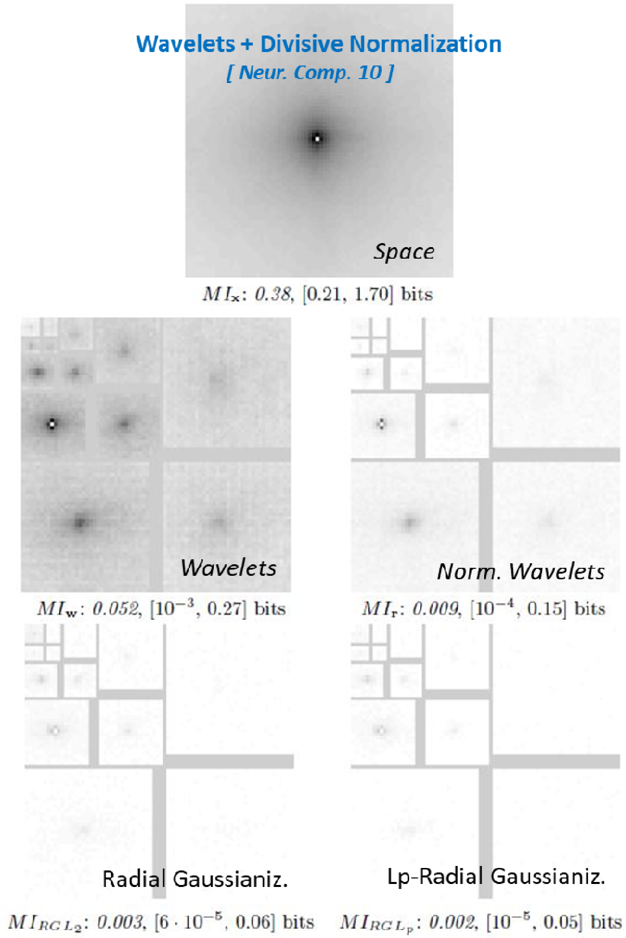
4 THE TWO SIDES OF THE EFFICIENT CODING HYPOTHESIS:

Stats, $\rightarrow$ $\leftarrow$ Biology









Local-DCT + Divisive Normalization

[Im.Vis.Comp.97, IEEE TIP 06]

	pixels	*local-DCT*	*local-PCA*	*normalized-DCT*
I_r	0.69	0.28	0.29	0.06

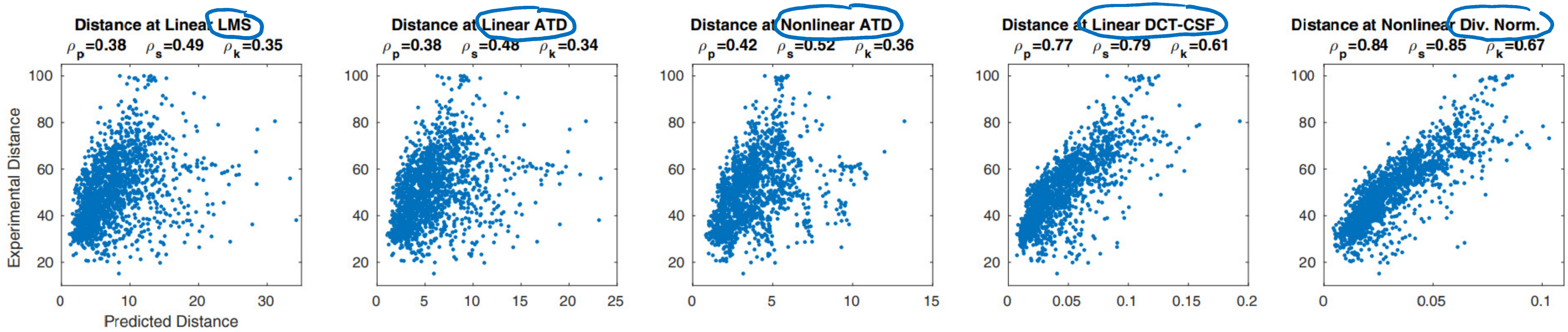
4 THE TWO SIDES OF THE EFFICIENT CODING HYPOTHESIS: Stats. $\rightleftarrows$ Biology

Information transmission in psychophysical models RBG

Malo **J. Math. Neurosci.** 2020

4 THE TWO SIDES OF THE EFFICIENT CODING HYPOTESIS: I

Stats. $\rightleftarrows$ Biology



Pearson correlation with human viewers using different building blocks (or model layers)

	Spatial Extent	$r^{(1)}$	$r^{(2)}$	$x^{(2)}$	$r^{(3)}$	$x^{(3)}$
More flexible model	(0.27 deg)	0.38	0.38	0.42	0.77	0.51
Baseline model	(0.27 deg)	0.38	0.38	0.42	0.77	**0.84**
More rigid model	(0.27 deg)	0.38	0.38	0.42	0.77	0.79
Totally rigid model	(0.27 deg)	0.38	0.38	0.38	0.68	0.68
Baseline model	(0.05 deg)	0.26	0.27	0.31	0.37	0.40

Assumptions:

(1) single-step transforms

$$\begin{tikzcd} x \arrow[r] \arrow[r, bend left] \arrow[rr, bend left] \arrow[rrr, bend left] & y \arrow[r] & z \arrow[r] & k \end{tikzcd}$$

(2) Constant SNR (5% noise)

④ THE TWO SIDES OF THE EFFICIENT CODING HYPOTESIS: Stats. $\rightleftarrows$ Biology

Redundancy Reduction

$\Delta T(x^{input}, x^{resp.})$

Pearson correlation with human viewers using different building blocks (or model layers).

	Spatial Extent	$\boldsymbol{r}^{(1)}$	$\boldsymbol{r}^{(2)}$	$\boldsymbol{x}^{(2)}$	$\boldsymbol{r}^{(3)}$	$\boldsymbol{x}^{(3)}$
More flexible model	(0.27 deg)	0.38	0.38	0.42	0.77	0.51
Baseline model	(0.27 deg)	0.38	0.38	0.42	0.77	**0.84**
More rigid model	(0.27 deg)	0.38	0.38	0.42	0.77	0.79
Totally rigid model	(0.27 deg)	0.38	0.38	0.38	0.68	0.68
Baseline model	(0.05 deg)	0.26	0.27	0.31	0.37	0.40

Transmitted Information

$I(x^{in}, x^{resp})$

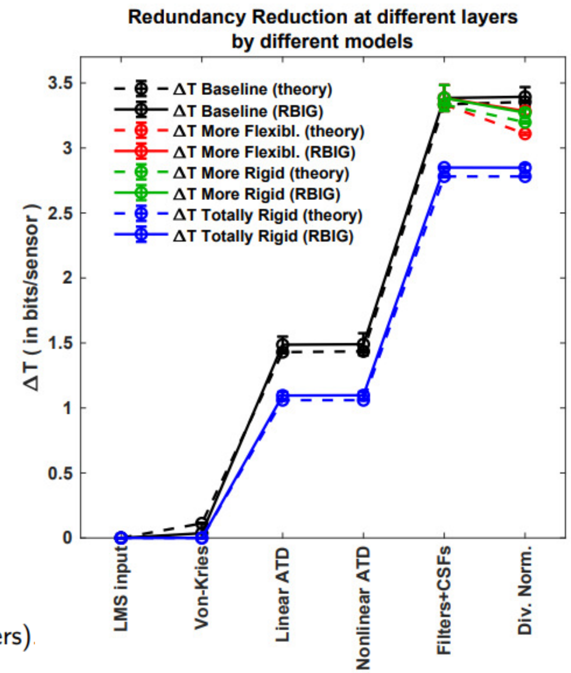
④ Stats, ⟶ Biology
⟵

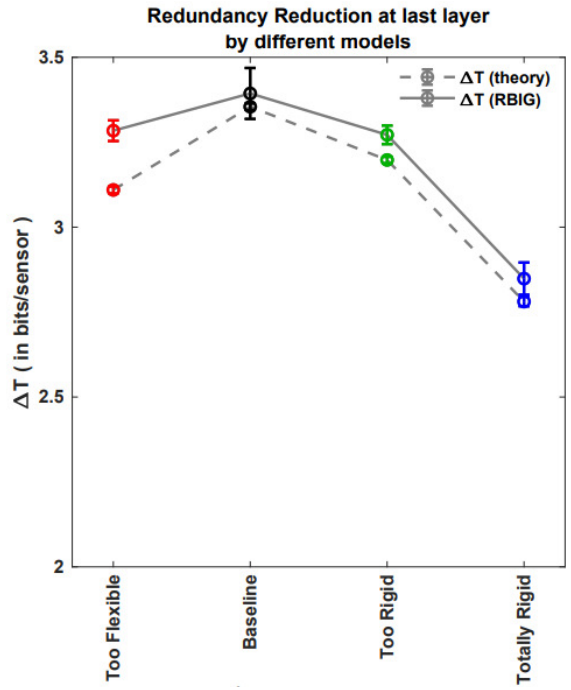
Redundancy Reduction

$\Delta T(x^{input}, x^{resp.})$

ΔT_{RBIG}

$\Delta T_{theor} = \sum_i h(x_i^{in}) - h(x_i^{resp}) + E_x\left[\log_2 |\nabla S|\right]$





Pearson correlation with human viewers using different building blocks (or model layers)

	Spatial Extent	$r^{(1)}$	$r^{(2)}$	$x^{(2)}$	$r^{(3)}$	$x^{(3)}$
More flexible model	(0.27 deg)	0.38	0.38	0.42	0.77	0.51
Baseline model	(0.27 deg)	0.38	0.38	0.42	0.77	**0.84**
More rigid model	(0.27 deg)	0.38	0.38	0.42	0.77	0.79
Totally rigid model	(0.27 deg)	0.38	0.38	0.38	0.68	0.68
Baseline model	(0.05 deg)	0.26	0.27	0.31	0.37	0.40

Transmitted Information

$I(x^{in}, x^{resp})$

4 Stats, → Biology ←

Redundancy Reduction

$\Delta T(x^{input}, x^{resp.})$

ΔT_{RBIG}

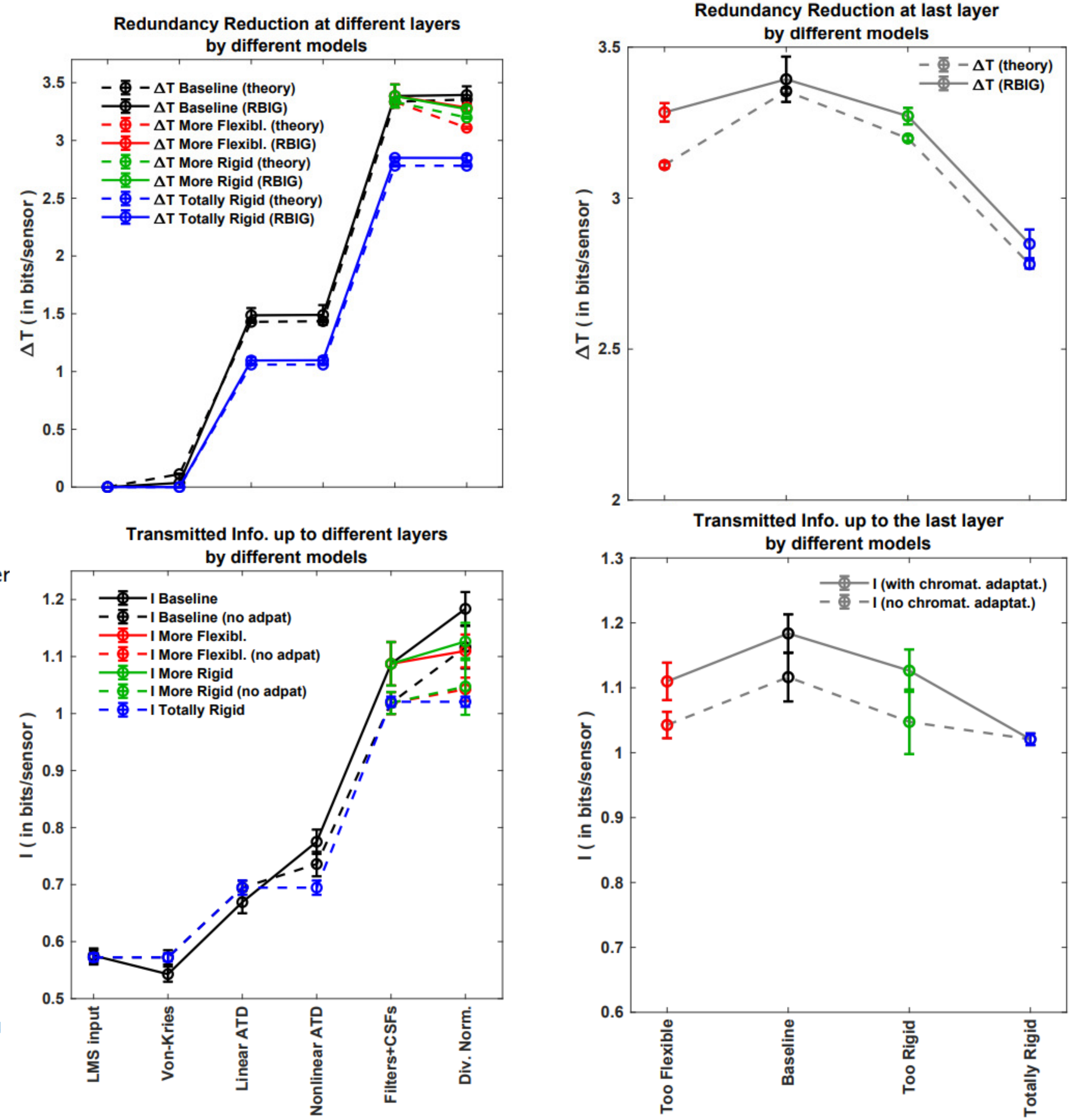
$\Delta T_{theor} = \sum_i h(x_i^{in}) - h(x_i^{resp}) + E_x\left[\log_2 |\nabla S|\right]$

Pearson correlation with human viewers using different building blocks (or model layer

	Spatial Extent	$r^{(1)}$	$r^{(2)}$	$x^{(2)}$	$r^{(3)}$	$x^{(3)}$
More flexible model	(0.27 deg)	0.38	0.38	0.42	0.77	0.51
Baseline model	(0.27 deg)	0.38	0.38	0.42	0.77	**0.84**
More rigid model	(0.27 deg)	0.38	0.38	0.42	0.77	0.79
Totally rigid model	(0.27 deg)	0.38	0.38	0.38	0.68	0.68
Baseline model	(0.05 deg)	0.26	0.27	0.31	0.37	0.40

Transmitted Information

$I(x^{in}, x^{resp})$



4 Stats $\rightleftarrows$ Biology

Redundancy Reduction

$\Delta T(x^{input}, x^{resp.})$

ΔT_{RBIG}

$\Delta T_{theor} = \sum_i h(x_i^{in}) - h(x_i^{resp}) + E_x\left[\log_2 |\nabla S|\right]$

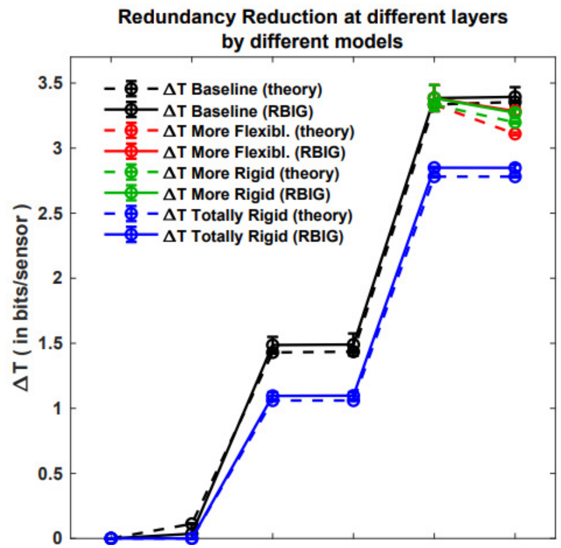
Pearson correlation with human viewers using different building blocks (or model layer

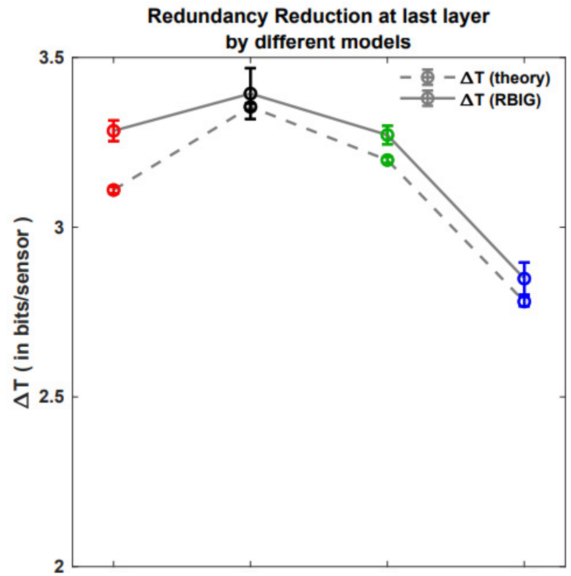
	Spatial Extent	$r^{(1)}$	$r^{(2)}$	$x^{(2)}$	$r^{(3)}$	$x^{(3)}$
More flexible model	(0.27 deg)	0.38	0.38	0.42	0.77	0.51
Baseline model	(0.27 deg)	0.38	0.38	0.42	0.77	**0.84**
More rigid model	(0.27 deg)	0.38	0.38	0.42	0.77	0.79
Totally rigid model	(0.27 deg)	0.38	0.38	0.38	0.68	0.68
Baseline model	(0.05 deg)	0.26	0.27	0.31	0.37	0.40

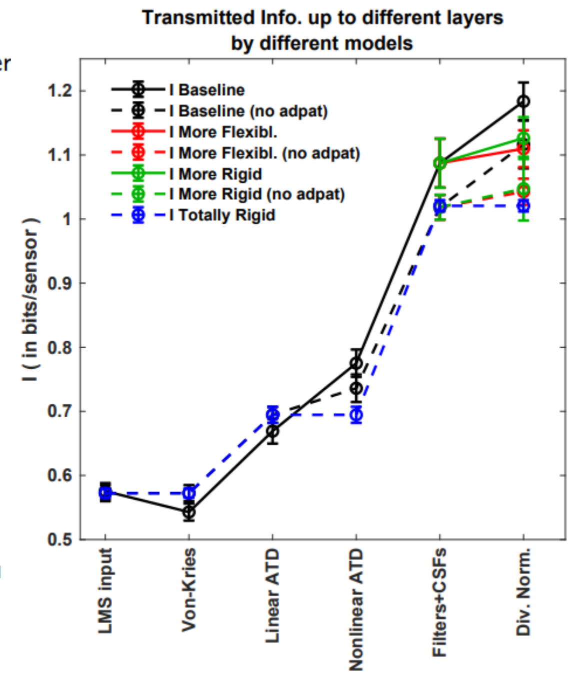
- Deeper is better
- Baseline is better
- Space vs Color

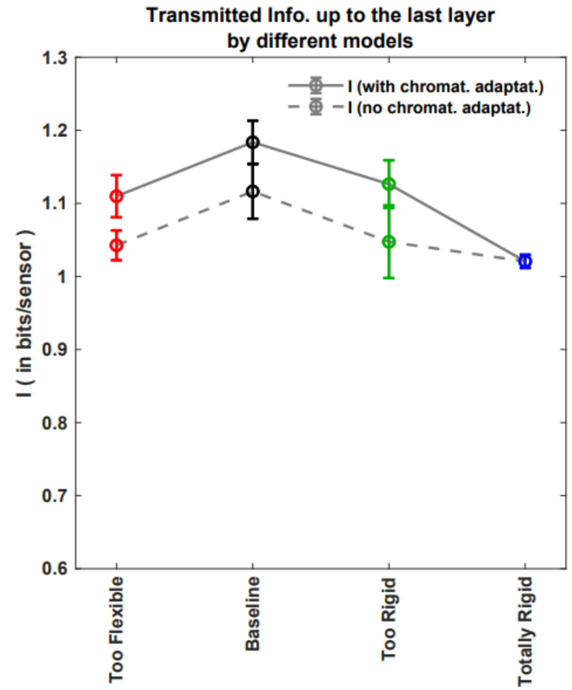
Transmitted Information

$I(x^{in}, x^{resp})$









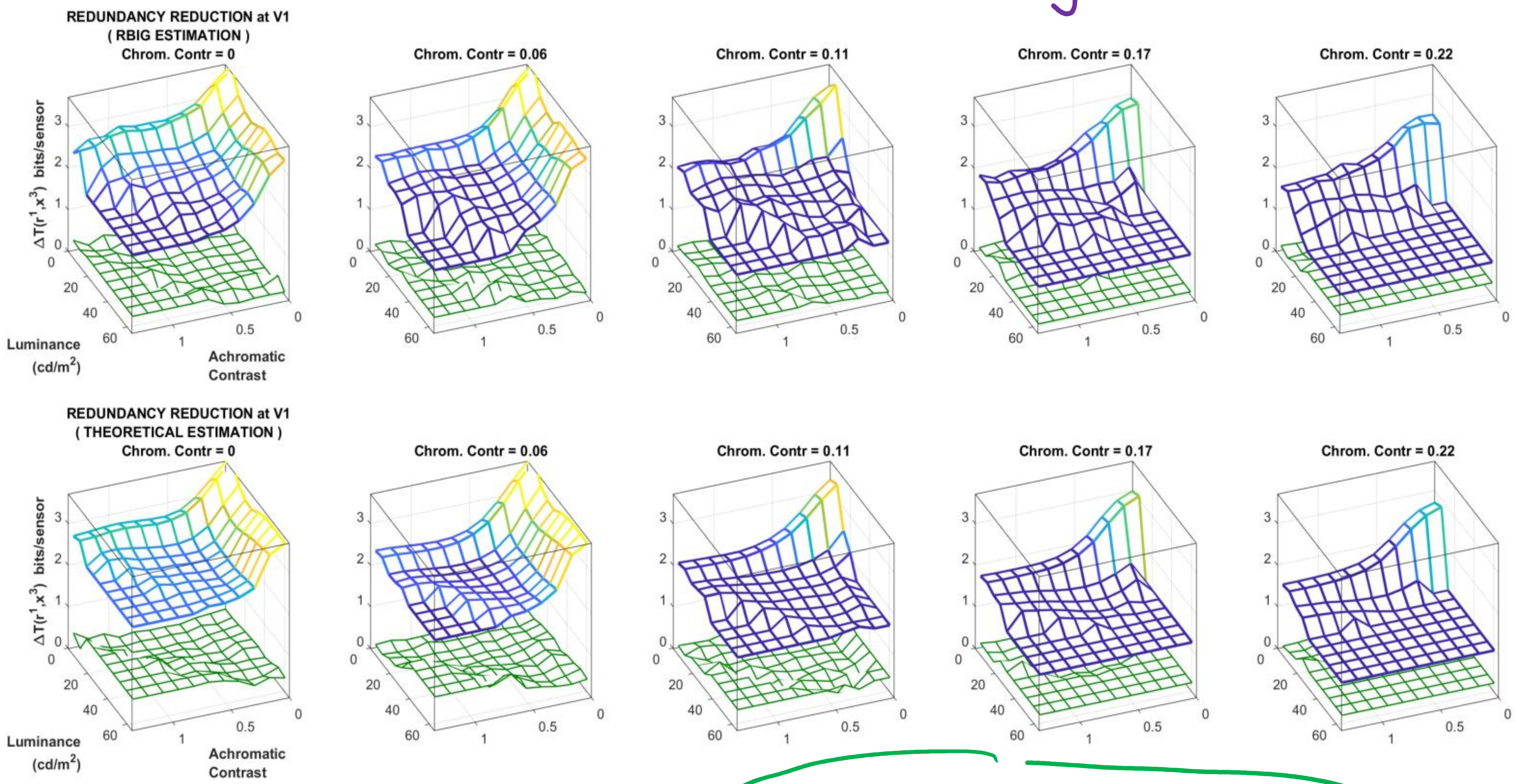
4 THE TWO SIDES OF THE EFFICIENT CODING HYPOTESIS:

Stats, $\rightleftarrows$ Biology

Redundancy Reduction

RBIG

THEORY



RBIG estimates work!

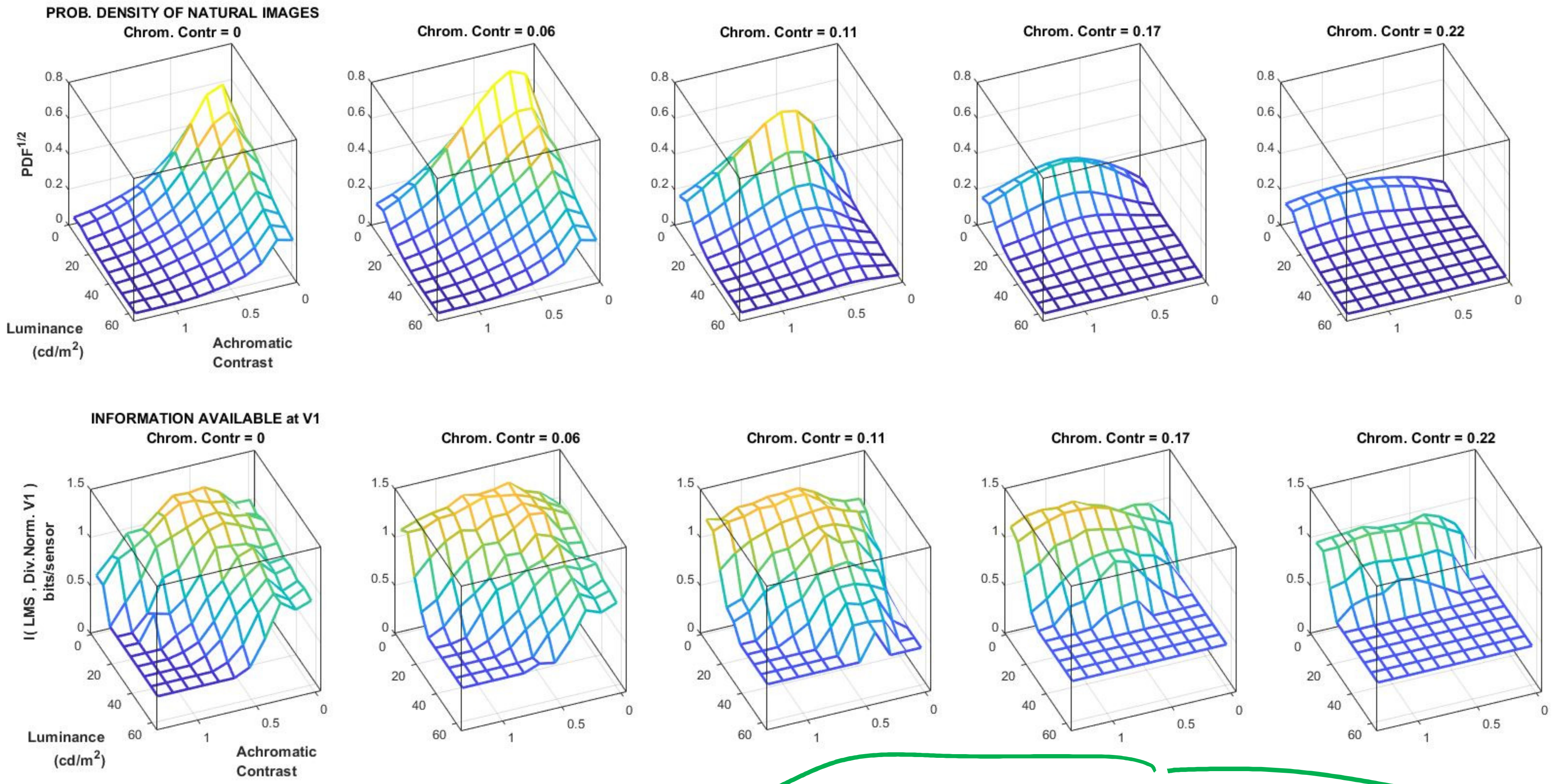
4 THE TWO SIDES OF THE EFFICIENT CODING HYPOTESIS:

Stats. ⟷ Biology

Transmitted Information

PDF Natural Images

$I(x^{in}, x^{resp})$



Transmitted info matches PDF

5 CONCLUSIONS & OPEN ISSUES

5 CONCLUSIONS & OPEN ISSUES

⑤ CONCLUSIONS & OPEN ISSUES

* The WHY question goes beyond empirical models

* It connects Artificial Intelligence with Neuroscience

Statistics $\rightleftarrows$ Biology

5 CONCLUSIONS & OPEN ISSUES

* The WHY question goes beyond empirical models

* It connects Artificial Intelligence with Neuroscience

  Statistics ⇄ Biology

* Information theory provides a framework but there are others

5 CONCLUSIONS & OPEN ISSUES

* The WHY question goes beyond empirical models

* It connects Artificial Intelligence with Neuroscience

  Statistics Biology

* Information theory provides a framework but there are others

  ICA explains linear receptive fields and their spatio. chrom. adapt.

  Classification goals explain receptive fields too!

  SPCA explains non-linearities in color / texture / motion

  Error minimization explains CSF

5 CONCLUSIONS & OPEN ISSUES

* The WHY question goes beyond empirical models

* It connects Artificial Intelligence with Neuroscience

Statistics Biology

* Information theory provides a framework but there are others

ICA explains linear receptive fields and their spatio. chrom. adapt.

Classification goals explain receptive fields too!

SPCA explains non-linearities in color / texture / motion

Error minimization explains CSF

RBIG measures show that psychophysically meaningful models have remarkable properties!

5 CONCLUSIONS & OPEN ISSUES

* The WHY question goes beyond empirical models

* It connects Artificial Intelligence with Neuroscience

  Statistics ⇄ Biology

* Information theory provides a framework but there are others

  ICA explains linear receptive fields and their spatio.chrom.adapt.

  → Classification goals explain receptive fields too!

  SPCA explains non.linearities in color/texture/motion

  Error minimization explains CSF

  ← RBIG measures show that psychophysically meaningful models have remarkable properties!

that is WHY!

5 CONCLUSIONS & OPEN ISSUES

STATISTICS & MODELS ← BIOLOGY

STATISTICS → BIOLOGY

5 CONCLUSIONS & OPEN ISSUES

STATISTICS & MODELS ← BIOLOGY

- Non reproducible behavior → improve models
  - Stronger nonlinearities
  - Resolution levels
    - Physiol.
    - Psychophys.
  - Fit models automatic differentiation

INRF
Wilson-Cowan
Div. Norm.

Bertalmío et al. **Sci. Rep.** 2020
Esteve, Bertalmío, Malo **arxiv** 2020

- Noise models
  - Discrimination
  - Information transmission

Esteve et al. **arxiv** 2020

- Better elements for deep-learning

STATISTICS → BIOLOGY

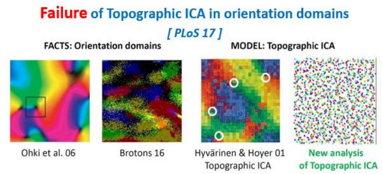
5 CONCLUSIONS & OPEN ISSUES

STATISTICS & MODELS ← BIOLOGY

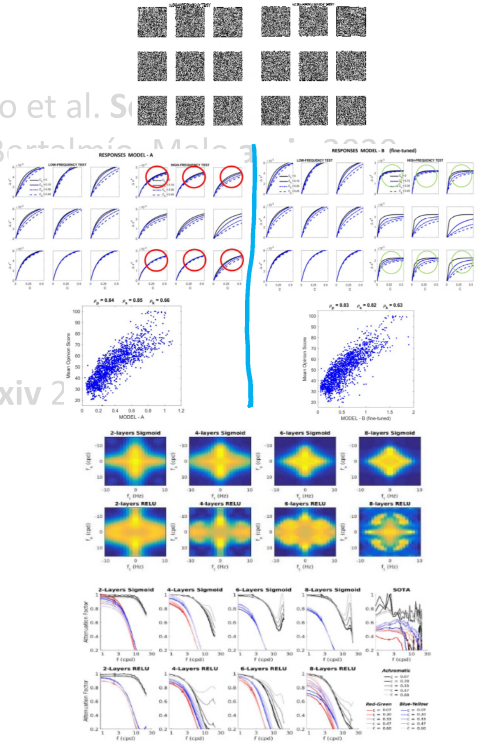
- Non reproducible behavior → improve models
  - Stronger nonlinearities
  - Resolution levels { · Physiol. · Psychophys.
  - Fit models automatic differentiation

INRF
Wilson-Cowan
Div. Norm.

- Noise models { · Discrimination · Information transmission
- Better elements for deep-learning



Bertalmío et al. Sc
Esteve, Bertalmío, Malo ... 2020



Esteve et al. arxiv 2

STATISTICS → BIOLOGY

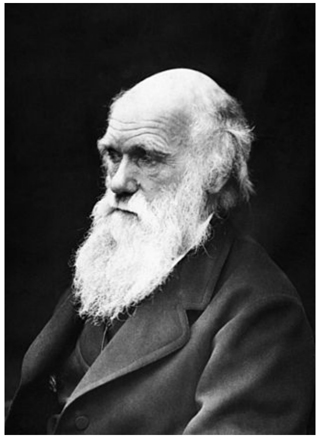
- Don't missuse deep learning!

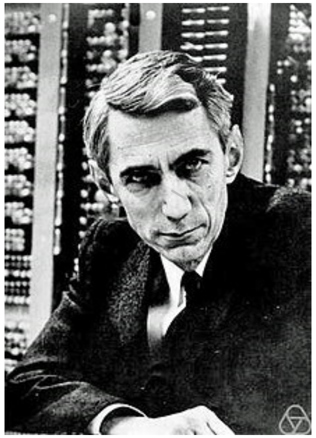
Geirhos et al. **Nature** 2020
Martinez et al. **PLOS** 2017
Martinez et al. **Front. Neurosci.** 2019
Li, Gomez, Bertalmio & Malo, **submit. JoV**. 2021

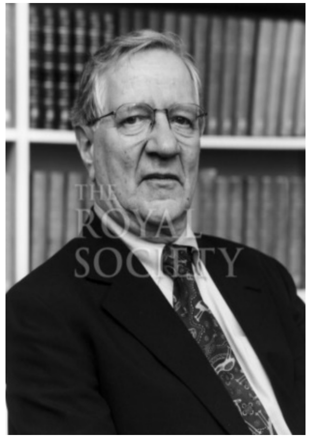
- Model-based new psychophysics { · Geometry MAD · Optimal stimuli

Malo & Simoncelli **SPIE** 2015
Martinez et al. **PLOS** 2018


QUESTIONS?
COMMENTS?






THE ROYAL SOCIETY