How do concepts relate to each other?

Readings for today

- Sims, C. R. (2018). Efficient coding explains the universal law of generalization in human perception. Science, 360(6389), 652-656

Topics

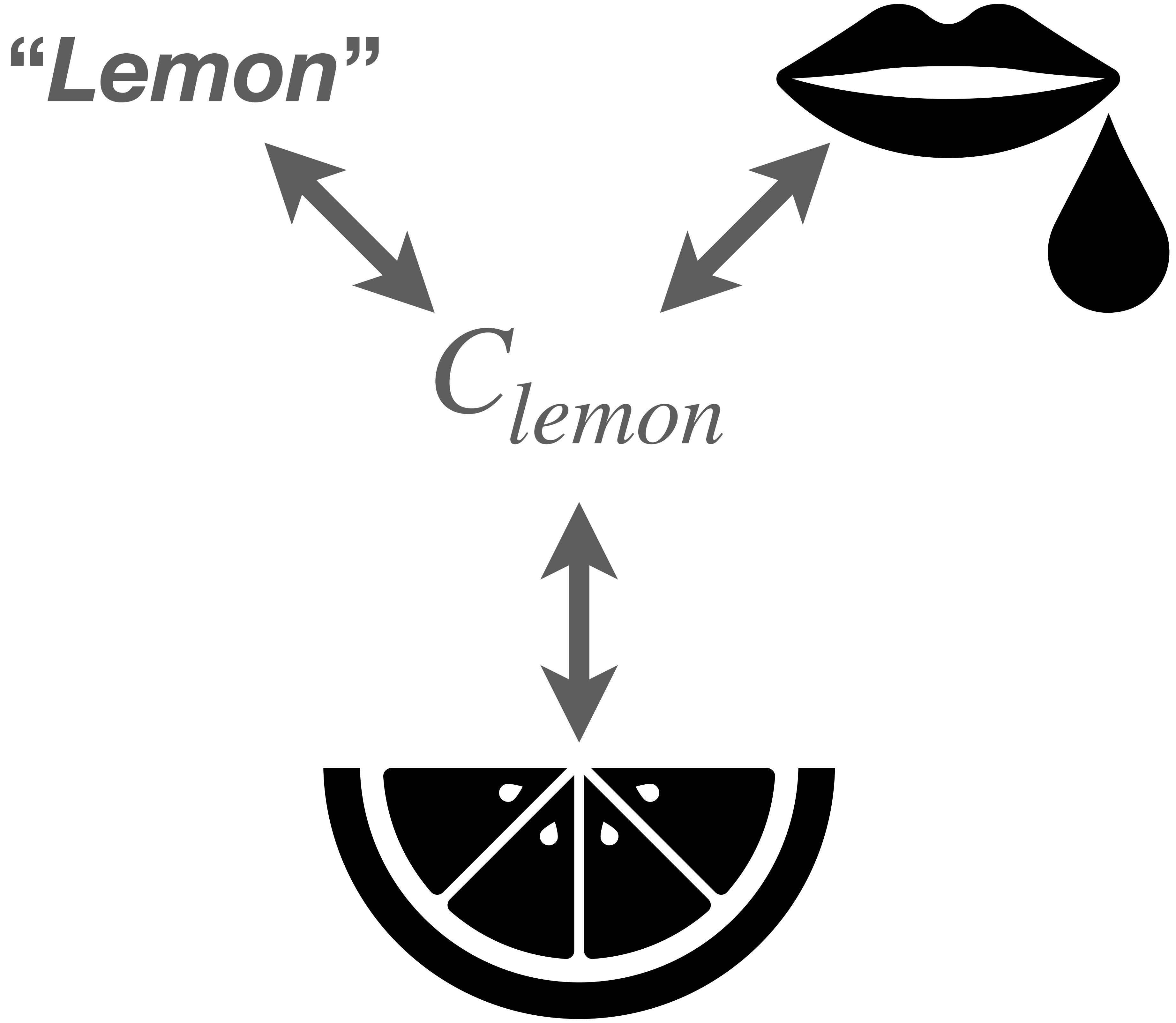
- How do concepts relate to each other?
- The efficient coding hypothesis.

Concepts & their relations

What is a concept?

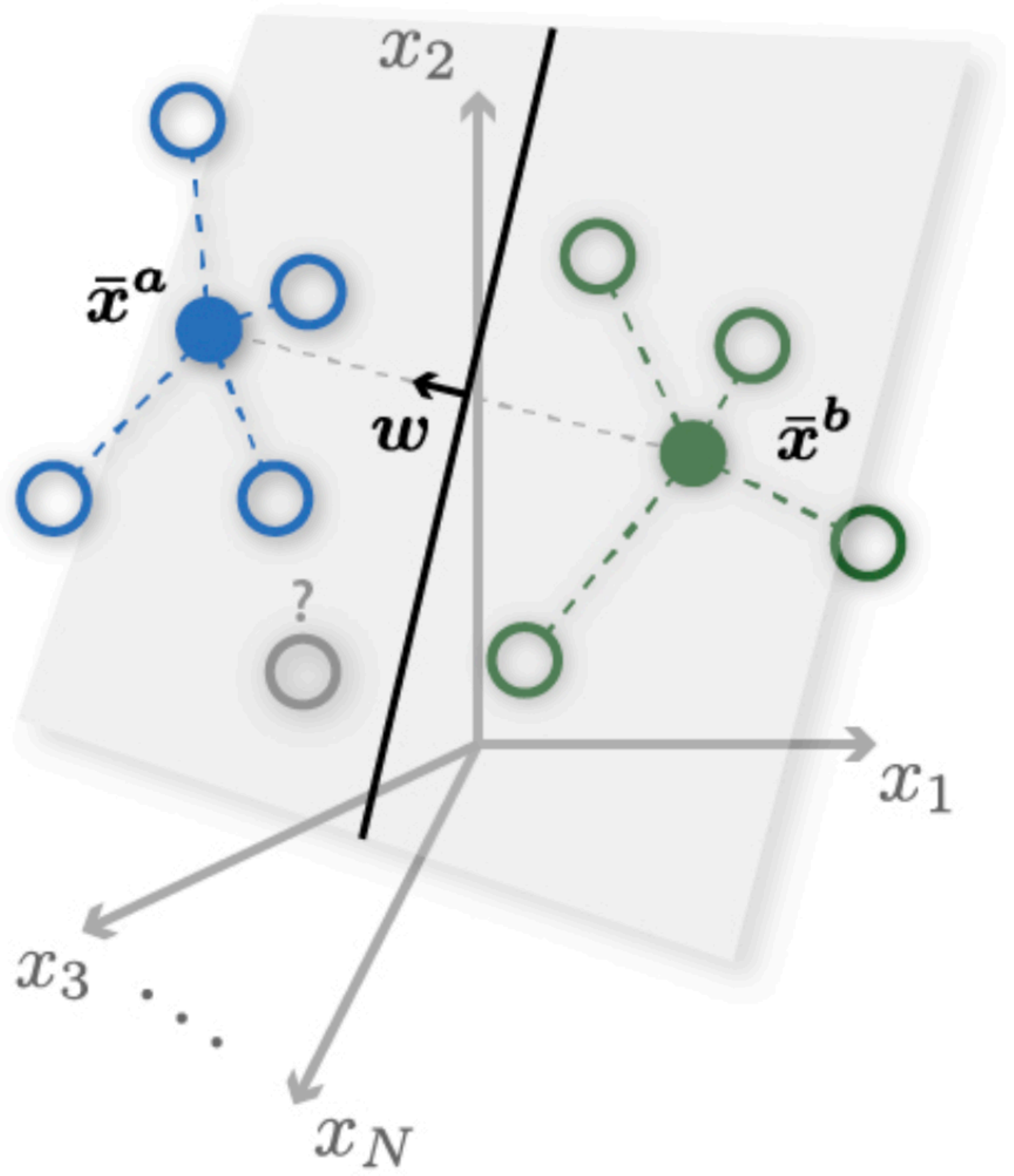
Definition: "[A] type of internal structure: one whose semantic content, when instantiated, exercises control over system output." - F. Dretske

- **Backward facing**: to the source of informational content.
- **Forward facing**: to the effects and consequences.



(Dretske, 1981)

What is a concept geometry?



<u>Concept geometry</u>:

Concepts are defined by tight geometric regions in the space of high-dimensional underlying (neural) representations.

- New concepts are learned in context of prior concept relations.
- Similarity of concepts is defined by shorter distances in representational space.

Universal law of generalization

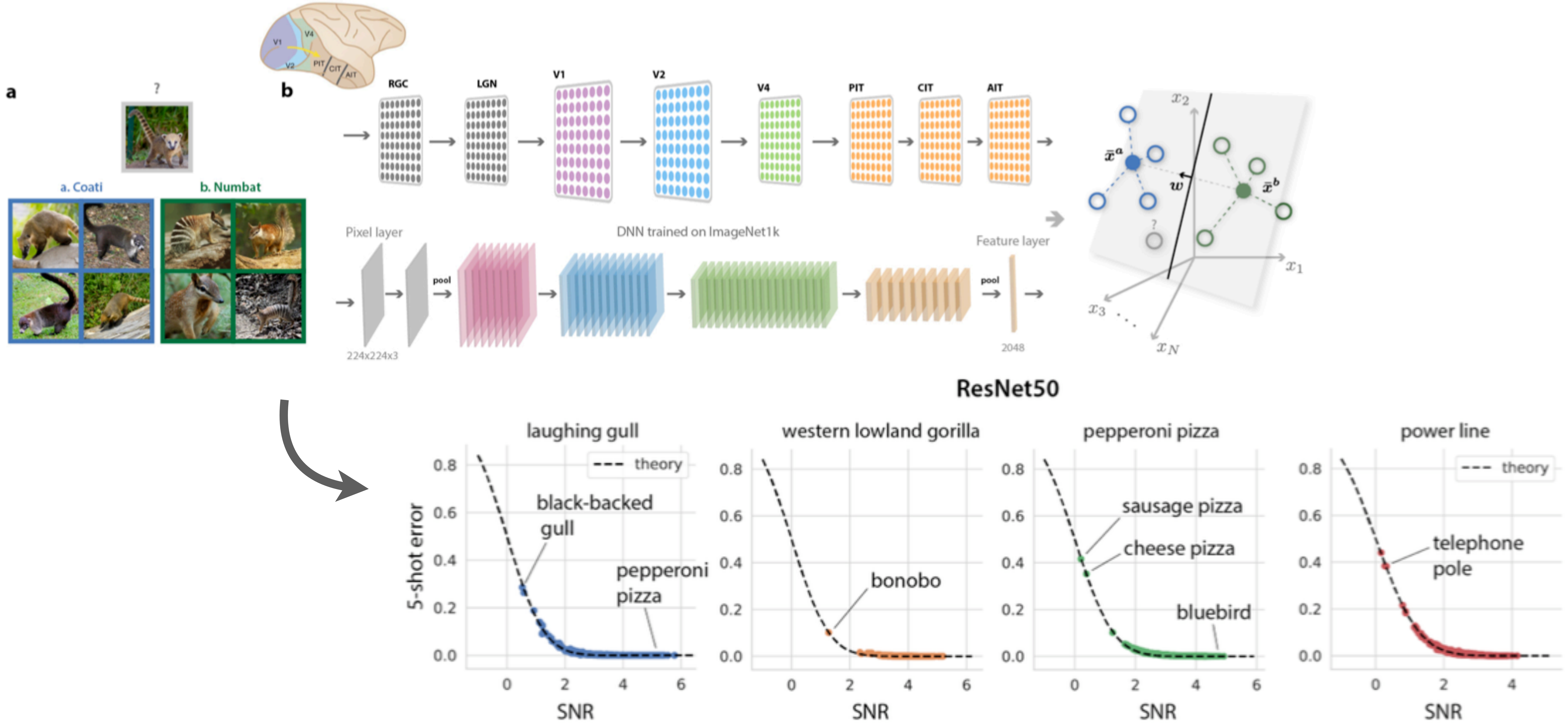
"A psychological space is established for any set of stimuli by determining metric distances between the stimuli such that the probability that a response learned to any stimulus will generalize to any other is an invariant [monotonic function] of the distance between them" - R. Shepard

$$\mathcal{G}_{x\hat{x}} \triangleq \left(\frac{p_{x\hat{x}} \cdot p_{\hat{x}x}}{p_{\hat{x}\hat{x}} \cdot p_{xx}}\right)^{\frac{1}{2}} = \exp[\underbrace{s}_{\text{channel capacity}}\underbrace{\mathscr{L}}_{\text{loss function}}(x, \hat{x})]$$

$p_{x\hat{x}} \longrightarrow$ Probability that a response associated with $\hat{x}$ is made to x

(Shepard, 1987; Sims, 2018)

Exponential relationship

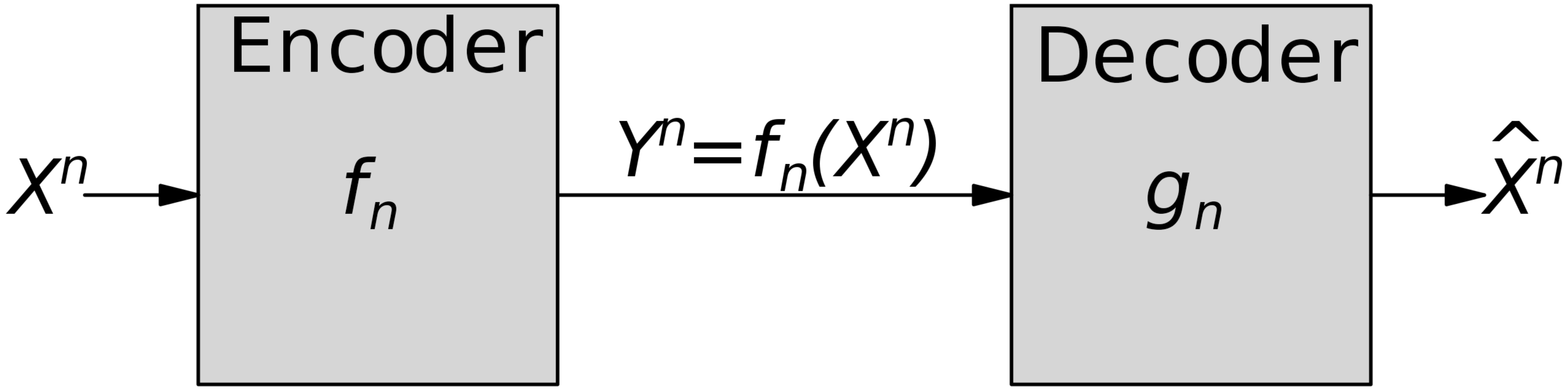


(Shepard, 1987)

The efficient coding hypothesis

Rate distortion theory

Problem:



Goal: $\min d(x, \hat{x}) = (x - \hat{x})^2$

distortion error

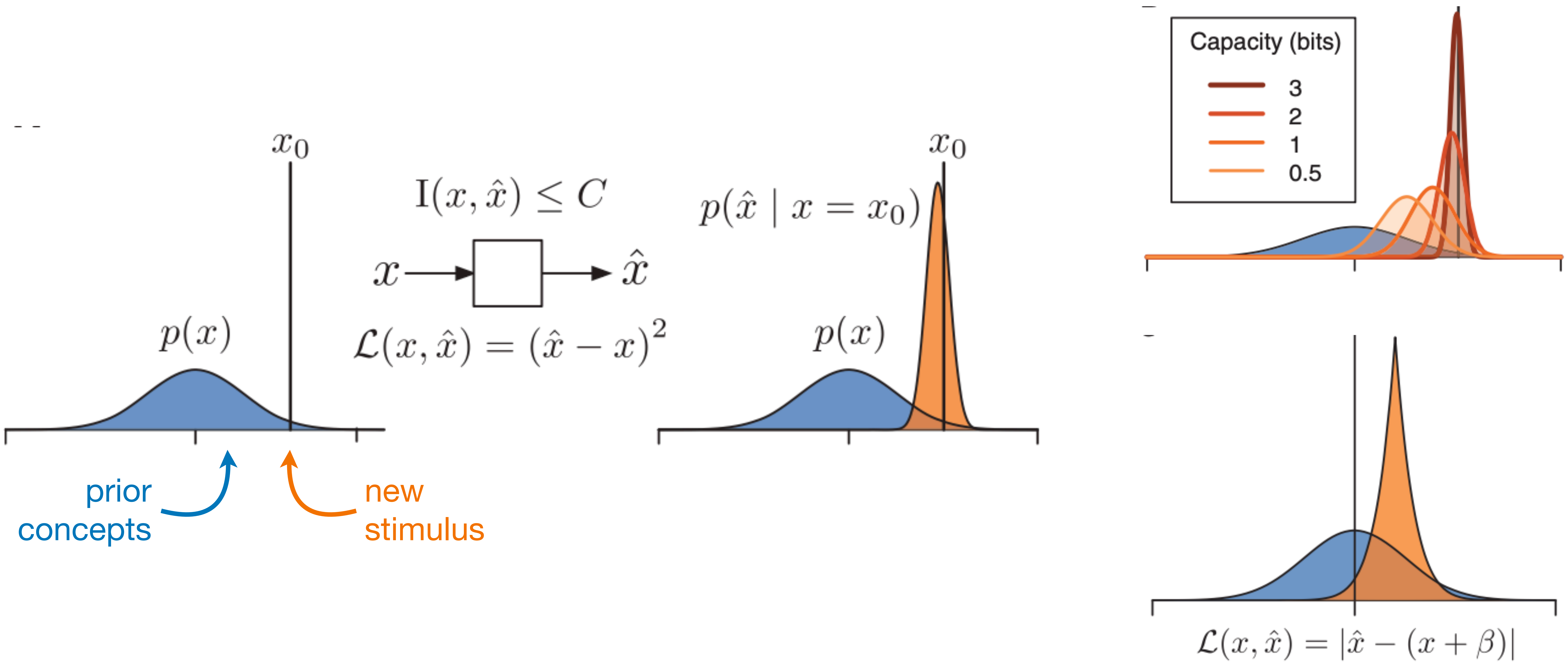
Form: $I_Q(X^n; \hat{X}^n) \leq R(D)$

mutual information

rate given distortion

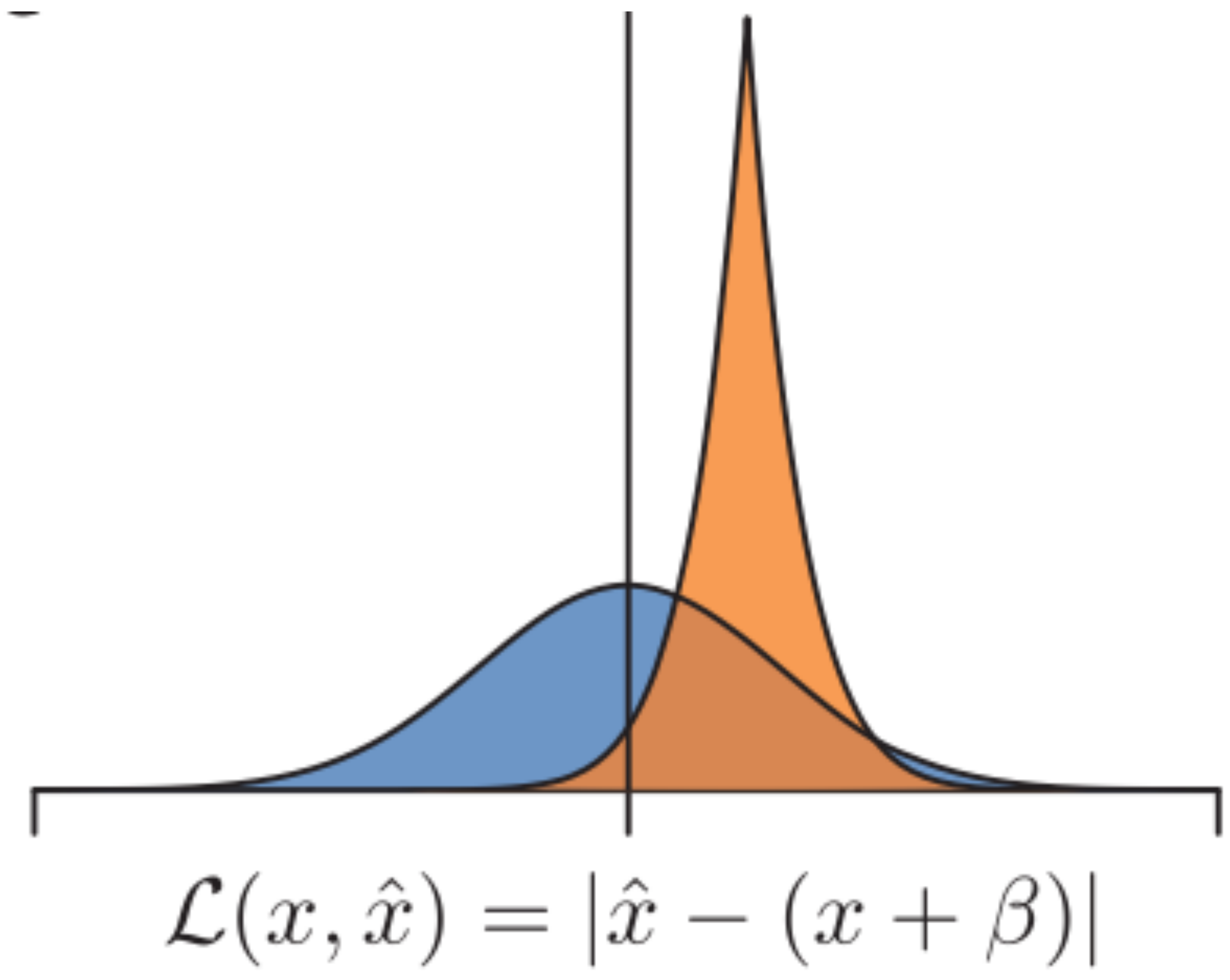
How much signal preservation can be achieved using lossy (i.e., signal not perfectly encoded) compression methods.

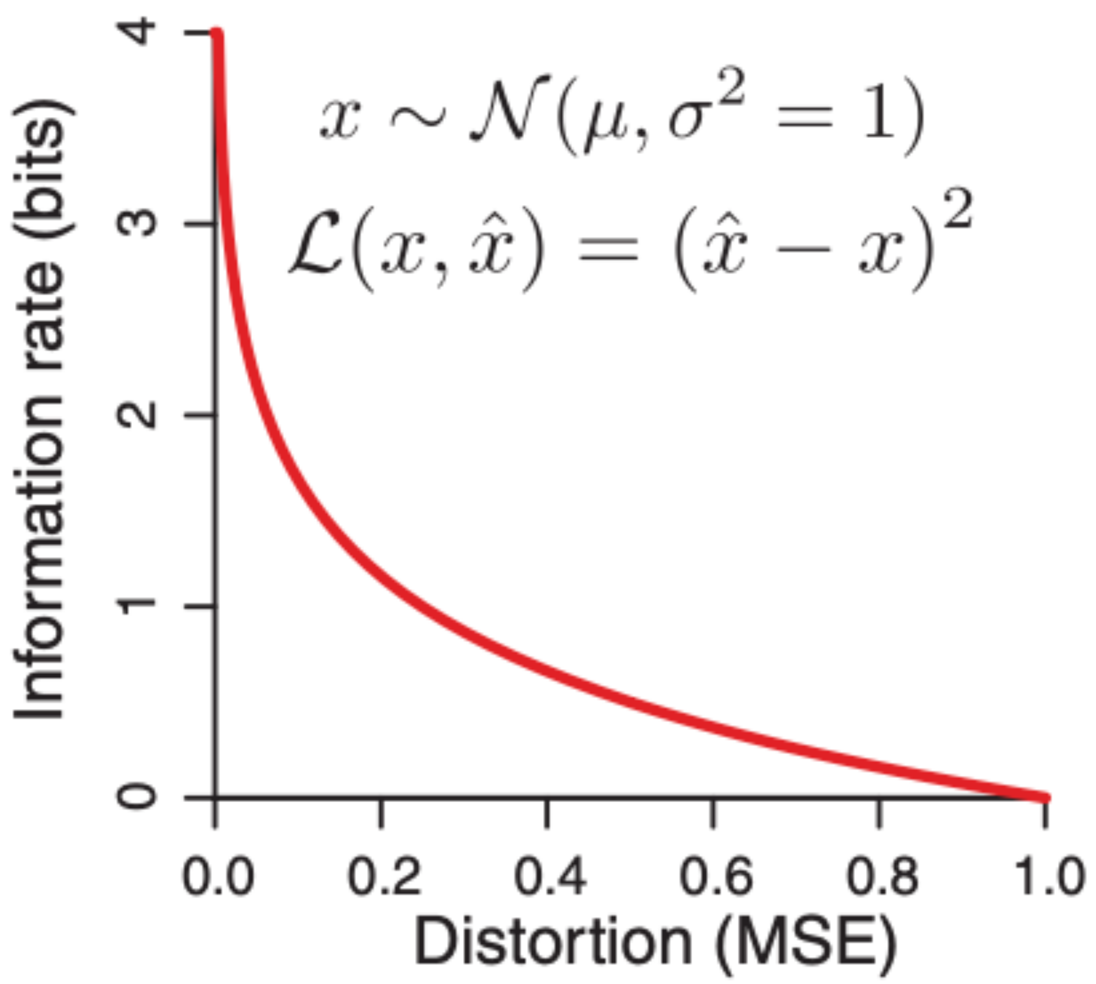
Perception as efficient coding

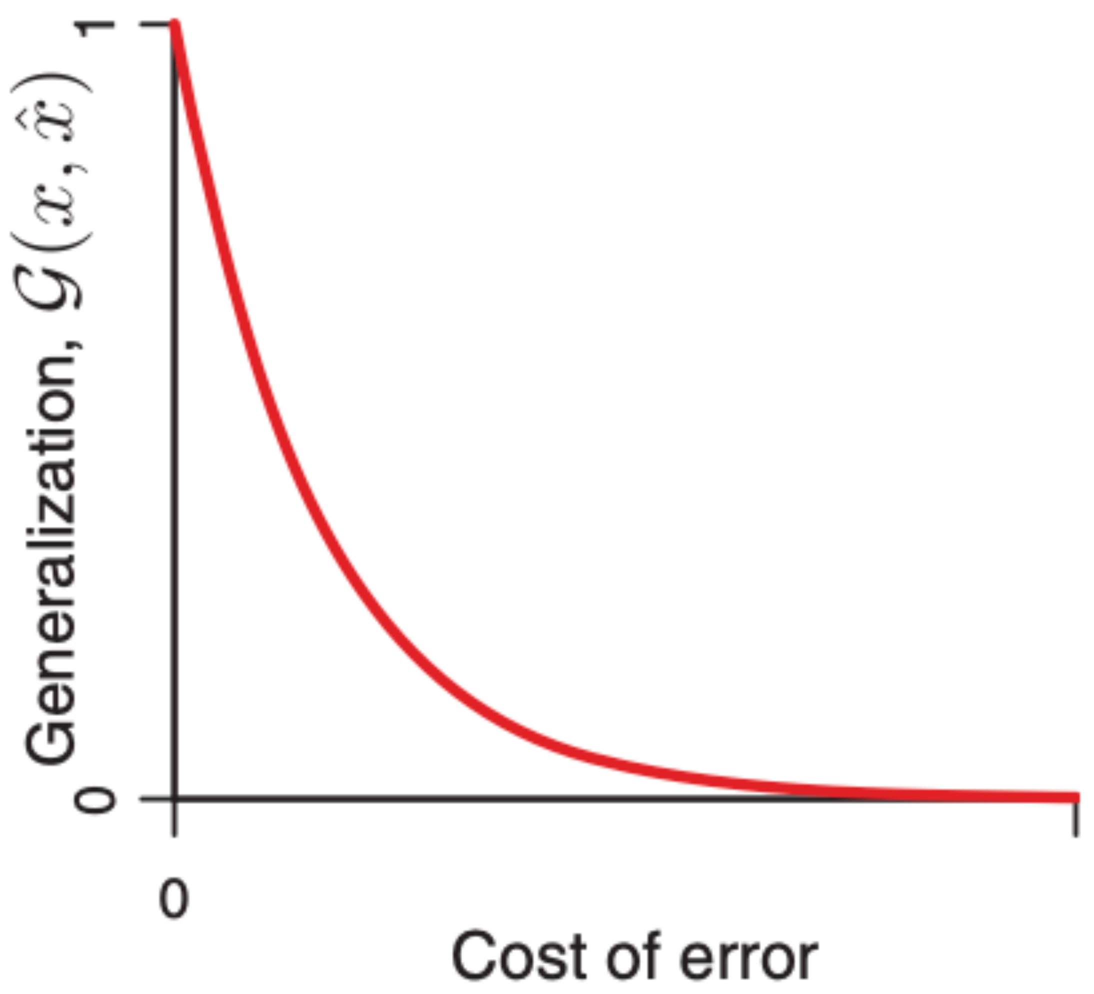


(Sims, 2018)

Perception as efficient coding





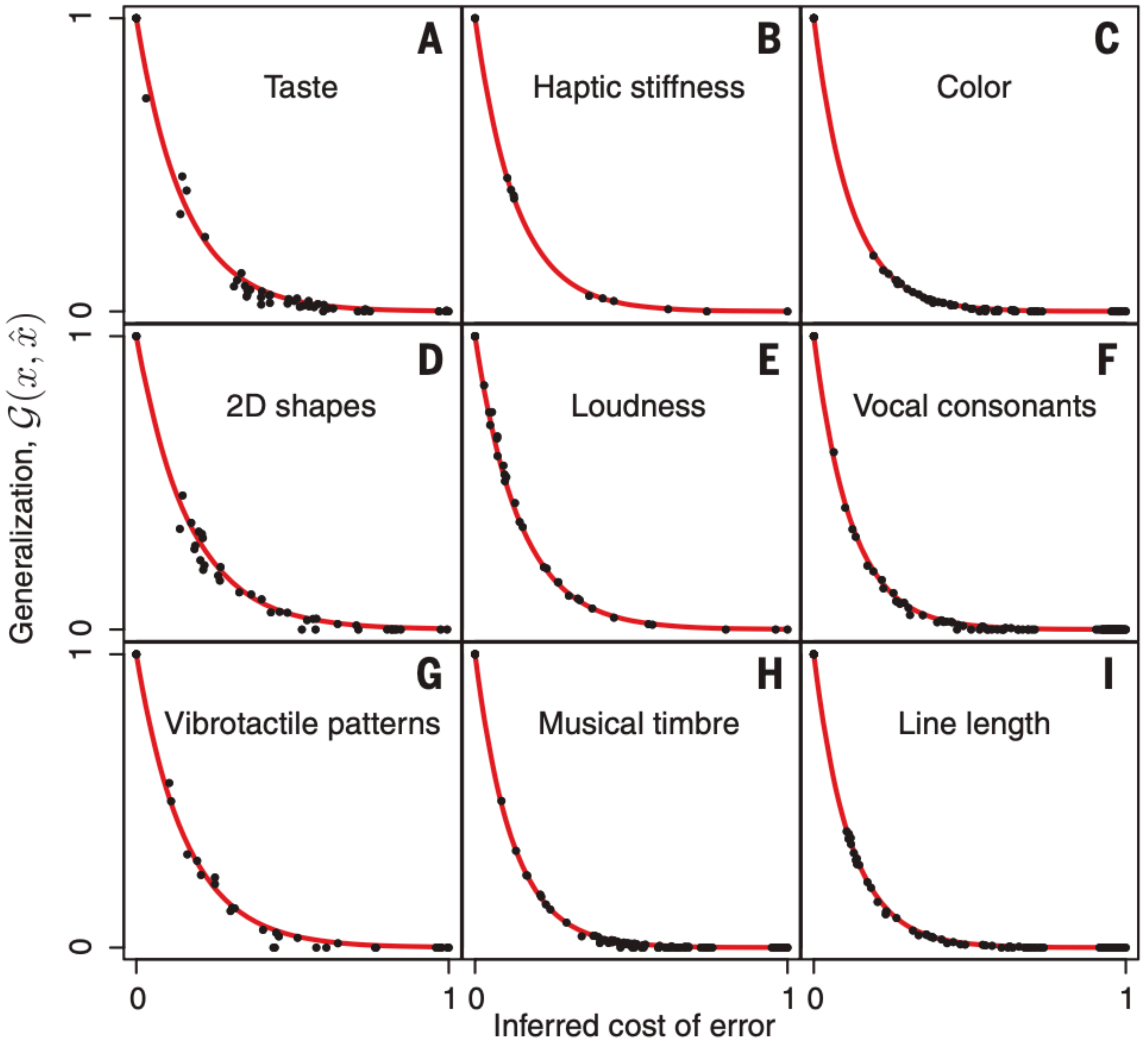


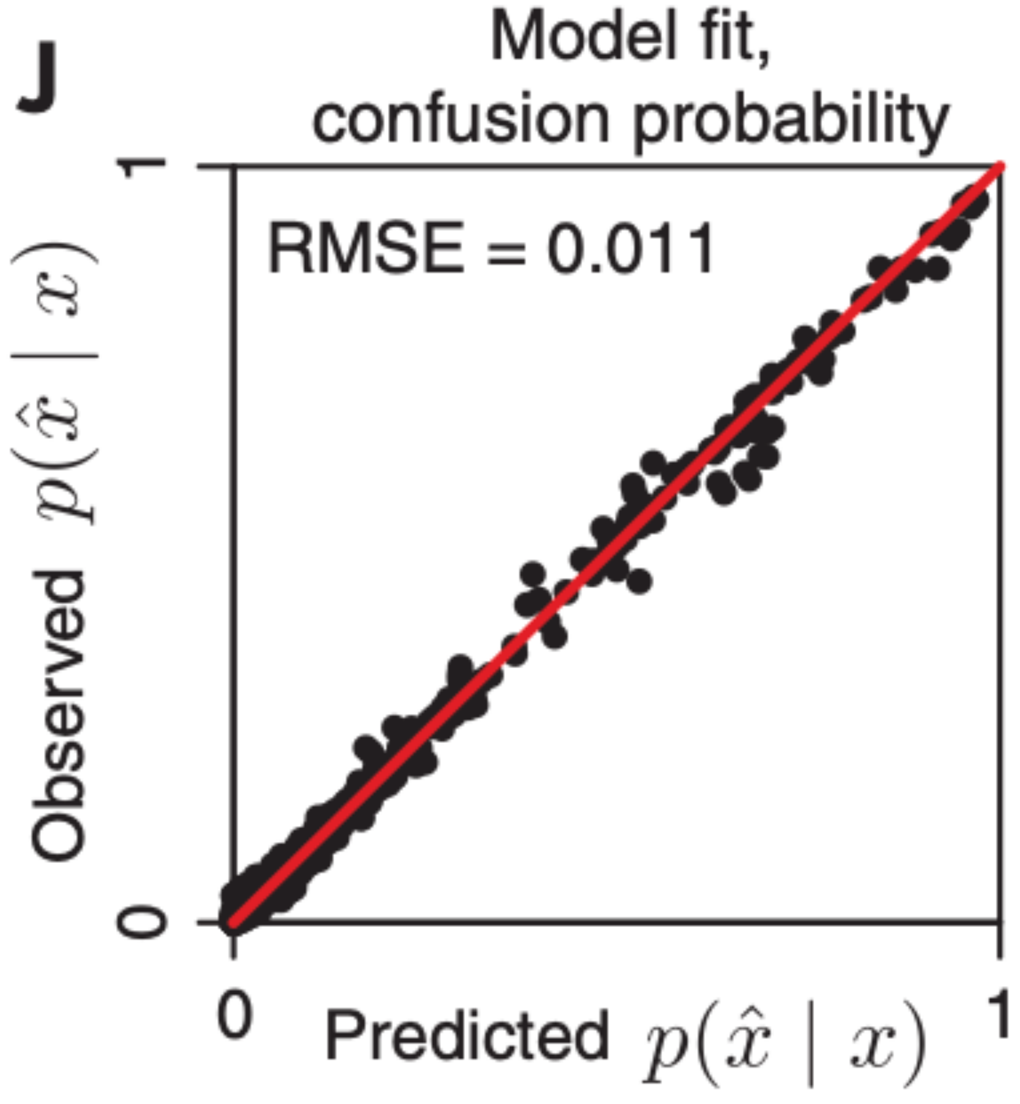
Assuming a simple lossy compression of concepts leads to a generalization pattern that resembles Shepard's law.

(Sims, 2018)

Information theoretic generalization

Model: $\mathcal{G}_{x\hat{x}} = \exp[s\frac{1}{2}(\mathcal{L}(x,\hat{x}) + \mathcal{L}(\hat{x},x) - \mathcal{L}(x,x) - \mathcal{L}(\hat{x},\hat{x}))]$

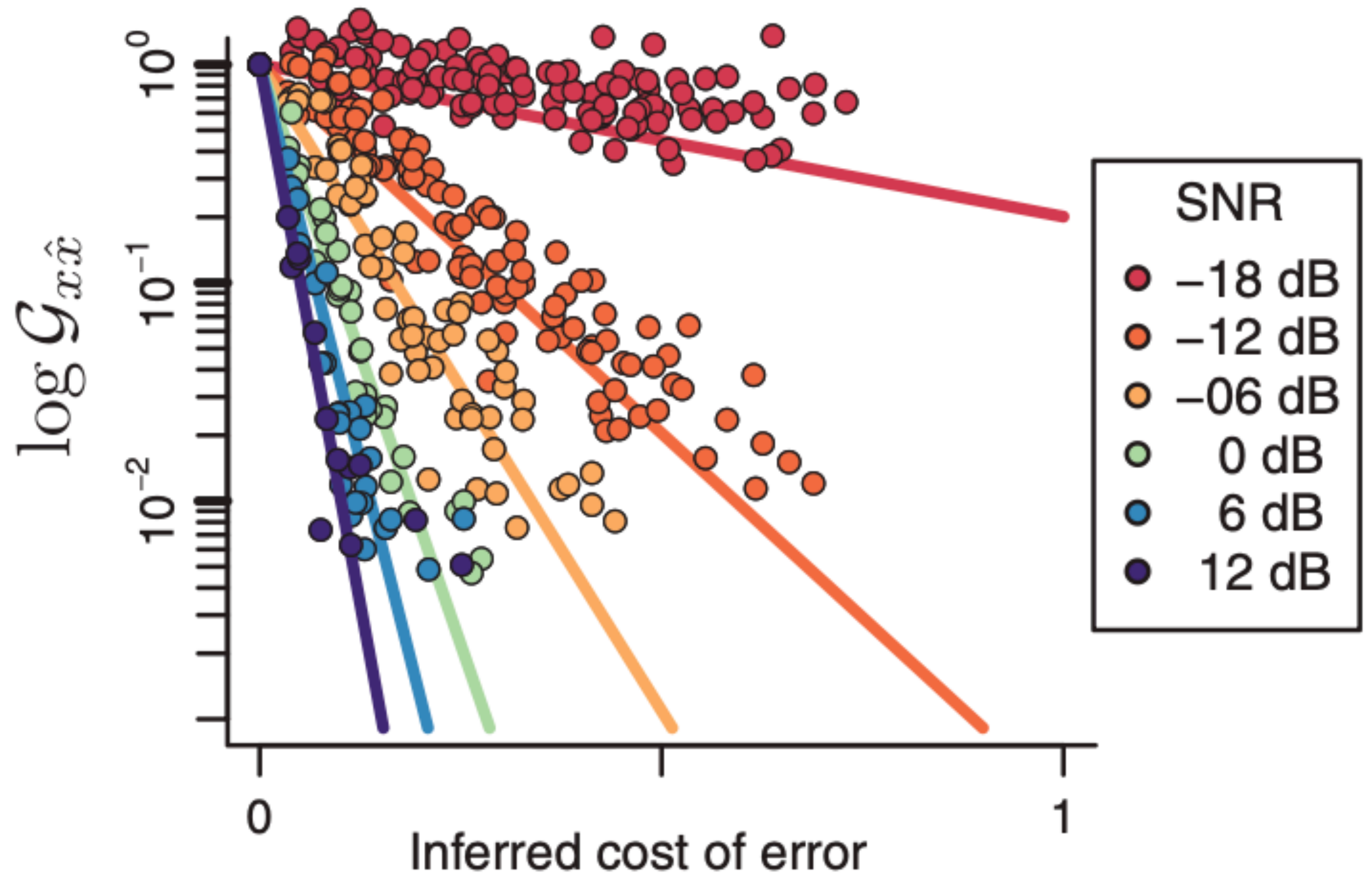


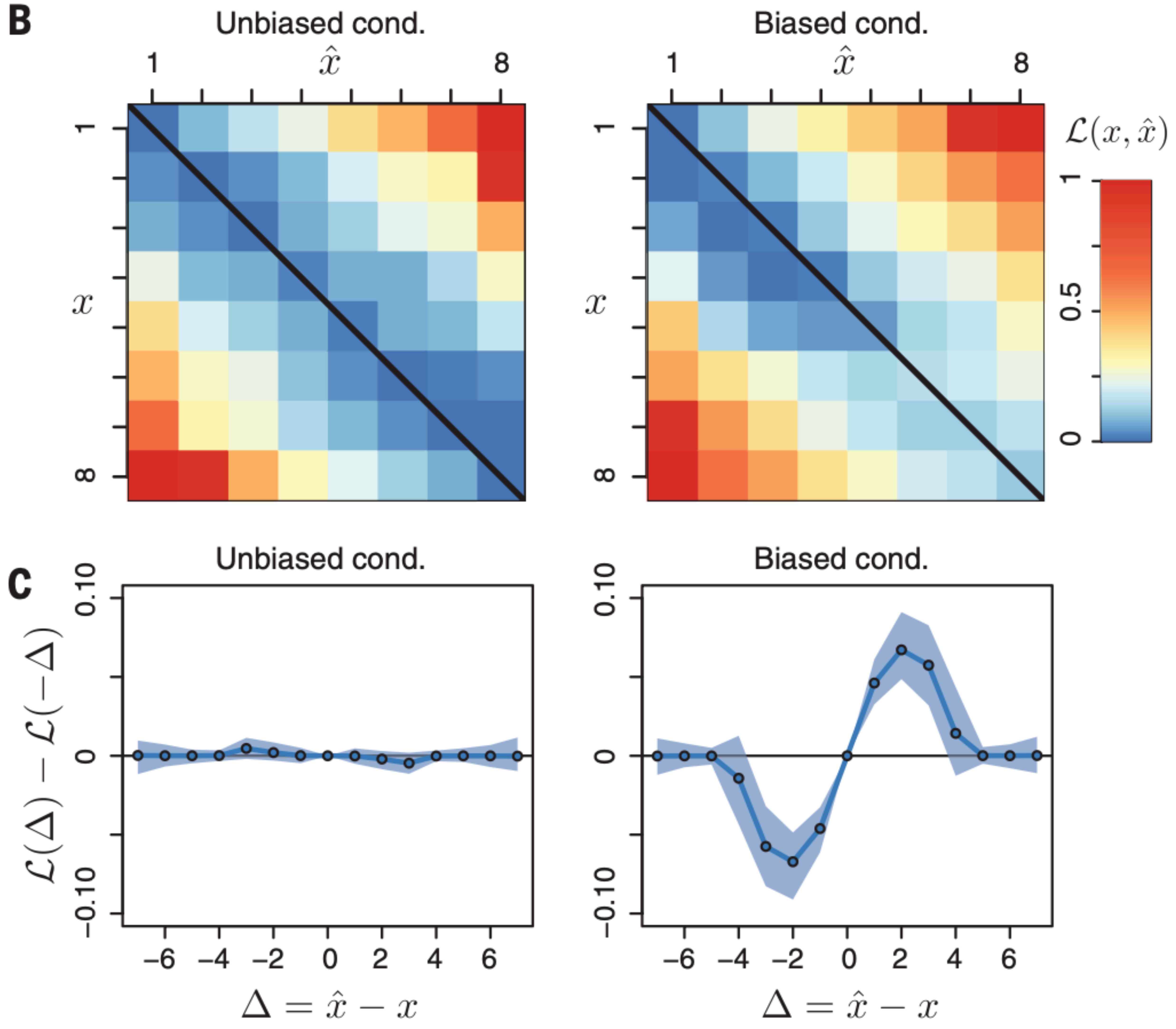


Efficient coding theory explains generalization effects across a variety of experiments

(Sims, 2018)

What rate-distortion theory predicts



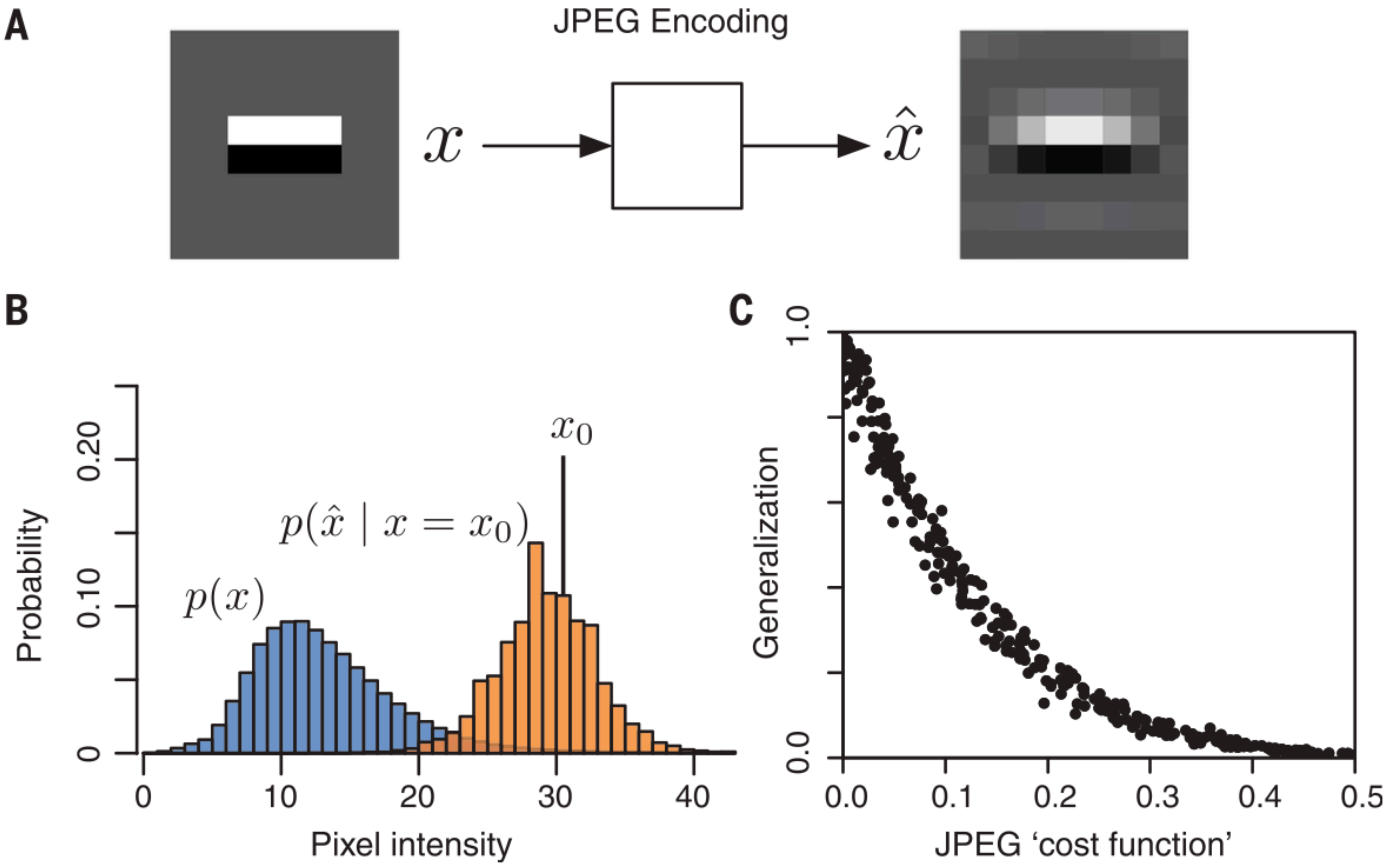


Predictions:

- Optimal generalization gradient: $s = \dfrac{\Delta R}{\Delta D}$
- No requirement for symmetry.
- Extends to artificial systems

(Sims, 2018)

Not just in biological brains



(Sims, 2018)

Take home message

- Concepts (as internal representations) form relational geometries.
- The generalization of new concepts depends on their relative perceptual distance from other concepts.
- This generalization effect can be explained as a result of lossy compression via efficiency coding.