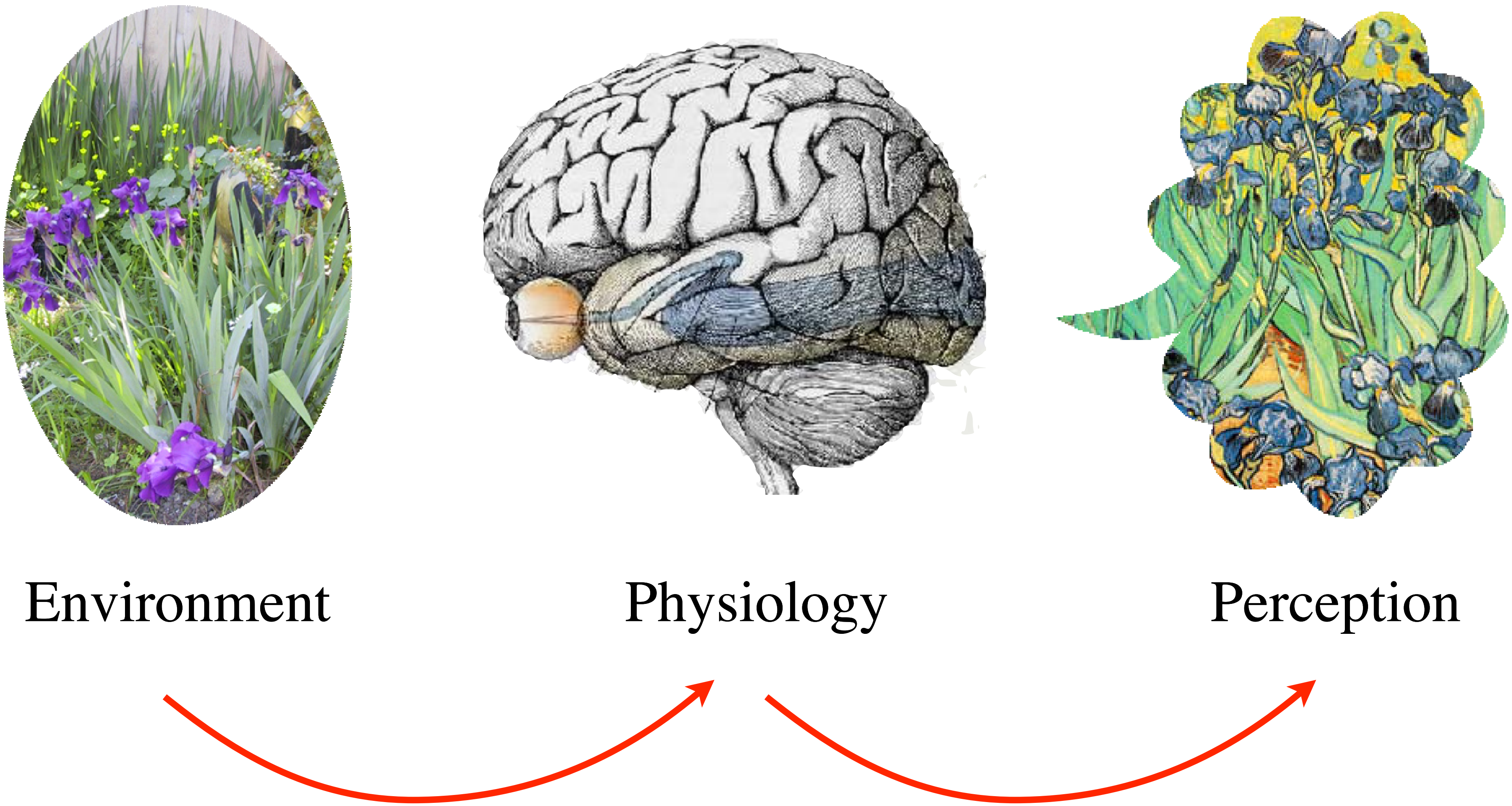
Efficient distribution of resources provides an embedding of environmental statistics

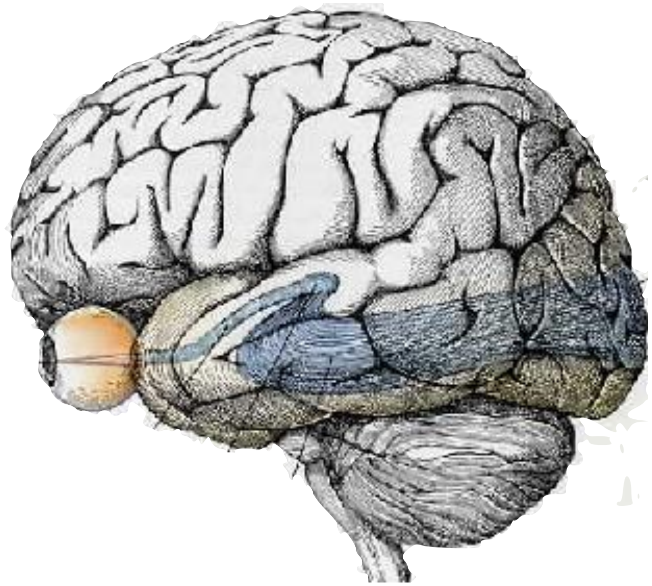
Eero Simoncelli

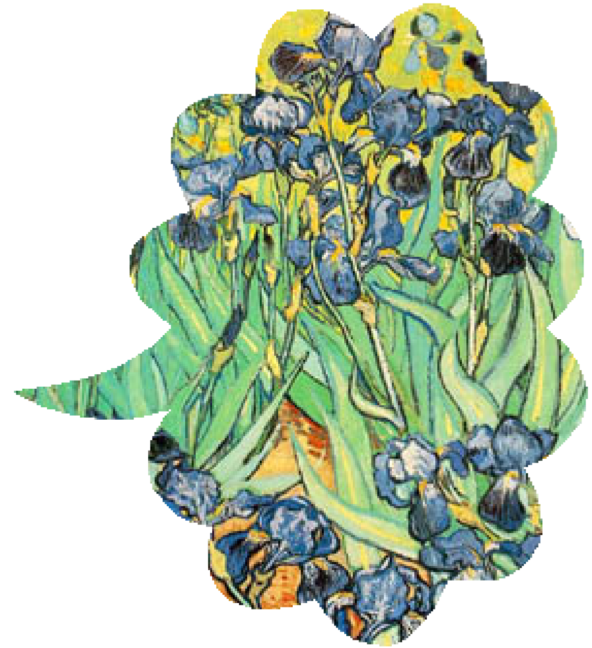
Howard Hughes Medical Institute
Center for Neural Science,
Courant Inst. of Mathematical Sciences,

New York University



[images: Hubel 1995; van Gogh 1889, c/o Getty]



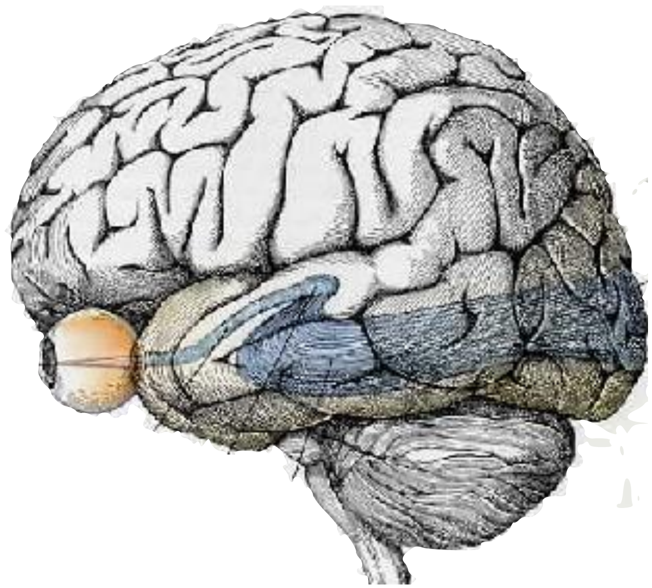


Environment Physiology Perception

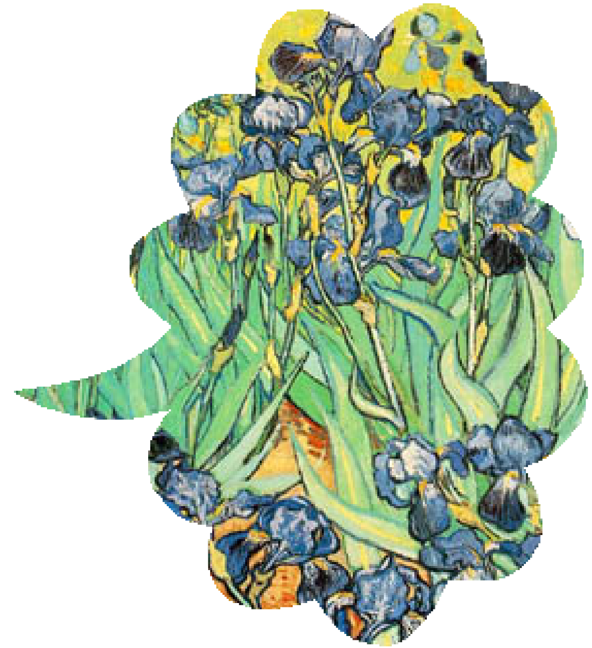
Efficient coding: Sensory systems allocate resources efficiently to capture information about inputs that are likely to occur.

[Barlow '61]

Environment



Physiology

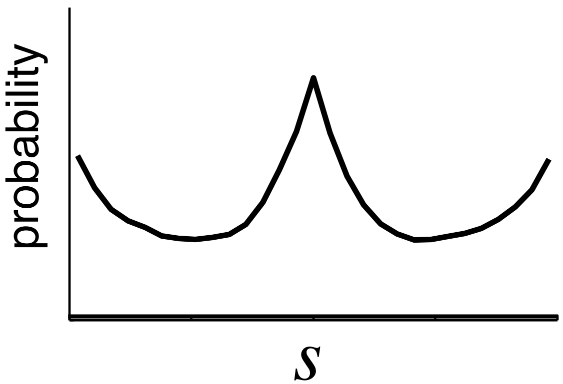


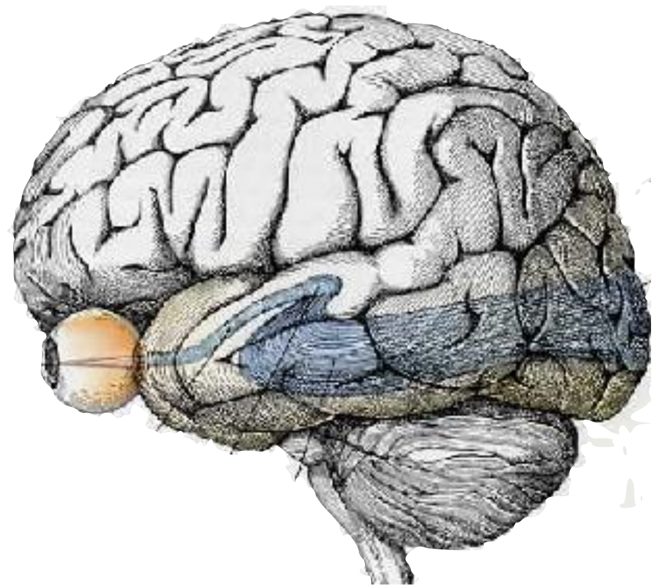
Perception

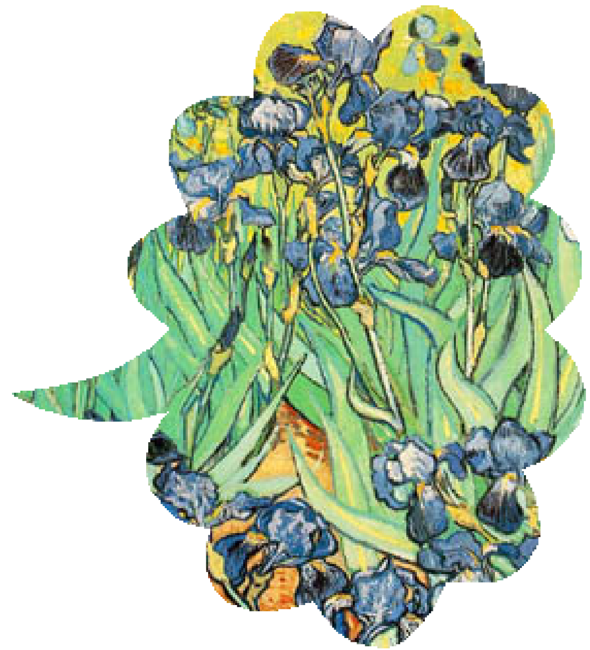
Optimal inference: Perception is a best guess as to what is in the world, given current sensory responses and prior experience

[Al Hazan, 1040; Helmholtz, 1866]

For today's talk:

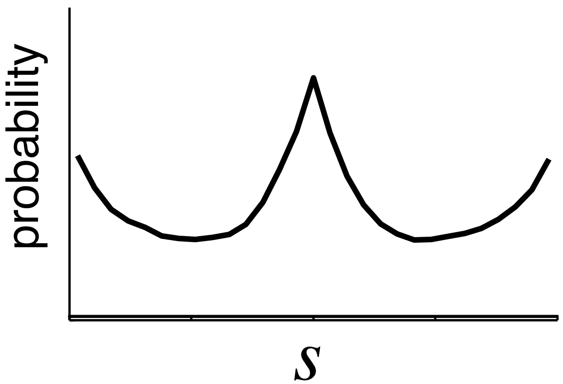


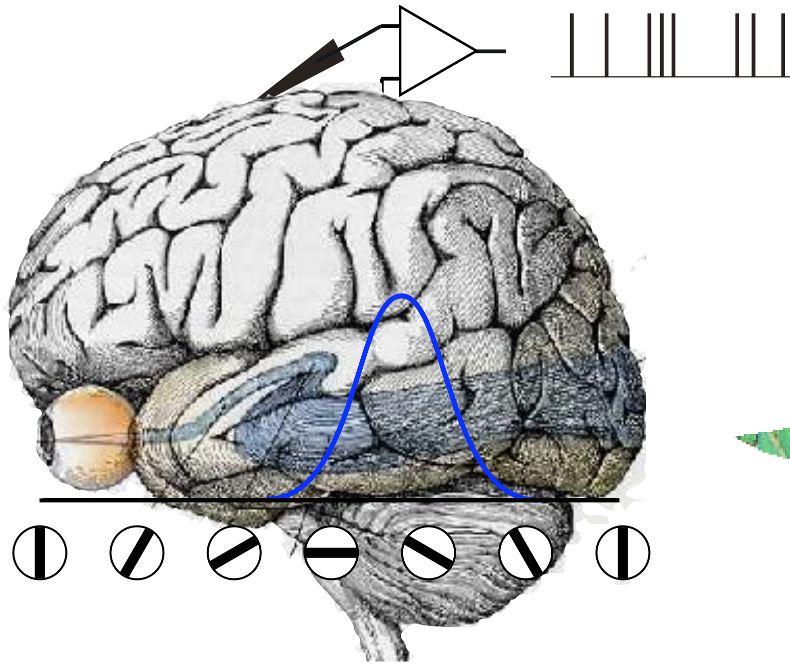


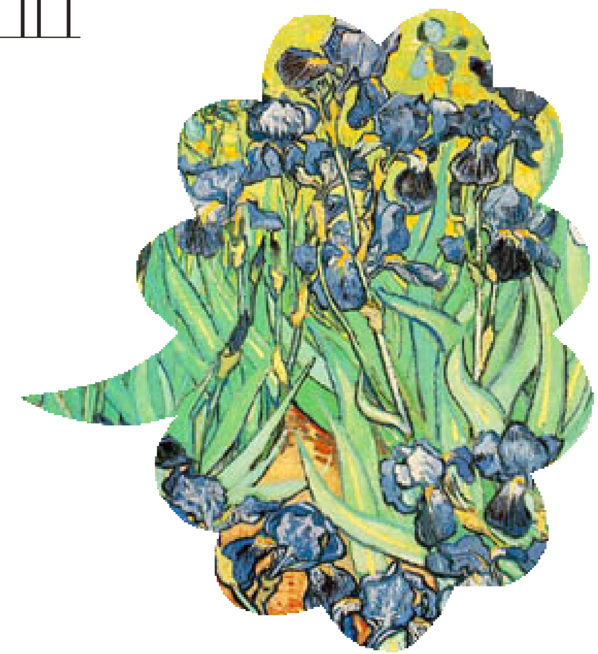


Environment Physiology Perception

For today's talk:





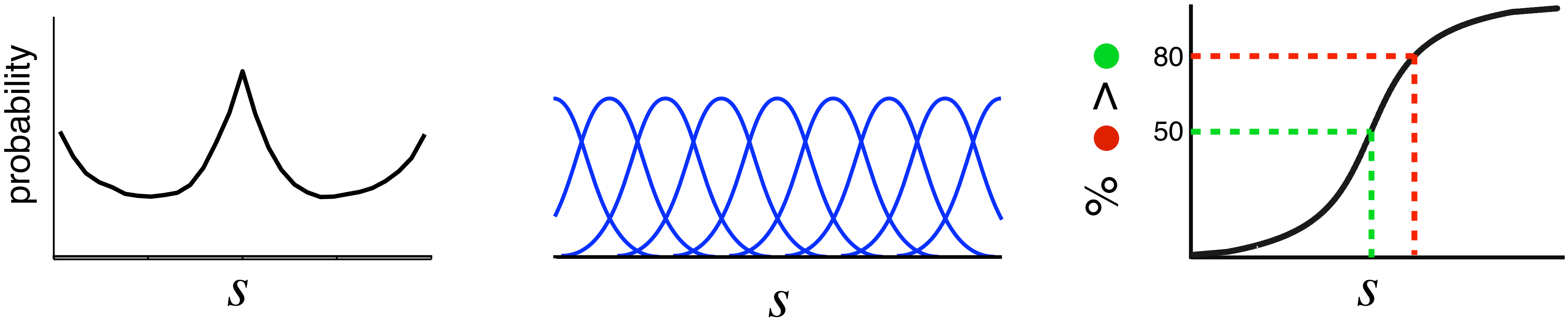


Environment

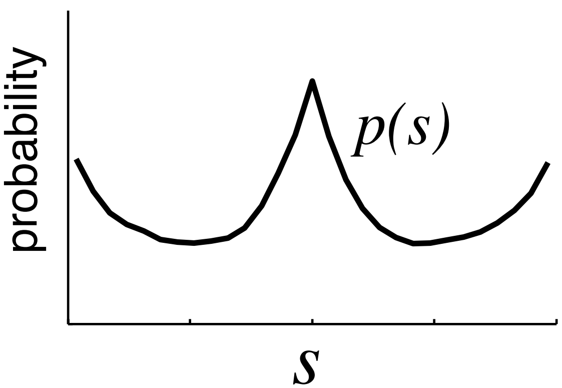
Physiology

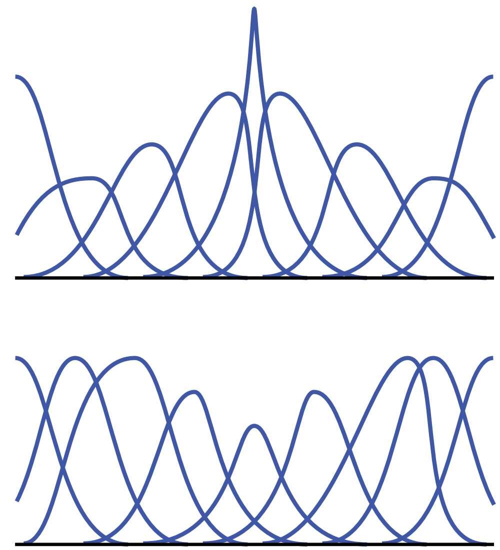
Perception

For today's talk:



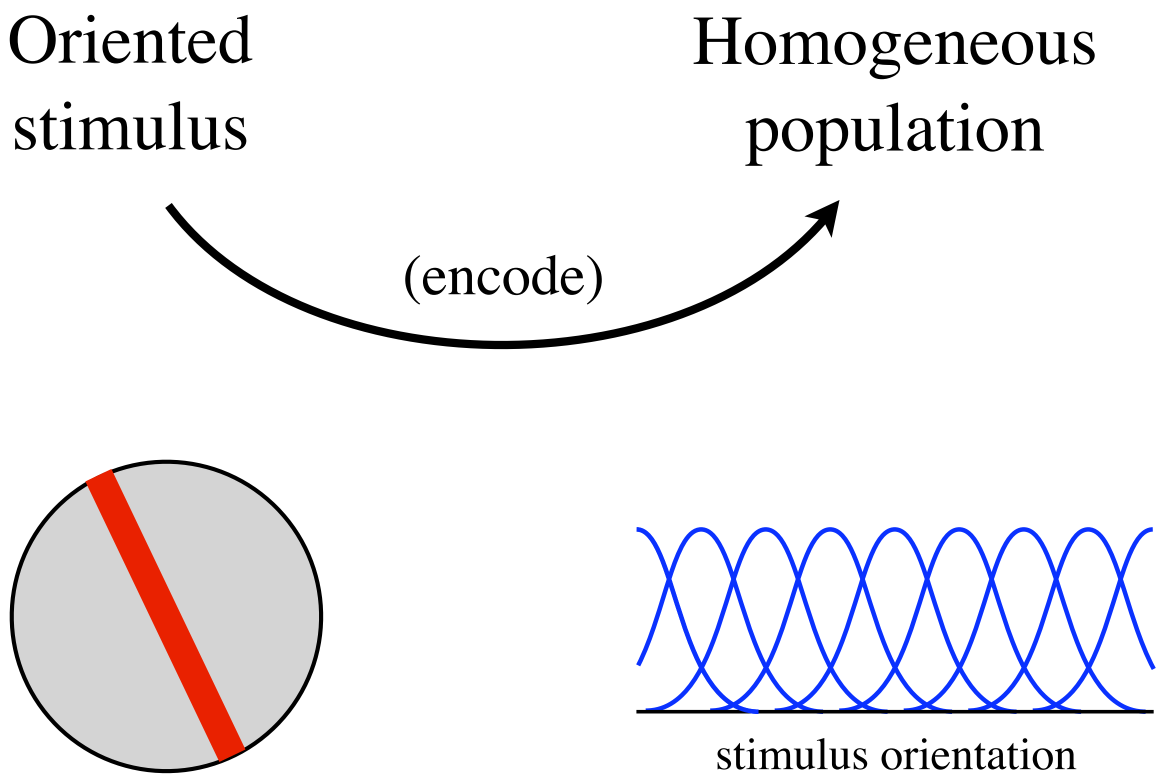
Environment Physiology Perception





Given N neurons, with total average firing rate R

- what set of tuning curves transmits the most information about s?
- what are the perceptual implications of this solution?


Oriented stimulus
Homogeneous population
(encode)
stimulus orientation

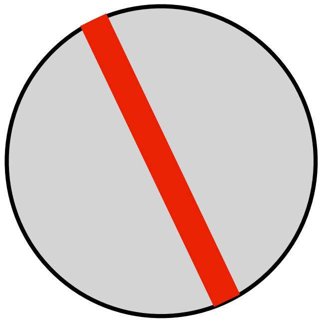
Oriented stimulus

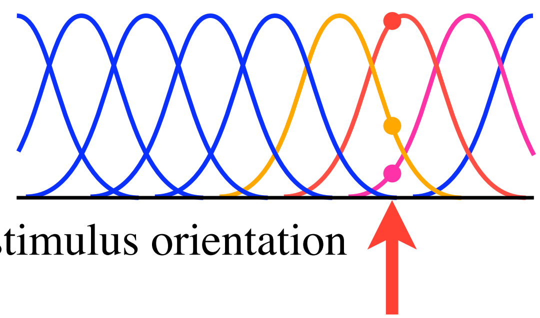
Homogeneous population

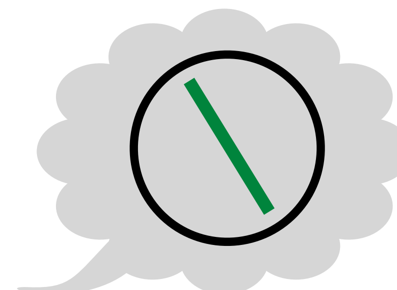
"Perceived" orientation

(encode)

(decode)



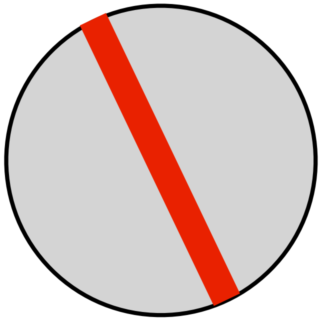


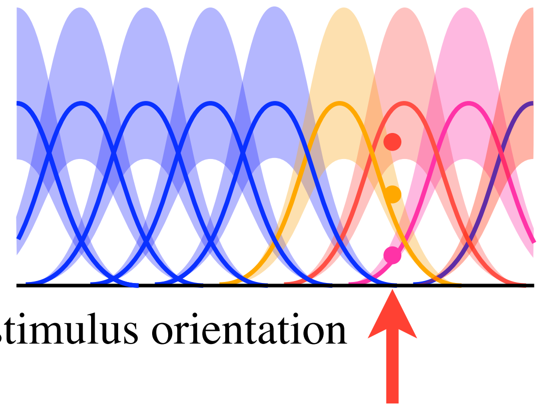


[Földiák, 1989;
Seung & Sompolinsky, 1993;
Salinas & Abbott, 1994;
Sanger, 1996;
Zemel, Dayan, Pouget, 1998;
Kang & Sejnowski, 1999;
Brown & Bäcker, 2006;
Jazayeri & Movshon, 2006;
Ma et. al., 2006; etc]

Oriented stimulus → Homogeneous population → "Perceived" orientation

(encode) (decode)

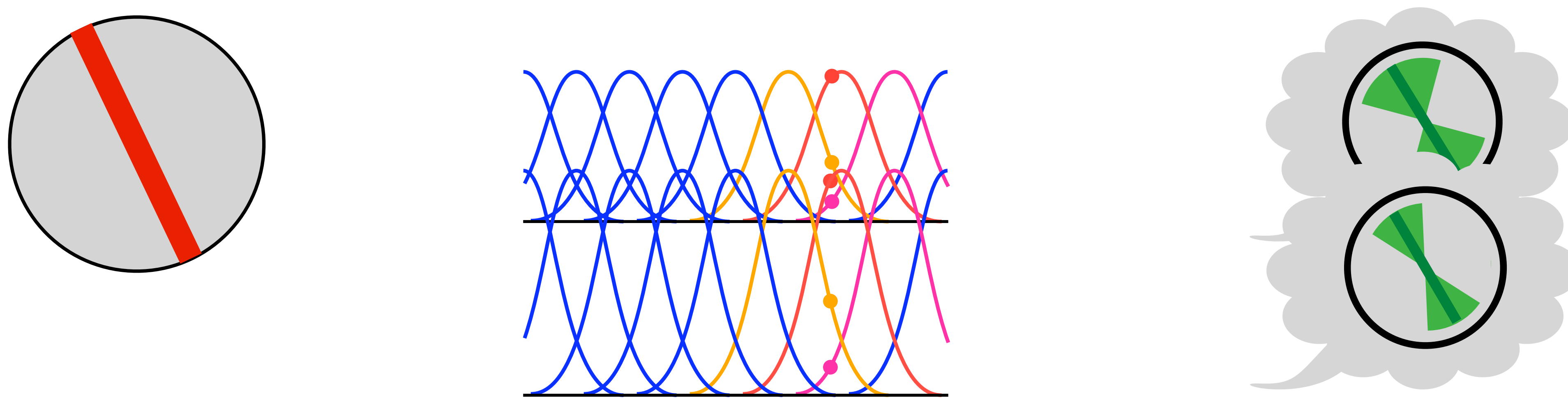




Accuracy limited by noise ...

[Földiák, 1989;
Seung & Sompolinsky, 1993;
Salinas & Abbott, 1994;
Sanger, 1996;
Zemel, Dayan, Pouget, 1998;
Kang & Sejnowski, 1999;
Brown & Bäcker, 2006;
Jazayeri & Movshon, 2006;
Ma et. al., 2006; etc]

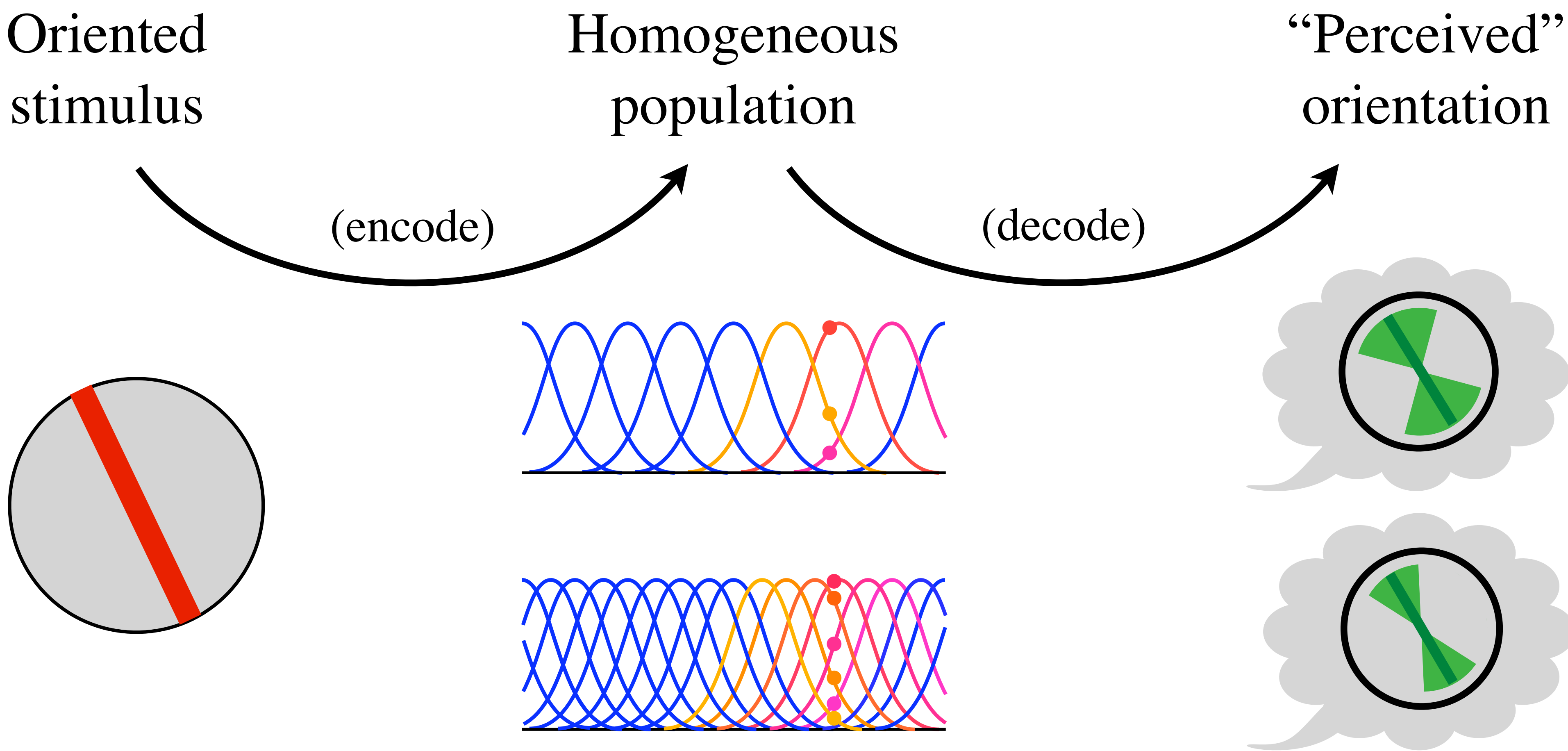
Oriented stimulus → (encode) → Homogeneous population → (decode) → "Perceived" orientation



Accuracy limited by noise ...

... but improves with additional resources:

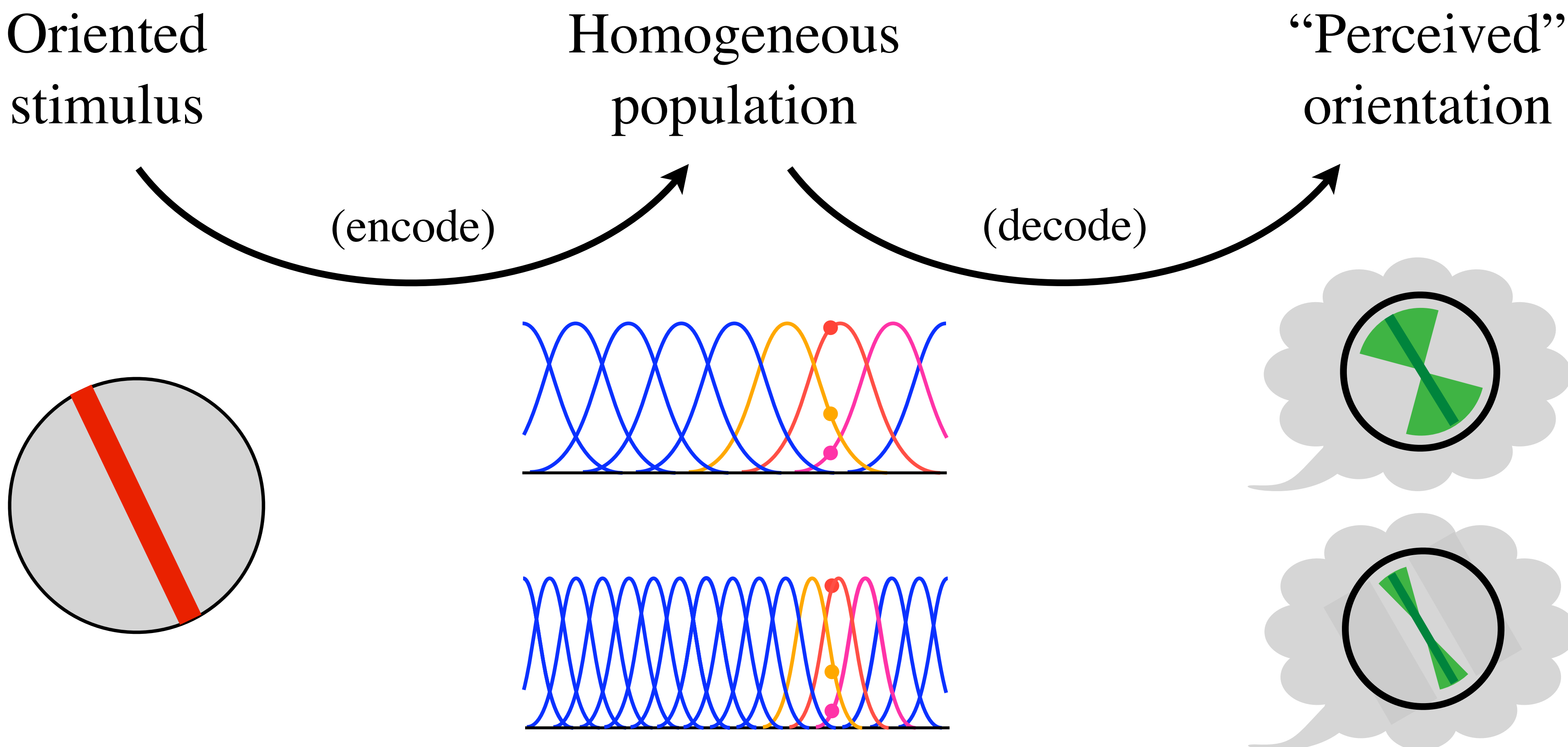
- Max firing rate



Accuracy limited by noise ...

... but improves with additional resources:

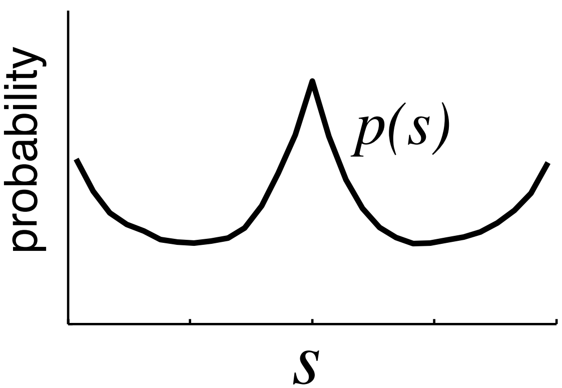
- Max firing rate
- Number of cells

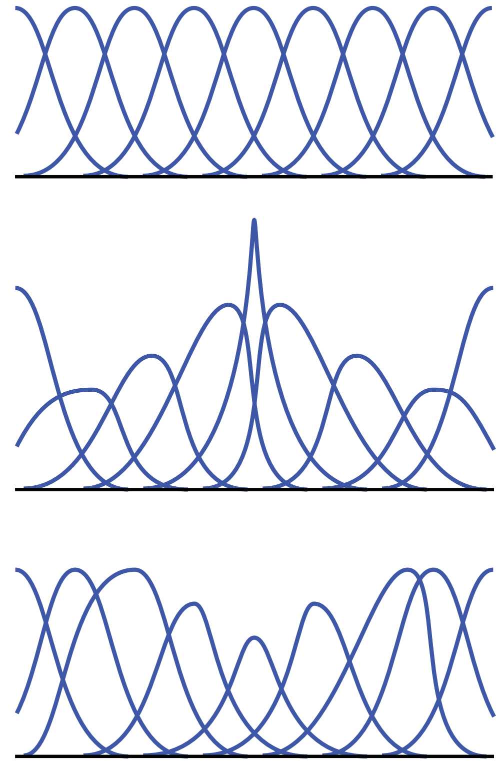


Accuracy limited by noise ...

... but improves with additional resources:

- Max firing rate
- Number of cells (& tuning width)





Given N neurons, with total average firing rate R

- what set of tuning curves transmits the most information about s?
- what are the perceptual implications of this solution?

[Ganguli & Simoncelli, 2014]

1. Independent Poisson responses [e.g., Seung & Sompolinsky, 1993]

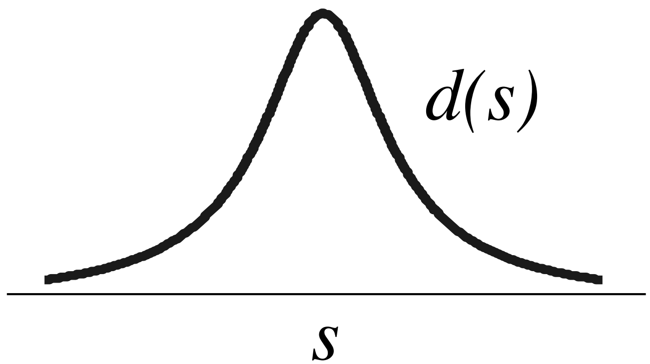
$$p(\vec{r}|s) = \prod_{n=1}^{N} \frac{h_n(s)^{r_n} \; e^{-h_n(s)}}{r_n!}$$

(generalizes to exponential families, w/ local correlations)

[cf., Jazayeri & Movshon, 2006; Ma et. al., 2006]

1. Independent Poisson responses
(generalizes to exponential families, w/ local correlations)

2. Population resource allocation parameterized by **density** and **gain**, with enforced "tiling":

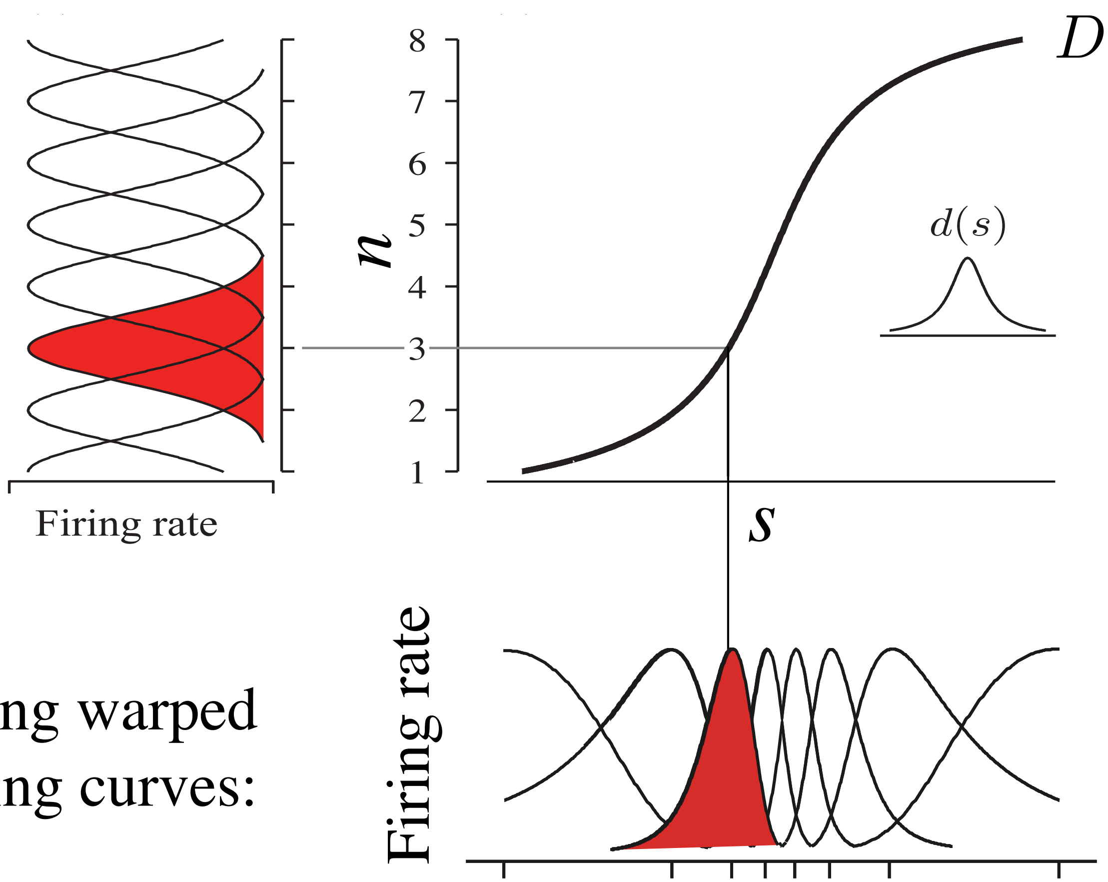


Tuning curve density function

1. Independent Poisson responses
(generalizes to exponential families, w/ local correlations)

2. Population of tuning curves parameterized by **density** and **gain**, with enforced “tiling”:

$$D(s) = \int_{-\infty}^{s} d(x)\, \mathrm{d}x$$

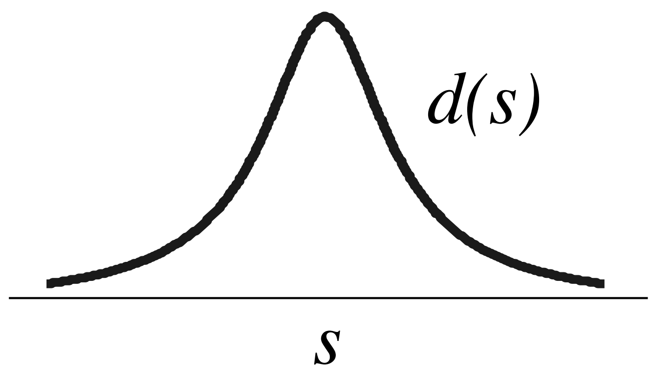


Resulting warped tuning curves:

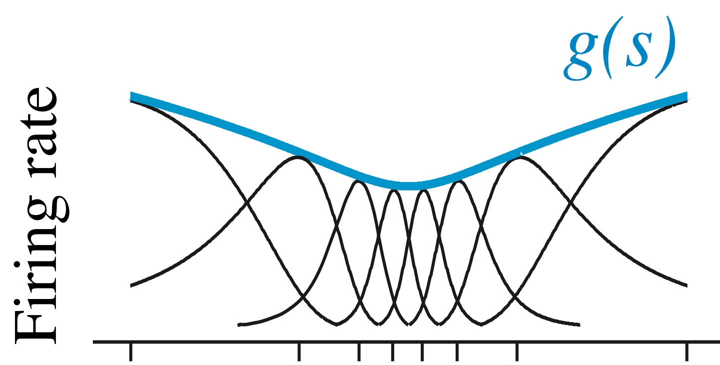
$$h(D(s) - n)$$

1. Independent Poisson responses (generalizes to exponential families, w/ local correlations)

2. Population of tuning curves parameterized by **density** and **gain**, with enforced “tiling”:



Tuning curve density function



Tuning curve gain function

Resulting **warped** and **modulated** tuning curves:

$$h_n(s) = g(s)\, h(D(s) - n)$$

1. Independent Poisson responses
   (generalizes to exponential families, w/ local correlations).

2. Parameterize tuning curve population in terms of **gain** and **density**, with enforced "tiling".

3. Optimize (Fisher bound on) transmitted information, for *N* neurons with total average firing rate *R* [Brunel & Nadal '98]:

$$I_f(s) = \int p(\vec{r}|s)\, \frac{\partial^2}{\partial s^2} \log p(\vec{r}|s)\, \mathrm{d}\vec{r}$$

$$\propto d^2(s) g(s) \qquad \text{(tiling parameterization)}$$

1. Independent Poisson responses (generalizes to exponential families, w/ local correlations).

2. Parameterize tuning curve population in terms of **gain** and **density**, with enforced "tiling".

3. Optimize (Fisher bound on) transmitted information, for N neurons with total average firing rate R [Brunel & Nadal '98]:

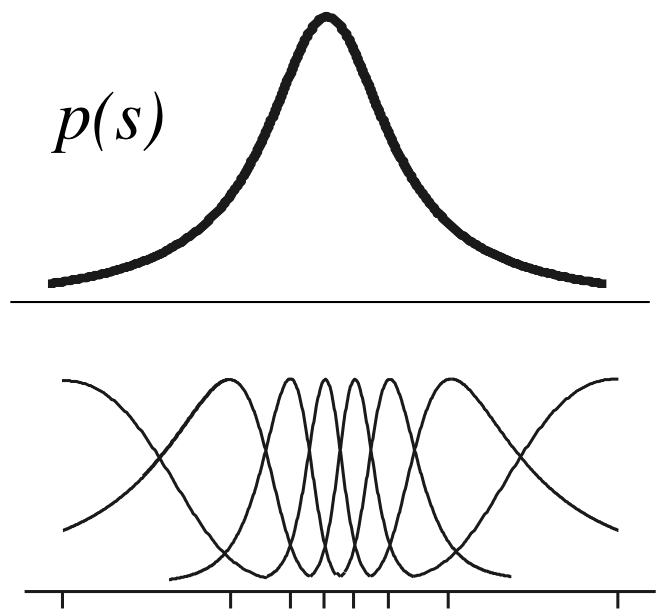
$$\arg\max_{d,g} \int p(s) \ \log\left(d^2(s)g(s)\right) \mathrm{d}s$$

such that $$\int d(s)\,\mathrm{d}s = N, \qquad \int p(s)g(s)\mathrm{d}s = R$$

Closed-form optimal solution

$$d(s) = Np(s), \quad g(s) = R$$

- More cells, with narrower tuning, in higher probability regions of stimulus space
- All neurons have same gain (max. firing rate)



[Ganguli & Simoncelli, 2014]

Closed-form optimal solution

$$d(s) = N p(s), \quad g(s) = R$$

physiological prediction

- More cells, with narrower tuning, in higher probability regions of stimulus space
- All neurons have same gain (max. firing rate)
- All neurons have same **mean** firing rate
- Bonus: limit on perceptual discrimination thresholds:

[Seriés, Stocker & Simoncelli 2009]

$$\delta(s) \geq \frac{\alpha}{\sqrt{R}\, N\, p(s)}$$

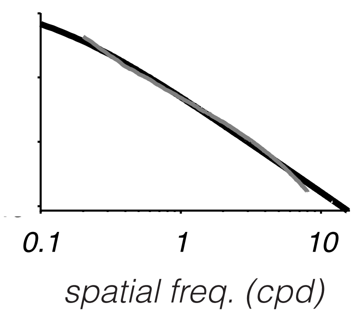
perceptual prediction

[Ganguli & Simoncelli, 2014]

Example: local spatial frequency

Environment

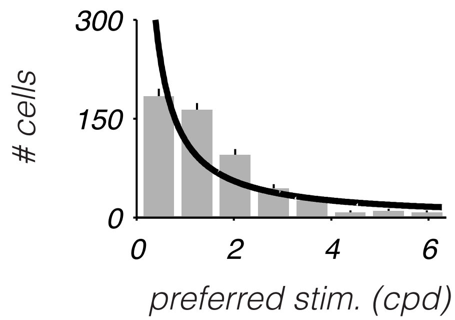
p(s)



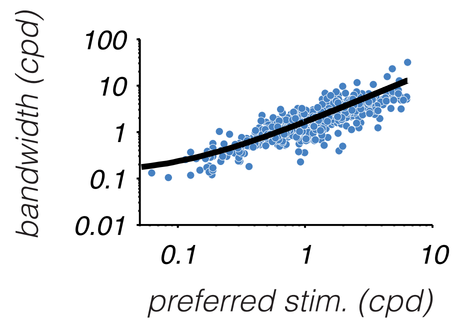
(photographs)
[van Hateren et. al. 98]
[Doi et. al. 02]
[Olmos et. al. 05]

Physiology

density



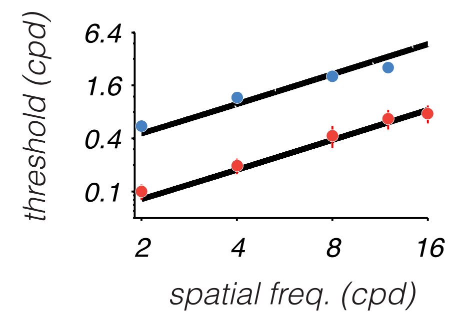
tuning width



(anesthetized macaque, V1)
[Cavanaugh et. al, 02]

Perception

discrim. thresh.

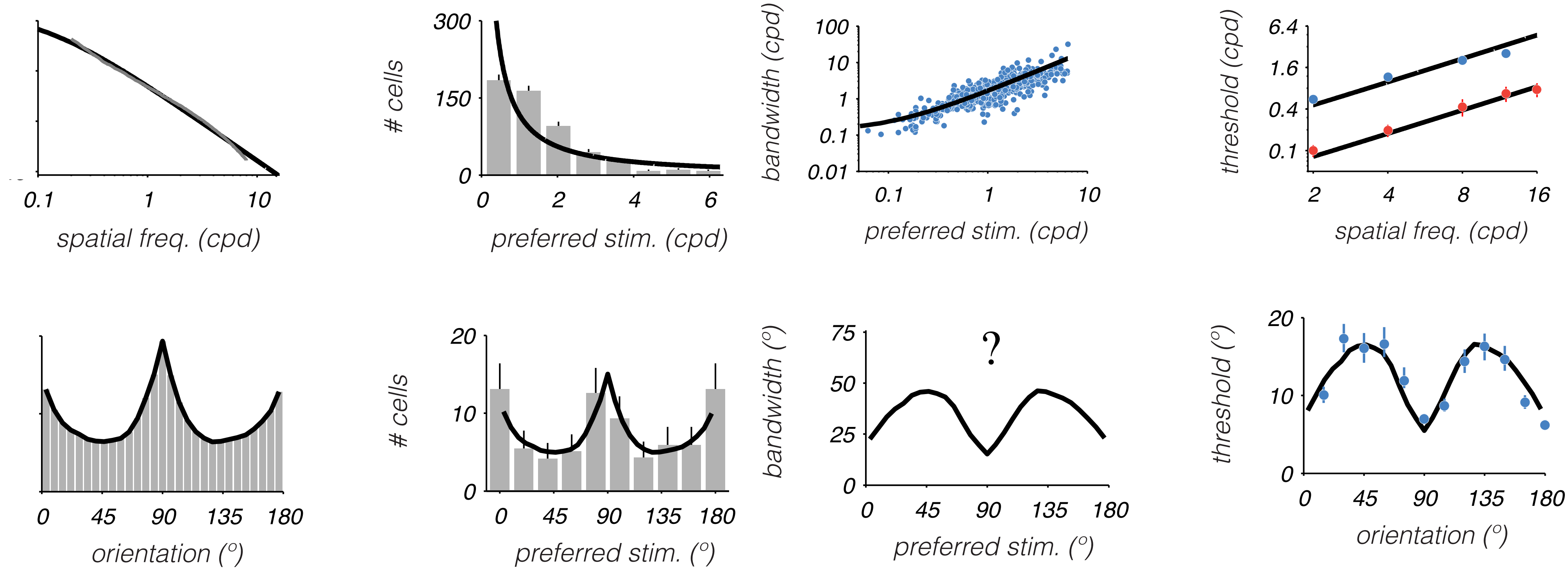


(human)
[Caelli et. al, 83]
[Regan et. al, 82]

[Ganguli & Simoncelli, Arxiv 2016]

Environment | Physiology | Perception

p(s) | density | tuning width | discrim. thresh.

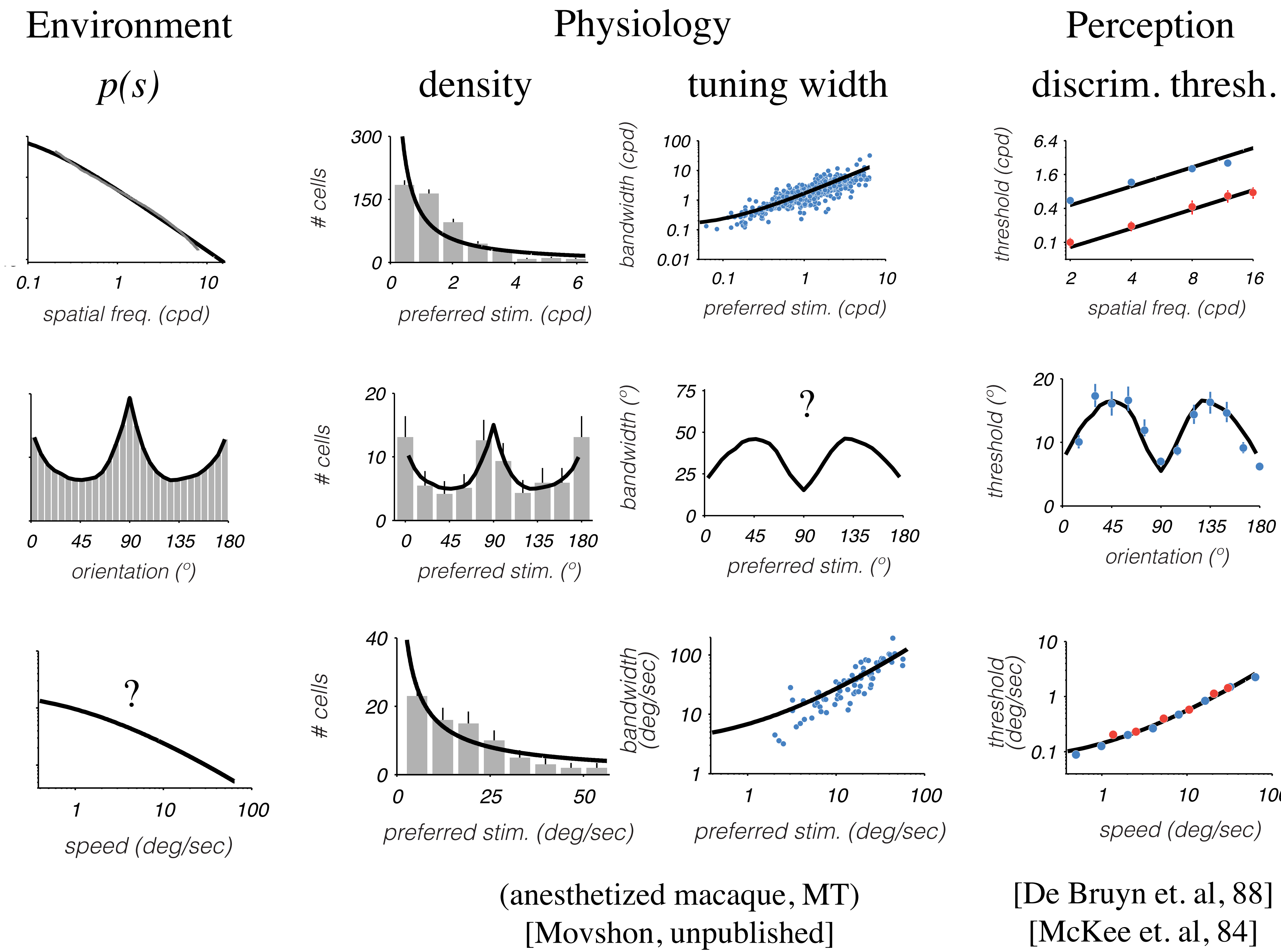


(photographs)
[Ganguli & Simoncelli, 2011]

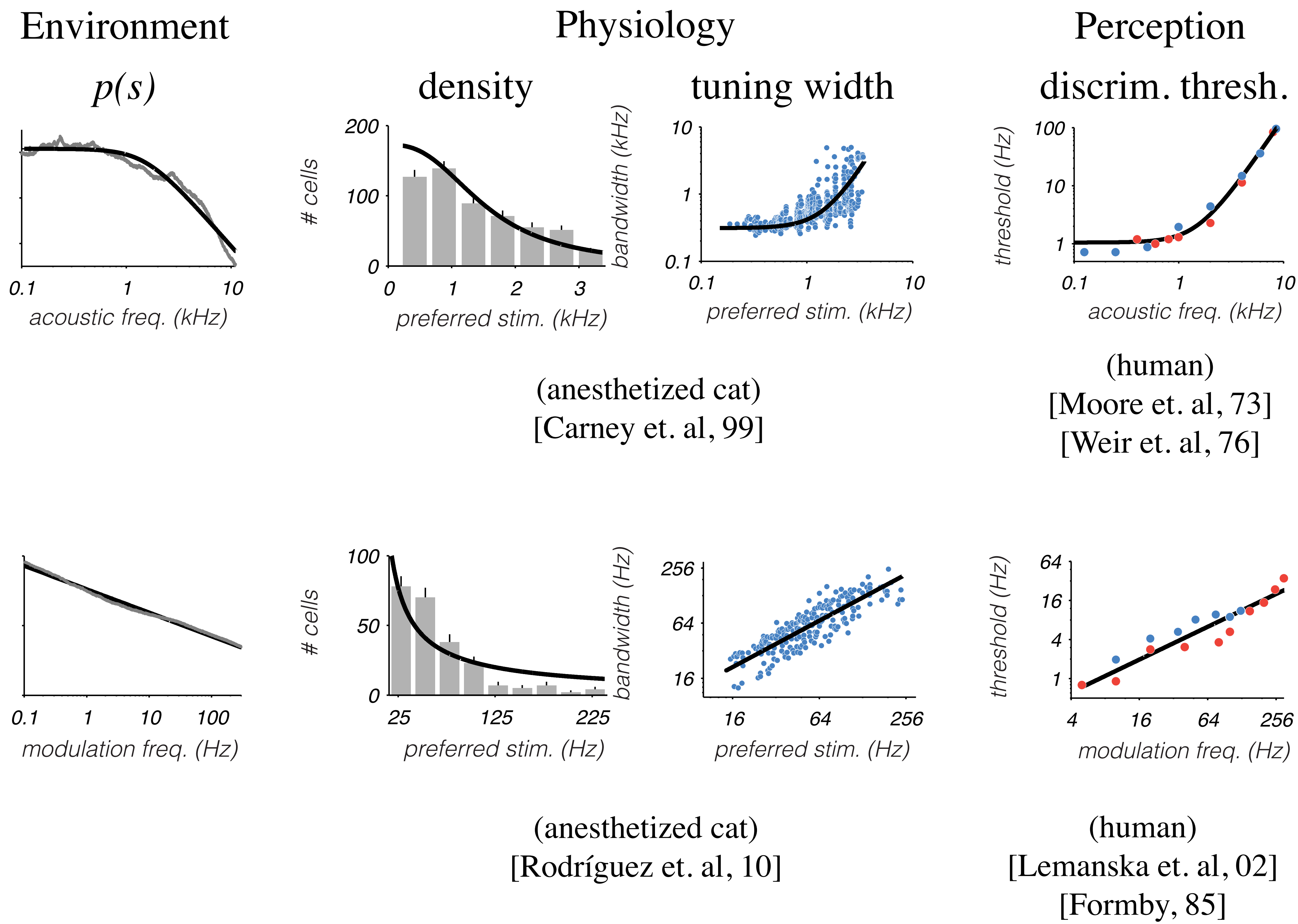
(anesthetized macaque, V1)
[Mansfield, 74]

(human)
[Girshick et. al, 11]

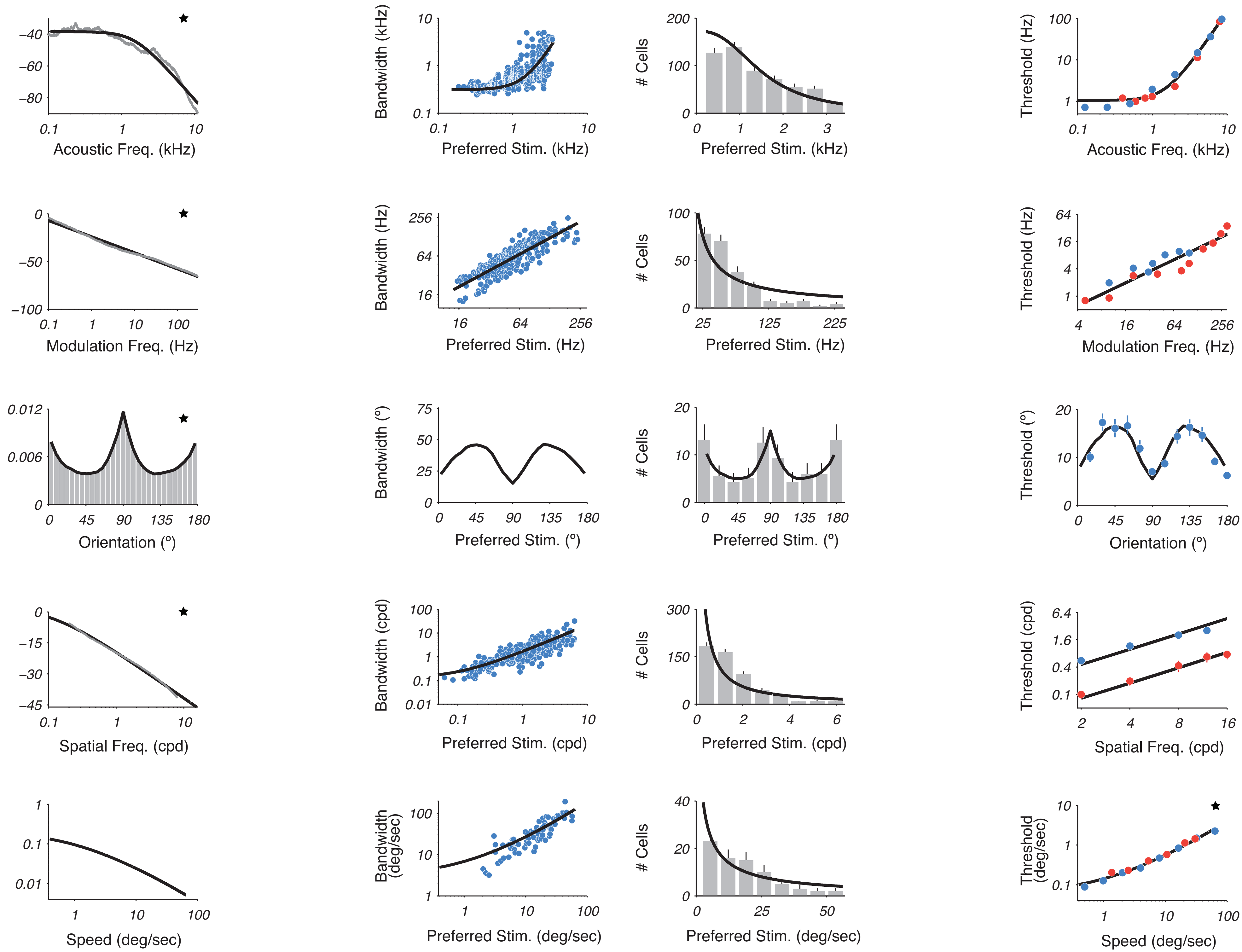
[Ganguli & Simoncelli, Arxiv 2016]



Environment
p(s)
Physiology
density
tuning width
Perception
discrim. thresh.
0.1
1
10
spatial freq. (cpd)
300
150
0
cells
0
2
4
6
preferred stim. (cpd)
bandwidth (cpd)
100
10
1
0.1
0.01
preferred stim. (cpd)
threshold (cpd)
6.4
1.6
0.4
0.1
2
4
8
16
spatial freq. (cpd)
0
45
90
135
180
orientation (°)
20
10
cells
preferred stim. (°)
bandwidth (°)
75
50
25
?
preferred stim. (°)
threshold (°)
orientation (°)
?
speed (deg/sec)
40
cells
25
50
preferred stim. (deg/sec)
bandwidth (deg/sec)
preferred stim. (deg/sec)
threshold (deg/sec)
speed (deg/sec)
(anesthetized macaque, MT)
[Movshon, unpublished]
[De Bruyn et. al, 88]
[McKee et. al, 84]



Environment
p(s)
Physiology
density
tuning width
Perception
discrim. thresh.
0.1
1
10
acoustic freq. (kHz)
200
100
0
cells
0
1
2
3
preferred stim. (kHz)
bandwidth (kHz)
10
1
0.1
0.1
1
10
preferred stim. (kHz)
threshold (Hz)
100
10
1
0.1
1
10
acoustic freq. (kHz)
(anesthetized cat)
[Carney et. al, 99]
(human)
[Moore et. al, 73]
[Weir et. al, 76]
0.1
1
10
100
modulation freq. (Hz)
100
50
0
cells
25
125
225
preferred stim. (Hz)
bandwidth (Hz)
256
64
16
16
64
256
preferred stim. (Hz)
threshold (Hz)
64
16
4
1
4
16
64
256
modulation freq. (Hz)
(anesthetized cat)
[Rodríguez et. al, 10]
(human)
[Lemanska et. al, 02]
[Formby, 85]



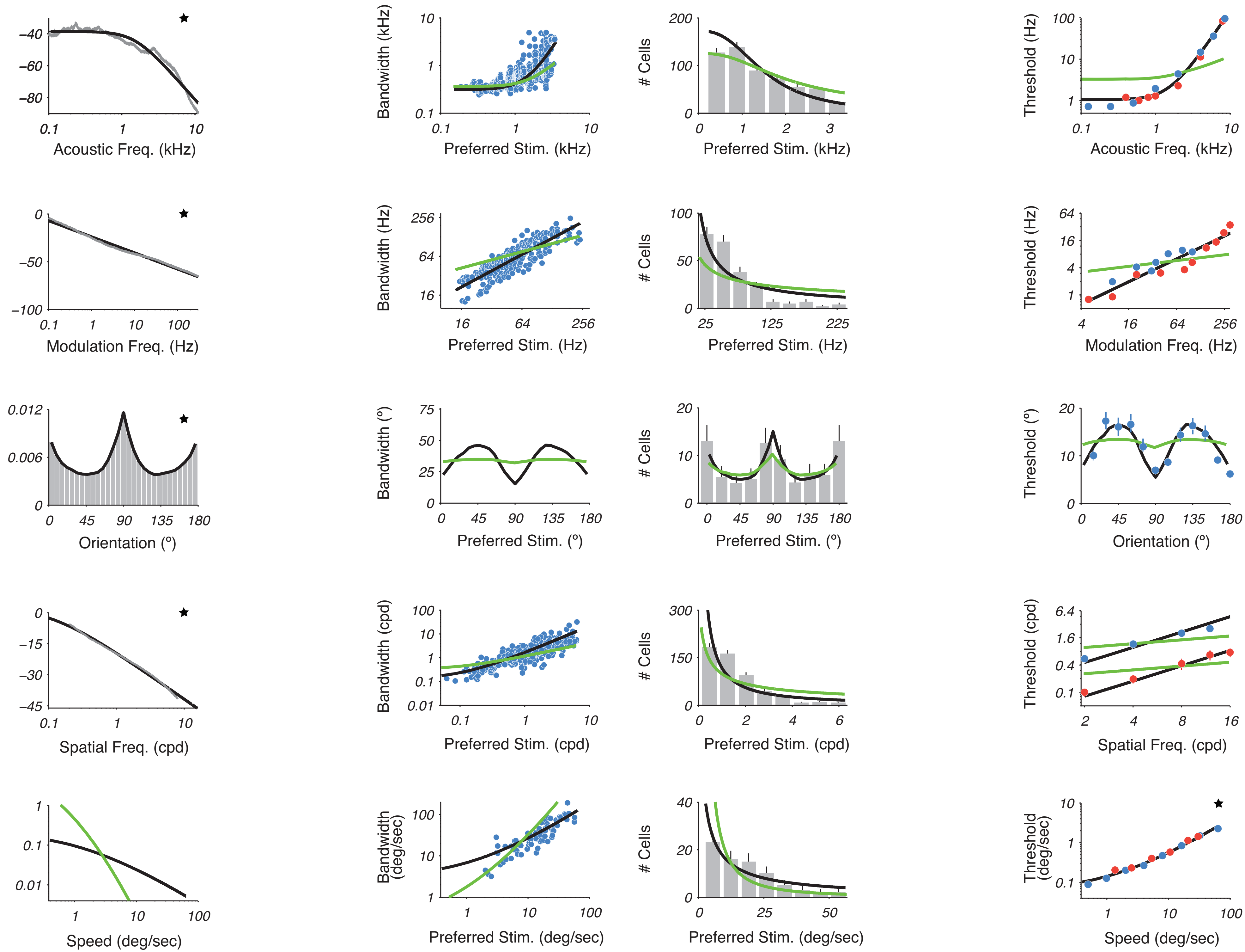
Acoustic Freq. (kHz)
Bandwidth (kHz)
Preferred Stim. (kHz)
Cells
Threshold (Hz)
Modulation Freq. (Hz)
Bandwidth (Hz)
Preferred Stim. (Hz)
Orientation (°)
Bandwidth (°)
Preferred Stim. (°)
Threshold (°)
Spatial Freq. (cpd)
Bandwidth (cpd)
Preferred Stim. (cpd)
Threshold (cpd)
Speed (deg/sec)
Bandwidth (deg/sec)
Preferred Stim. (deg/sec)
Threshold (deg/sec)

More generally:

$$\arg\max_{d(s),g(s)} \int p(s)\ f\left(d^2(s)g(s)\right)\ \mathrm{d}s$$

$$\text{s.t.} \quad \int d(s)\mathrm{d}s = N, \quad \int p(s)g(s)\mathrm{d}s = R$$

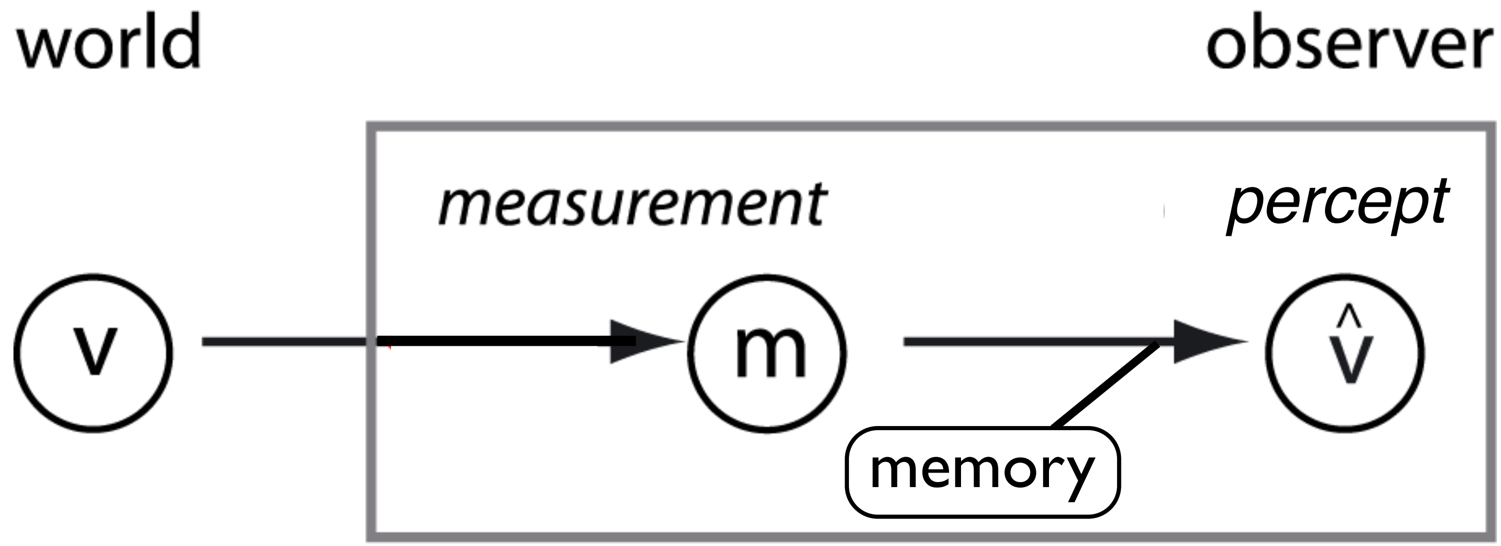
		Infomax	Discrimax	Power law
	Optimized function:	$f(x) = \log x$	$f(x) = -x^{-1}$	$f(x) = -x^{\alpha},\ \alpha < \frac{1}{3}$
Density (Tuning width)$^{-1}$	$d(s)$	$Np(s)$	$\propto Np^{\frac{1}{2}}(s)$	$\propto Np^{\frac{\alpha-1}{3\alpha-1}}(s)$
Gain	$g(s)$	R	$\propto Rp^{-\frac{1}{2}}(s)$	$\propto Rp^{\frac{2\alpha}{1-3\alpha}}(s)$
Fisher information	$I_f(s)$	$\propto RN^2p^2(s)$	$\propto RN^2p^{\frac{1}{2}}(s)$	$\propto RN^2p^{\frac{2}{1-3\alpha}}(s)$
Discriminability bound	$\delta_{\min}(s)$	$\propto p^{-1}(s)$	$\propto p^{-\frac{1}{4}}(s)$	$\propto p^{\frac{1}{3\alpha-1}}(s)$



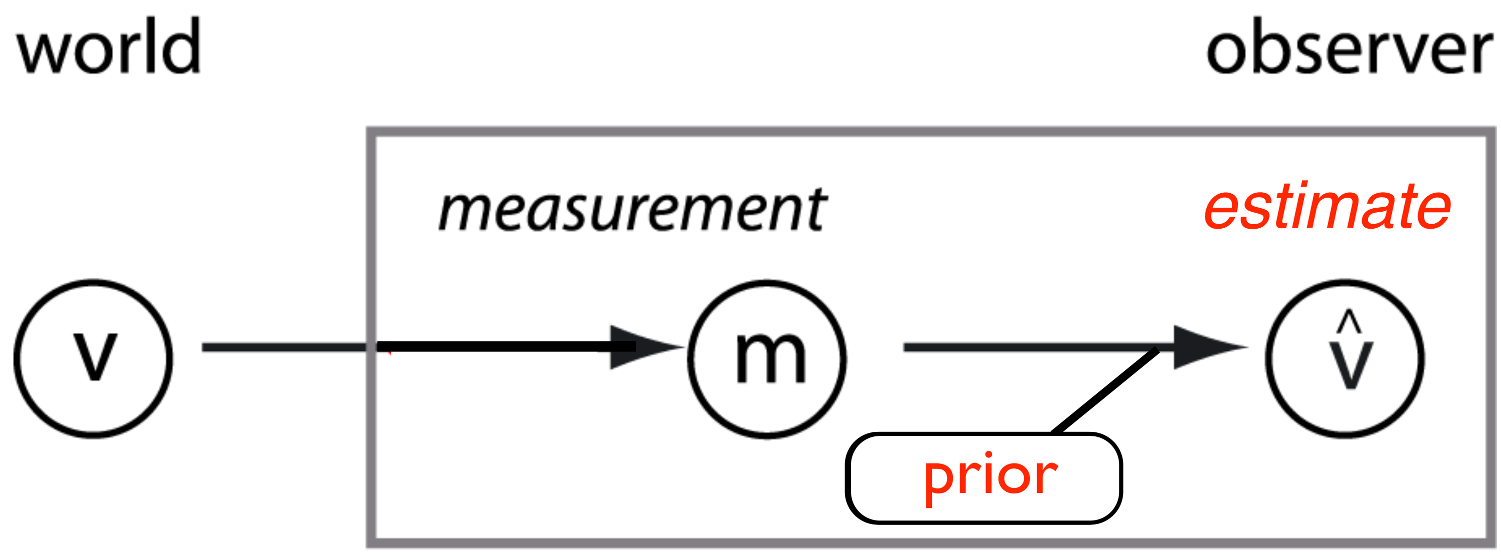
Acoustic Freq. (kHz)
Preferred Stim. (kHz)
Bandwidth (kHz)
Cells
Threshold (Hz)
Modulation Freq. (Hz)
Bandwidth (Hz)
Preferred Stim. (Hz)
Orientation (°)
Bandwidth (°)
Preferred Stim. (°)
Threshold (°)
Spatial Freq. (cpd)
Bandwidth (cpd)
Preferred Stim. (cpd)
Threshold (cpd)
Speed (deg/sec)
Bandwidth (deg/sec)
Preferred Stim. (deg/sec)
Threshold (deg/sec)

Helmholtz (1866)

Perception is our best guess as to what is in the world, given our current sensory input and our prior experience [paraphrased]



Bayes (1750)



Bayesian perception

- Shading/lighting
  [Kersten 90; Knill, Kersten, Yuille 96; Mamassian, Landy, Maloney 01]
- Retinal speed/velocity
  [Simoncelli et. al. 91; Heeger & Simoncelli 93; Weiss etal. 02; Stocker & Simoncelli 06]
- Surface orientation [Bülthoff & Yuille 96; Saunders & Knill 01]
- Color constancy [Brainard & Freeman 97]
- Contours [Geisler, Perry, Super 01]
- Sensory-motor tasks [Körding & Wolpert 04]
- 2D orientation [Girshick, Landy & Simoncelli 11]
- Acceleration (vestibular) Laurens & Droulez 08

Physiological instantiation of priors?

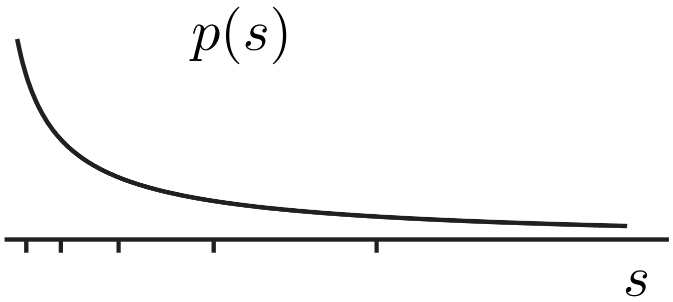
- Weighting in “readout” population
  [eg: Georgopoulos 1982; Salinas & Abbott 1994; ...]

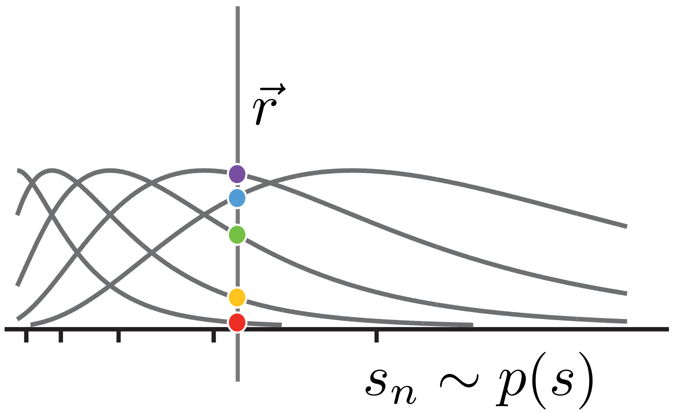
- Modulatory feedback signals
  [eg: Lee & Mumford 2003]

- Spiking activity in a separate population of cells
  [eg: Ma et. al., 2009]

- Dynamic network property (responses are samples)
  [eg: Tkacik et. al. 2010; Berkes et. al. 2011]

- Population inhomogeneities: baseline rate, gain, tuning width, and cell density
  [Stocker & Simoncelli 2006; Simoncelli, 2009; Ganguli & Simoncelli, 2010; Fischer & Pena 2011; Girshick, et. al. 2011]





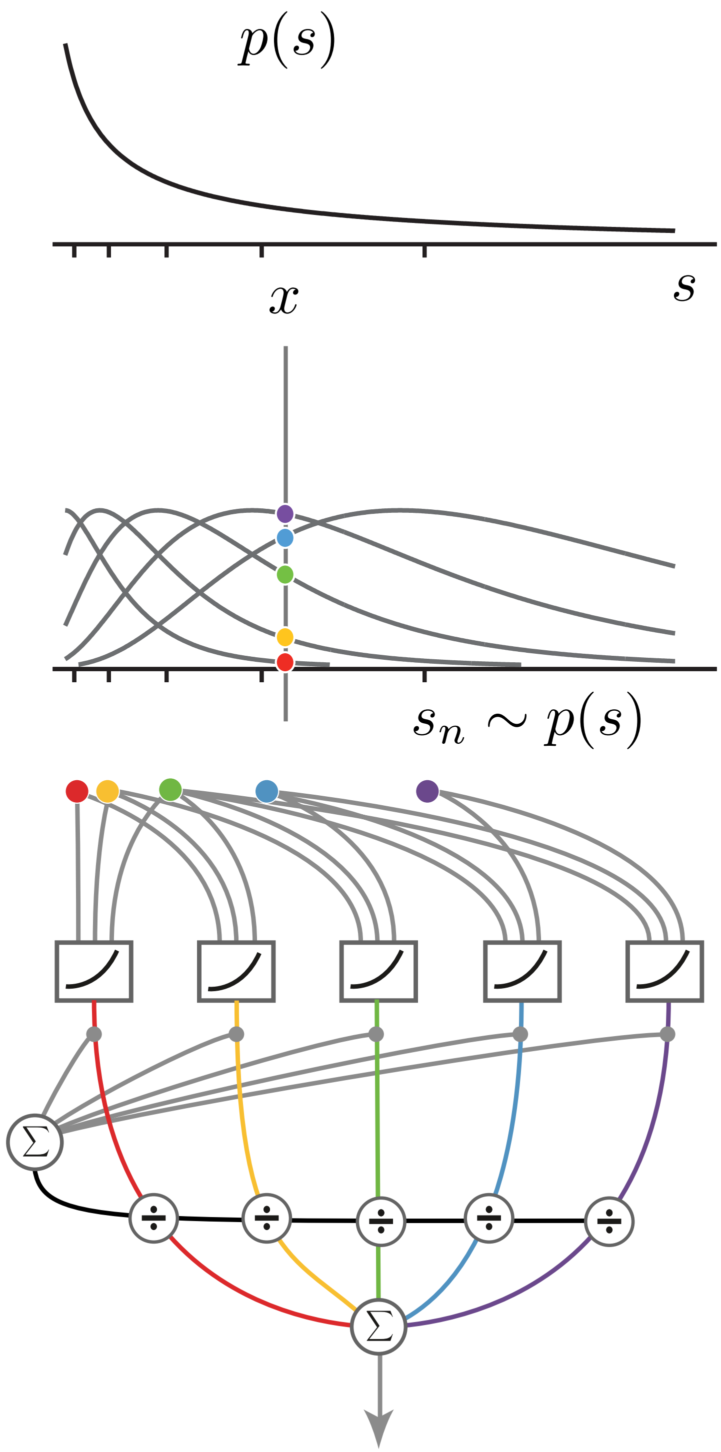
Bayes least squares estimator:

$$\hat{s}_{\mathrm{BLS}}(\vec{r}) = \frac{\int s\, p(\vec{r}|s)\, p(s)\, \mathrm{d}s}{\int p(\vec{r}|s)\, p(s)\, \mathrm{d}s}$$

Discrete (Riemann) approximation:

$$\begin{aligned} \hat{s}_{\mathrm{BLS}}(\vec{r}) &\approx \frac{\sum_{n=1}^{N} s_n\, p(\vec{r}|s_n)\, p(s_n)\, \Delta_n}{\sum_{n=1}^{N} p(\vec{r}|s_n)\, p(s_n)\, \Delta_n} \\ &= \frac{\sum_{n=1}^{N} s_n\, p(\vec{r}|s_n)}{\sum_{n=1}^{N} p(\vec{r}|s_n)} \end{aligned}$$

[Ganguli & Simoncelli, 2014]



$$\hat{s}_{\mathrm{BLS}}(\vec{r}) \approx \frac{\sum_{n=1}^{N} s_n \, p(\vec{r}|s_n)}{\sum_{n=1}^{N} p(\vec{r}|s_n)}$$

$$= \frac{\sum_{n=1}^{N} s_n \exp\left(\sum_{m=1}^{N} r_m \cdot w_{m-n}\right)}{\sum_{n=1}^{N} \exp\left(\sum_{m=1}^{N} r_m \cdot w_{m-n}\right)}$$

Similar to the "population vector"

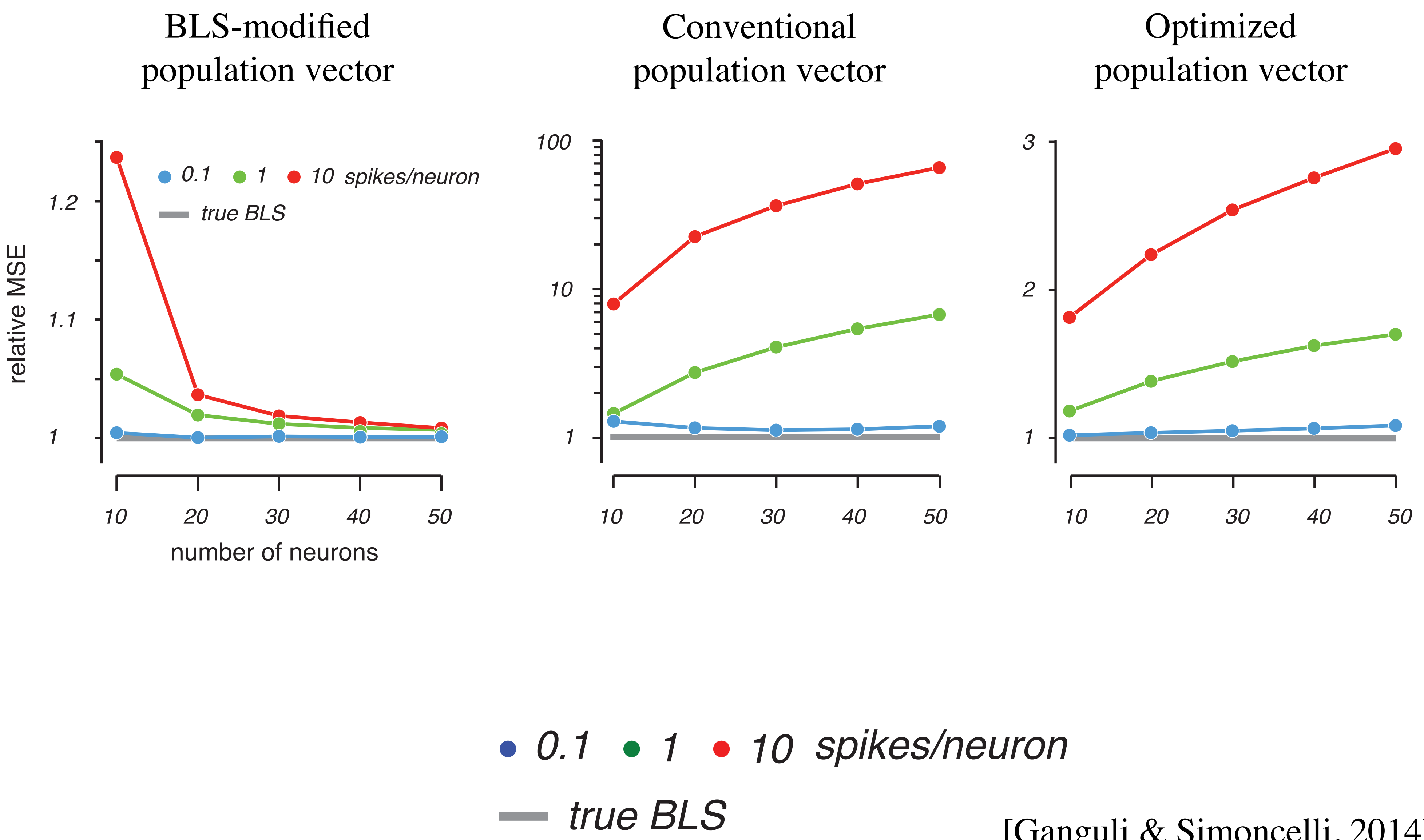
[Georgopoulos et. al. 86]

$$\hat{s}_{\mathrm{PV}}(\vec{r}(x)) \approx \frac{\sum_{n=1}^{N} s_n \, r_n}{\sum_{n=1}^{N} r_n}$$

... which has been previously proposed as an approximate Bayes estimator

[Shi & Griffiths 09; Fischer & Pena 11]

Example (power law stimulus distribution)



[Ganguli & Simoncelli, 2014]

Optimally efficient populations...

- provide a direct link between **measurable** quantities: environmental distribution, density/tuning of neurons, and perceptual discriminability

- suggest that **information maximization** is more consistent with physiological/perceptual data than **estimation/discrimination accuracy**

- provide an embedding of environmental probabilities, readily incorporated into Bayesian decoders

- Missing:
  - Multi-dimensional stimulus variables
  - Feature learning
  - Biological optimization (development, adaptation)