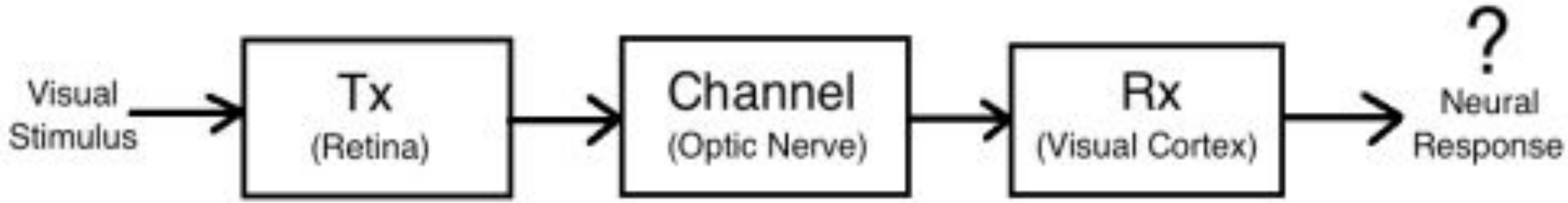
Neural Coding & Representations



Relevant Publications

Panzeri, Stefano, et al. "The structures and functions of correlations in neural population codes." *Nature Reviews Neuroscience* 23.9 (2022): 551-567.https://harveylab.hms.harvard.edu/pdf/Panzeri2022.pdf

Simoncelli & Olshausen. “NATURAL IMAGE STATISTICS AND NEURAL REPRESENTATION” Ann. Neuroscience 2001

REVIEWS

Check for updates

The structures and functions of correlations in neural population codes

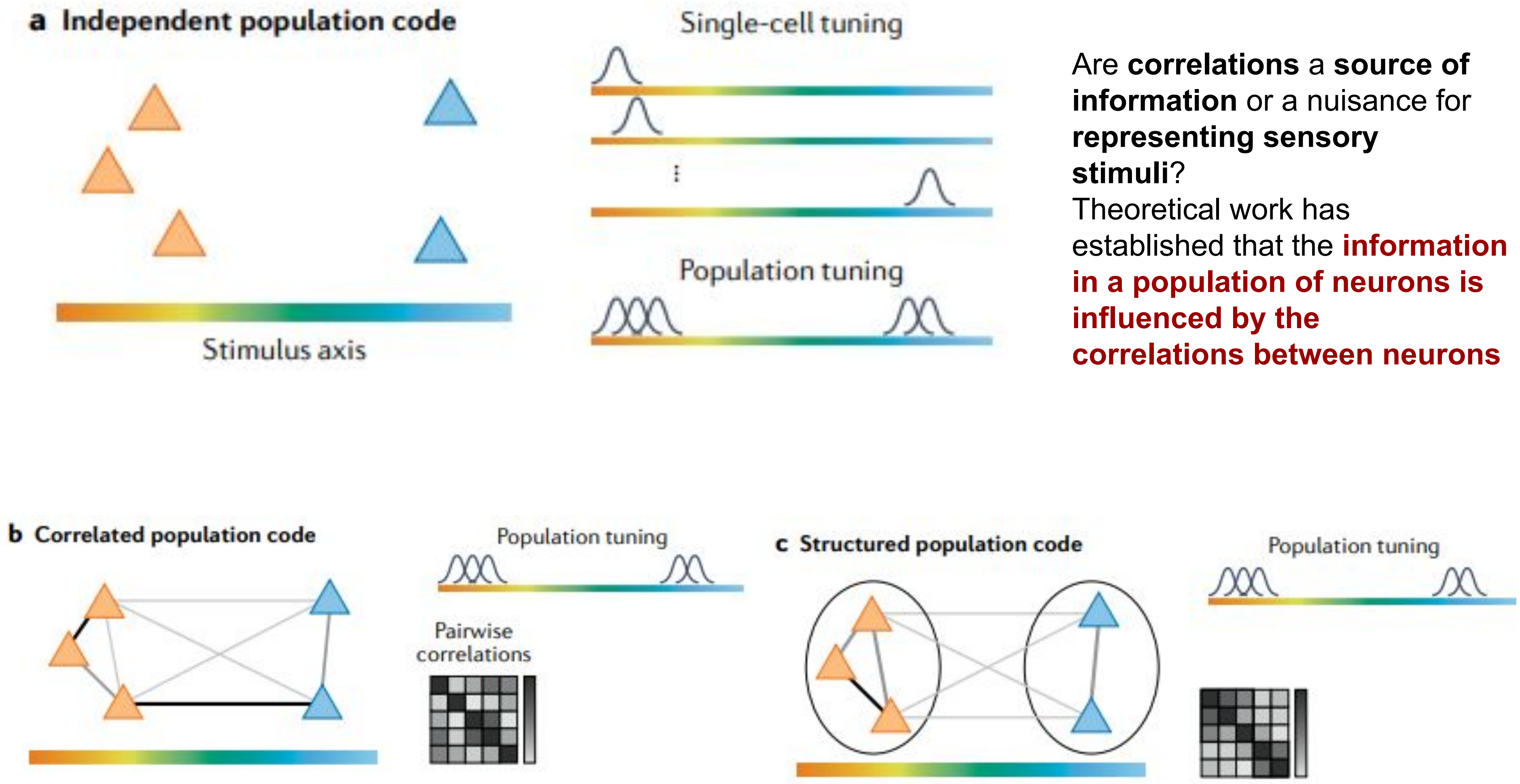
Stefano Panzeri[1,2]✉, *Monica Moroni*[2], *Houman Safaai*[3] *and Christopher D. Harvey*[3]✉

Annu. Rev. Neurosci. 2001. 24:1193–216

Natural Image Statistics and Neural Representation

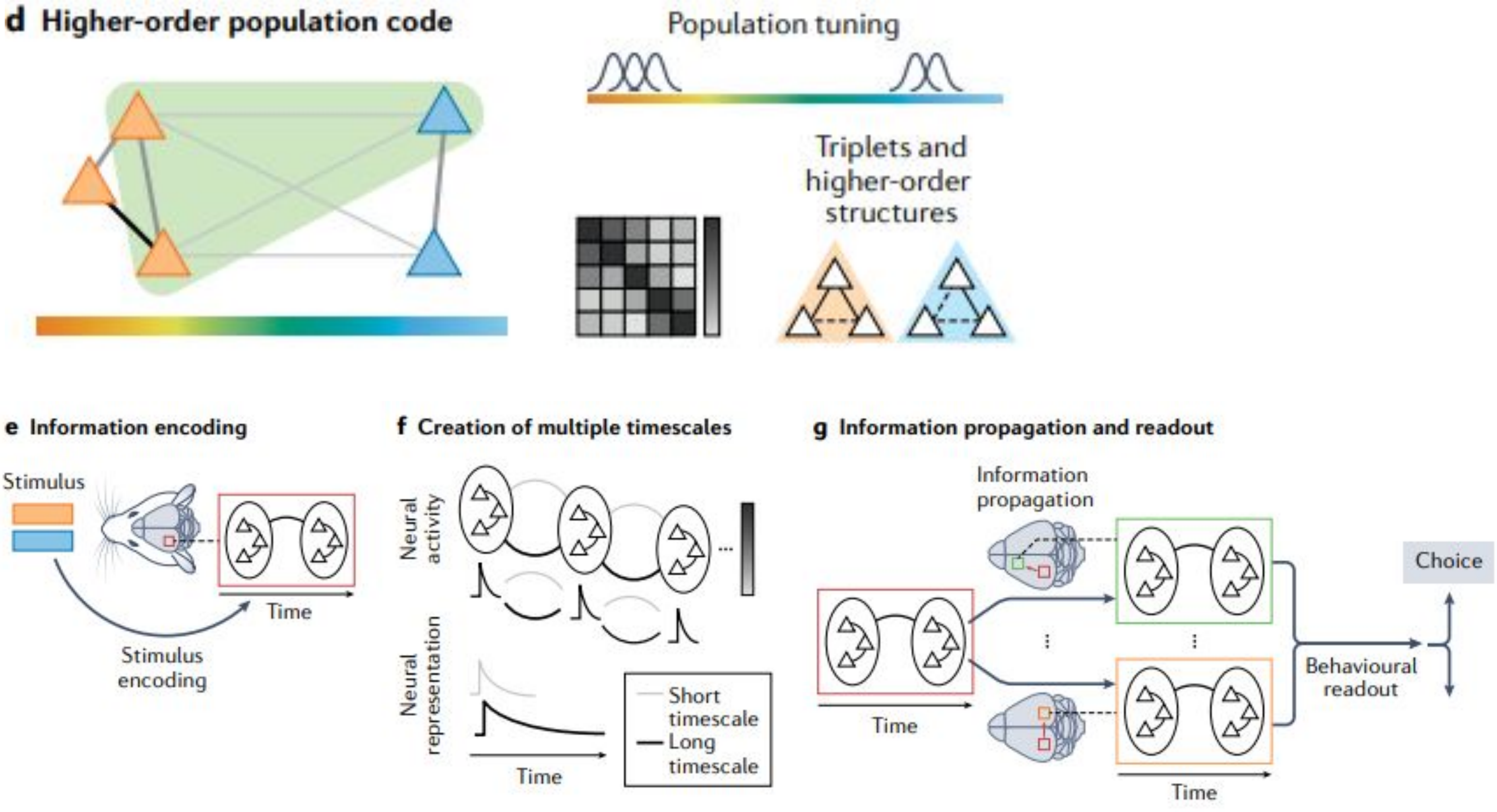
Eero P Simoncelli
Howard Hughes Medical Institute, Center for Neural Science, and Courant Institute of Mathematical Sciences, New York University, New York, NY 10003; e-mail: eero.simoncelli@nyu.edu

Bruno A Olshausen
Center for Neuroscience, and Department of Psychology, University of California, Davis, Davis, California 95616; e-mail: baolshausen@ucdavis.edu



Are **correlations** a **source of information** or a nuisance for **representing sensory stimuli**?

Theoretical work has established that the **information in a population of neurons is influenced by the correlations between neurons**



d Higher-order population code
Population tuning
Triplets and higher-order structures
e Information encoding
Stimulus
Time
Stimulus encoding
f Creation of multiple timescales
Neural activity
Neural representation
Time
Short timescale
Long timescale
g Information propagation and readout
Information propagation
Time
Behavioural readout
Choice

Signal and Noise Correlations

- **Signal correlations** indicate the **similarity of stimulus tuning** of different neurons, with high signal correlations for neurons tuned to the same stimuli.

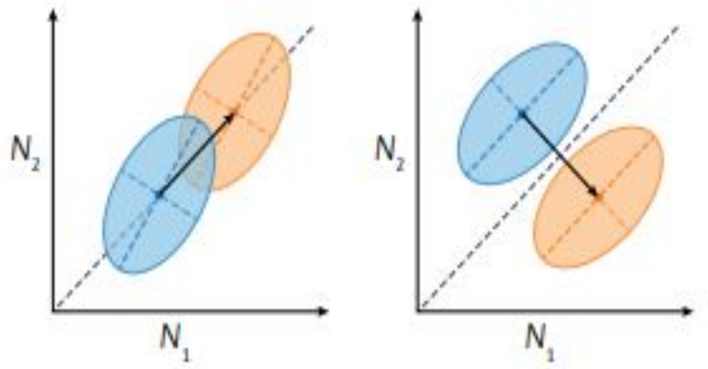
  Signal correlations can be defined even for independent population codes and can be measured in both simultaneously recorded and pseudo-population responses.

- **Noise correlations** measure **activity correlations *beyond* the stimulus tuning** shared by the neurons and are often quantified as the **correlation in *individual* trial responses between neurons for a *given stimulus***.

Noise Correlations and Information Encoding

Response distributions (ellipses) of **populations of neurons** (N1, N2) in the space of neural population activity (using n=2 or 3 neurons) in response to **two stimuli** (orange and blue).

a Stimulus-independent noise correlations

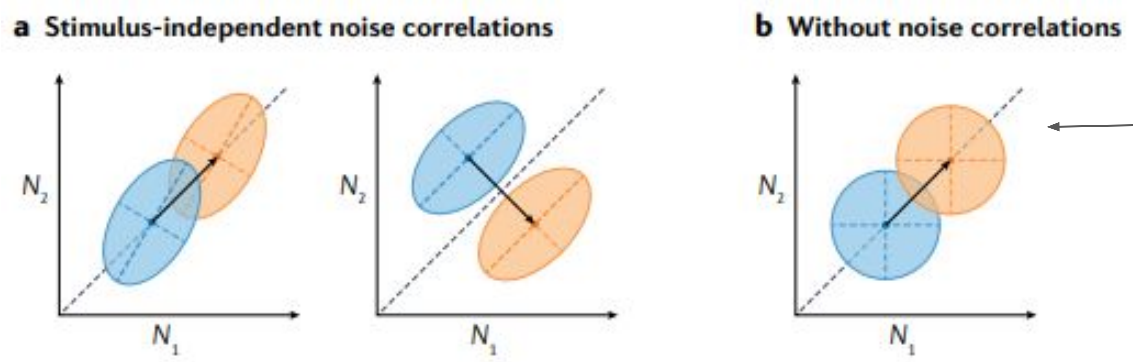


Stimulus-independent noise correlations can decrease (lef
or increase (right) the amount of encoded information about stimulus

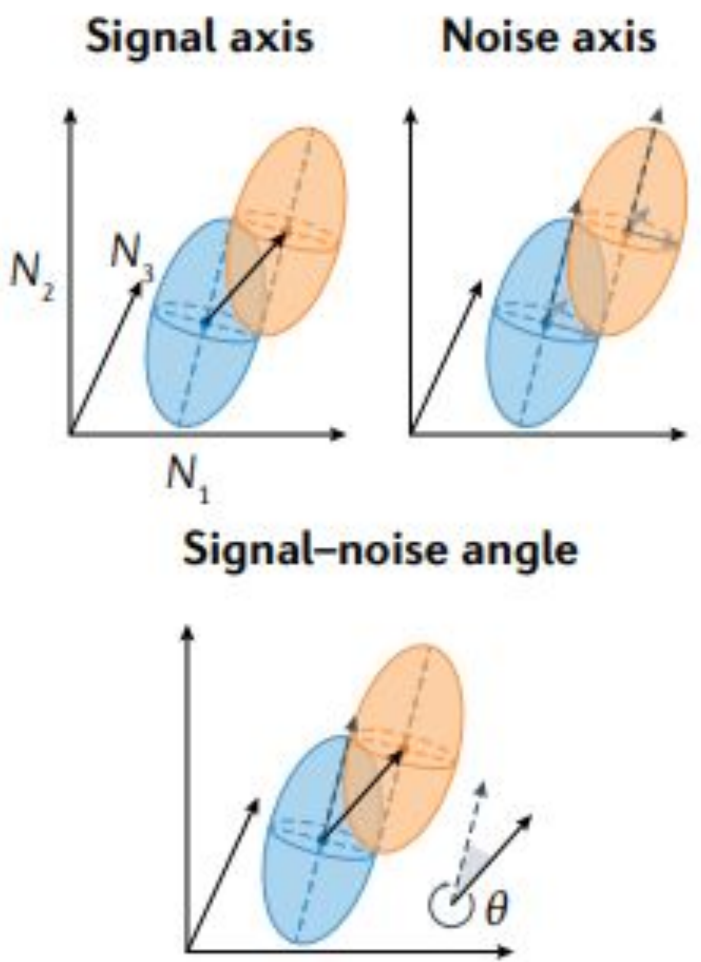
Overlapping distributions being harder to discriminate

Noise Correlations and Information Encoding

Response distributions (ellipses) of **populations of neurons** (N1, N2) in the space of neural population activity (using n=2 or 3 neurons) in response to **two stimuli** (orange and blue).



The properties of the individual neurons are the same but **noise correlations are absent**



Interplay between **signal and noise** can be described using the **signal axis and noise axis**

Signal axis connects the **average responses to different stimuli** (solid arrow)

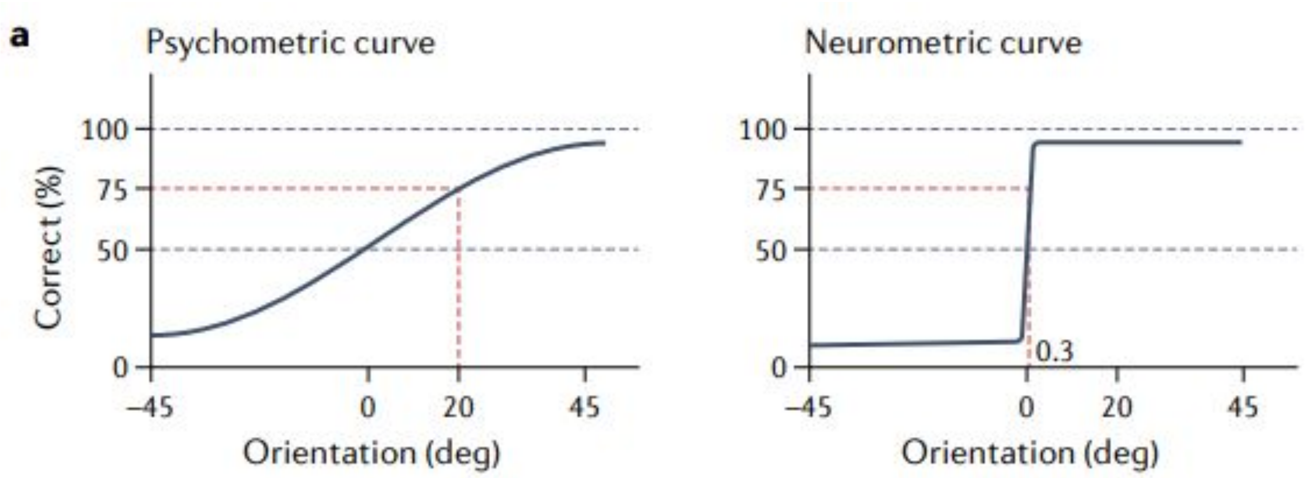
Noise axis: the **direction of maximum variability of a response to a fixed stimulus** (dashed arrow)

Signal–noise angle, which is the high-dimensional angle between the **signal axis and noise axis.**

Information transmission and Behaviour

- Correlated population codes have typically been studied from the perspective of information encoding but, ultimately, the importance of these codes depends on ***how they are transmitted to downstream brain regions*** **and** ***used to guide behaviour***.
- This issue is critical if the **reading out of population codes** is suboptimal. Thus, a major question concerns whether correlations in neural populations help or hinder the propagation of signals to downstream networks.

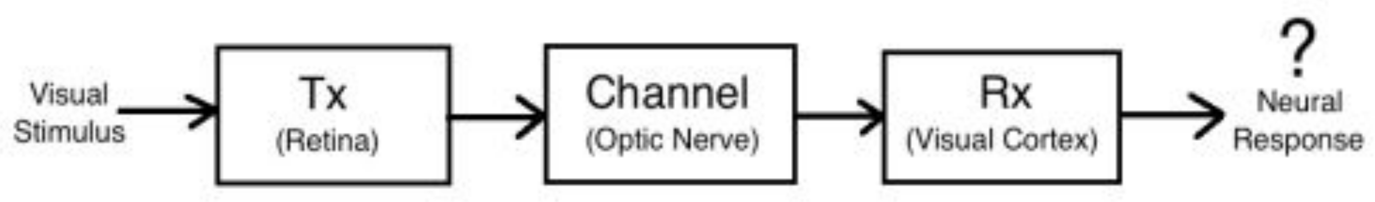
- In principle, correlations could aid the transmission of information even if they limit the encoding capacity.



Three levels for solving information processing tasks (Marr, 1982)

Computational theory	Representation and algorithm	Hardware implementation
What is the goal of the computation?	How can it be implemented?	How can the representation and algorithm be realized physically?
Why is it appropriate?	What is the representation for the input and output?	
What is its logic?	What is the algorithm for the transformation?	And how can they be represented in the brain?
How can it be carried out?		

This hypothesis suggests that a group of neurons should encode information as compactly as possible in order to **maximize efficient utilization of resources**, meaning that the information carried by one neuron should ***not*** be redundant to the information which is carried by the other neurons.



The Efficient Coding Hypothesis assumes that from the response of the neurons in the visual cortex, one can reconstruct the original stimulus with some accuracy.

Efficient Coding Hypothesis

- More than 40 years ago, motivated by developments in information theory, Attneave (1954) suggested that the goal of visual perception is to produce an ***efficient representation of the incoming signal*.**
- In a neurobiological context, Barlow (1961) hypothesized that the **role of early sensory neurons** is to ***remove statistical redundancy*** in the sensory input.
- Variants of this "**efficient coding**" hypothesis have been formulated by numerous other authors (e.g. Laughlin 1981, Atick 1992, van Hateren 1992, Field 1994, Riecke et al 1995).

Specification of a probability distribution over the space of input signals.

- Difficult problem in general!
- Easier: empirical statistics computed from a large set of example images that are representative of the relevant environment.
- Specify **time scale** over which the environment should shape the system.
- State which neurons are meant to satisfy the **efficiency criterion**, and how their responses are to be interpreted.

Efficiency of Neural Code

- Examine the statistical properties of environmental signals
- Show that a transformation derived according to some statistical optimization criterion provides a good description of the response properties of a set of sensory neurons

- Depends both on the **transformation that maps the input to the neural responses** and on **the statistics of the input.**
- Optimal efficiency of the neural responses for one input ensemble does ***not*** imply optimality over other input ensembles.

The efficient coding principle $\neq$ *optimal compression* (i.e. rate-distortion theory) or optimal estimation.

- $\neq$ accuracy with which the signals are represented
- **No requirement that the transformation from input to neural responses** is invertible.
- No assumption about the representation or cost of misrepresenting the input (but costs are relevant for real organisms!)
- **Uncertainty or variability** in the neural responses to identical stimuli
- Presence of **noise**!

Rate-Distortion Theory

- A major branch of information theory that establishes the mathematical foundations of **lossy data compression** for any communication channel (Shannon, 1959).
- Formalizes the link between **compression and communication** by determining the minimum amount of information that a source should transmit (the rate) for a target to approximately receive the input signal *without exceeding an expected amount of noise* (the distortion) (Shannon, 1959).

Lossy data compression is **reducing the amount of information transmitted (rate), accepting some loss of data fidelity (distortion).** Using a rate-distortion model, we sought to explain how the macroscale connectome supports efficient coding from minimal assumptions.

Efficiency of Neural Coding

Take noise into account, by **maximizing the information that the responses provide about the stimulus** (technically, the mutual information between stimulus and response).

Generally difficult to measure but there are approximation methods Bialek et al (1991) and Rieke et al (1995)

Efficient Coding in Single Neurons

- Determine whether the information conveyed by this neuron is maximal
- Impose a constraint on the response values (maximal value Rmax)

(if they can take on any real value, then the amount of information that can be encoded is unbounded)

- The distribution of responses that conveys maximal information is **uniform** over the interval [0, Rmax]. That is, an **efficient neuron** should make equal use of all of its available response levels.
- The **optimal distribution depends critically on the neural response constraint:** e.g.,

If the **variance is fixed**, the **information maximizing response distribution is a Gaussian**.

If the **mean of the response is fixed**, the information maximizing response distribution is an exponential*.

* Jaynes, E. T. (1957). "Information theory and statistical mechanics." Physical Review, 106(4), 620. (proof using Langrage multipliers)

Key Result from Information Theory: Maximum Entropy Principle

Among all probability distributions with a fixed variance σ^2, the **Gaussian (Normal) distribution** maximizes entropy.

$$X \sim \mathcal{N}(\mu, \sigma^2) \quad \Rightarrow \quad H(X) = \frac{1}{2}\log(2\pi e\sigma^2)$$

The Gaussian is the **least structured distribution given a mean and variance,** meaning it spreads out the information most efficiently.

Any other distribution with the same variance has lower entropy and thus conveys less information.

Experimental Data — Firing Rates

Baddeley et al (1998) showed that the instantaneous firing rates of spiking neurons in **primary and inferior temporal visual cortices of cats and monkeys** are **exponentially distributed (when visually stimulated with natural scenes)**, consistent with optimal coding with a constraint on the mean firing rate.

Jaynes, E. T. (1957). "Information theory and statistical mechanics." Physical Review, 106(4), 620.

Foundational paper on maximum entropy principles.

Shows that if only the mean is constrained, the probability distribution that maximizes entropy is exponential.

Does the visual system take advantage of the correlational structure of natural images ?

Srinivasan et al (1982) measured the autocorrelation function of natural scenes and then computed the **amount of subtractive inhibition** that would be required **from neighboring photoreceptors** to effectively **cancel out these correlations**.

They then compared the **predicted inhibitory surround fields** to those actually **measured from first-order interneurons in the compound eye of the fly**.

The correspondence was surprisingly good and provided the **first quantitative evidence for decorrelation in early spatial visual processing**

Does the visual system take advantage of the correlational structure of natural images ?

Atick & Redlich (1991, 1992) on the problem of whitening the power spectrum of natural images (equivalent to decorrelation) in the presence of white photoreceptor noise:

showed that both single-cell physiology and the psychophysically measured contrast sensitivity functions are consistent with the product of a whitening filter and an optimal lowpass filter for noise removal (known as the Wiener filter). Similar predictions and physiological comparisons were made by van Hateren (1992) for the fly visual system. The inclusion of the Wiener filter allows the behavior of the system to change with mean luminance level. Specifically, at lower luminance levels (and thus lower signal-to-noise ratios), the filter becomes more low-pass (intuitively, averaging over larger spatial regions in order to recover the weaker signal).

Another model for retinal horizontal cells, proposed by Balboa & Grzywacz (2000),, assumes a divisive form of retinal surround inhibition and shows that the changes in effective receptive field size are optimal for representation of intensity edges in the presence of photon-absorption noise.

Does the visual system take advantage of the correlational structure of natural images ?

Atick & Redlich (1991, 1992) on the problem of **whitening the power spectrum of natural images** (equivalent to **decorrelation**) in the presence of white photoreceptor noise:

showed that both **single-cell physiology** and the **psychophysically** measured contrast sensitivity functions are consistent with the **product of a whitening filter** and an o**ptimal low pass filter for noise removal** (known as the Wiener filter).

Similar predictions and physiological comparisons were made by van Hateren (1992) for the **fly visual system**. The inclusion of the **Wiener filter allows the behavior of the system to change with mean luminance level.**

Specifically, **at lower luminance levels** (and thus lower signal-to-noise ratios), the filter becomes more low-pass (intuitively, averaging over larger spatial regions in order to recover the weaker signal).

Another model for retinal horizontal cells, proposed by Balboa & Grzywacz (2000),, assumes a divisive form of retinal surround inhibition and shows that the changes in effective receptive field size are optimal for representation of intensity edges in the presence of photon-absorption noise.

How neurons might represent probabilities?

Several proposals for how neurons might represent probabilities have been presented (Pouget et al., 2013), the simplest of which **directly relate neural activity to the probability of a feature being present in the neuron's classical RF** (Barlow, 1969; Anastasio et al., 2000; Rao, 2004).

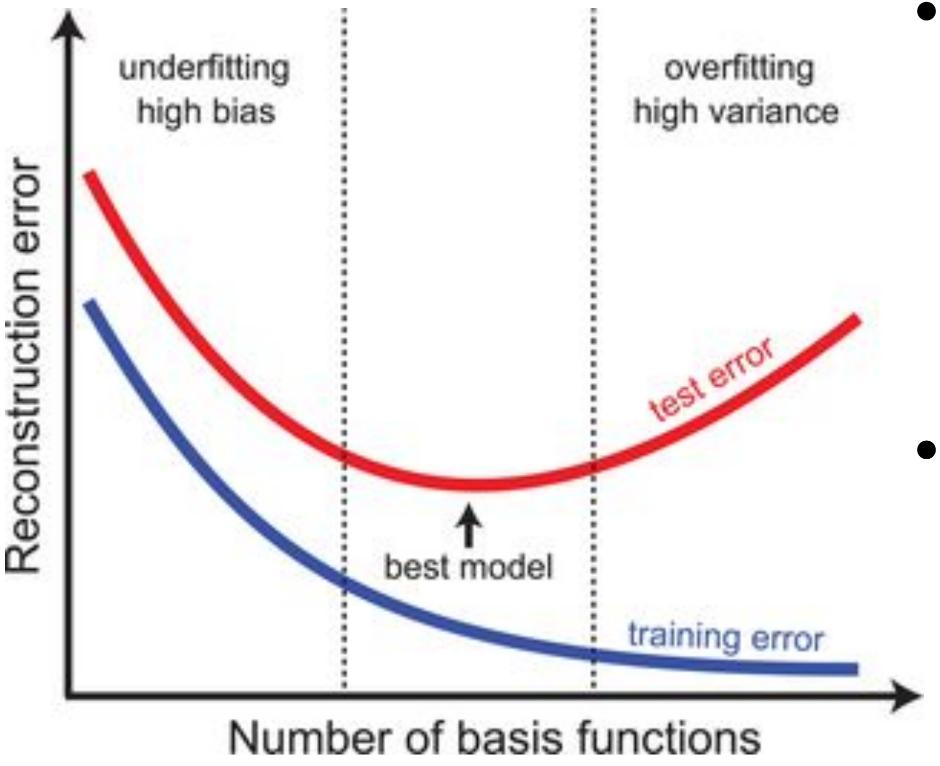
Many authors assume that **at the earliest stages of processing (e.g. retina and V1),** it is desirable for the system to provide a **generic image representation that preserves *as much information as possible about the incoming signal*.**

The success of efficient coding principles in accounting for response properties of neurons in the retina, LGN, and V1 may be seen as verification of this assumption.

However, a richer theoretical framework is required. A commonly proposed example of such a **framework is Bayesian decision estimation theory**, which includes both a ***prior* statistical model for the environment** and a ***loss or reward function*** that specifies the cost of different errors, or the desirability of different behaviors.

https://www.youtube.com/watch?v=vJG698U2Mvo

The Bias-Variance Dilemma



- **High Bias (Low Complexity Model)**:

  If a coding scheme is too simplistic (e.g., using a highly compressed representation with strong assumptions), it may **discard valuable details** and **fail** to capture **true variations in the signal**.

- **High Variance (High Complexity Model)**:

  If a coding scheme tries to capture every small fluctuation, it becomes highly **sensitive to noise**, making it inefficient in generalizing across different contexts.

The Bias-Variance Dilemma

- **High Bias (Low Complexity Model)**: If a coding scheme is too simplistic (e.g., using a highly compressed representation with strong assumptions), it may **discard valuable details** and **fail** to capture **true variations in the signal**.
- **High Variance (High Complexity Model)**: If a coding scheme tries to capture every small fluctuation, it becomes highly **sensitive to noise**, making it inefficient in generalizing across different contexts.

In **neural coding**, this trade-off is seen in:

- **Sparse vs. Redundant Representations**: Sparse coding in the visual and auditory systems reduces redundancy (higher bias), while more distributed coding (higher variance) captures fine details at the cost of noise sensitivity.
- **Predictive Coding Models**: The brain is thought to **minimize prediction errors** by **adjusting bias and variance dynamically**. It keeps high bias when stimuli are predictable and lowers bias (increasing variance) when encountering unexpected inputs.