Nonlinear mixed selectivity produces noise-tolerant representations

W. Jeffrey Johnston[1,2], Stephanie E. Palmer[1,3,4,5], and David J. Freedman[1,2,5]
1. Graduate Program in Computational Neuroscience 2. Department of Neurobiology
3. Department of Organismal Biology and Anatomy 4. Department of Physics
5. Grossman Institute for Neuroscience, Quantitative Biology, and Human Behavior
The University of Chicago, Chicago, IL, USA

THE UNIVERSITY OF CHICAGO

Introduction

Reliable transmission and storage of neural representations are essential to computation in the brain. However, little is known about what strategies the brain uses for error reduction and correction.

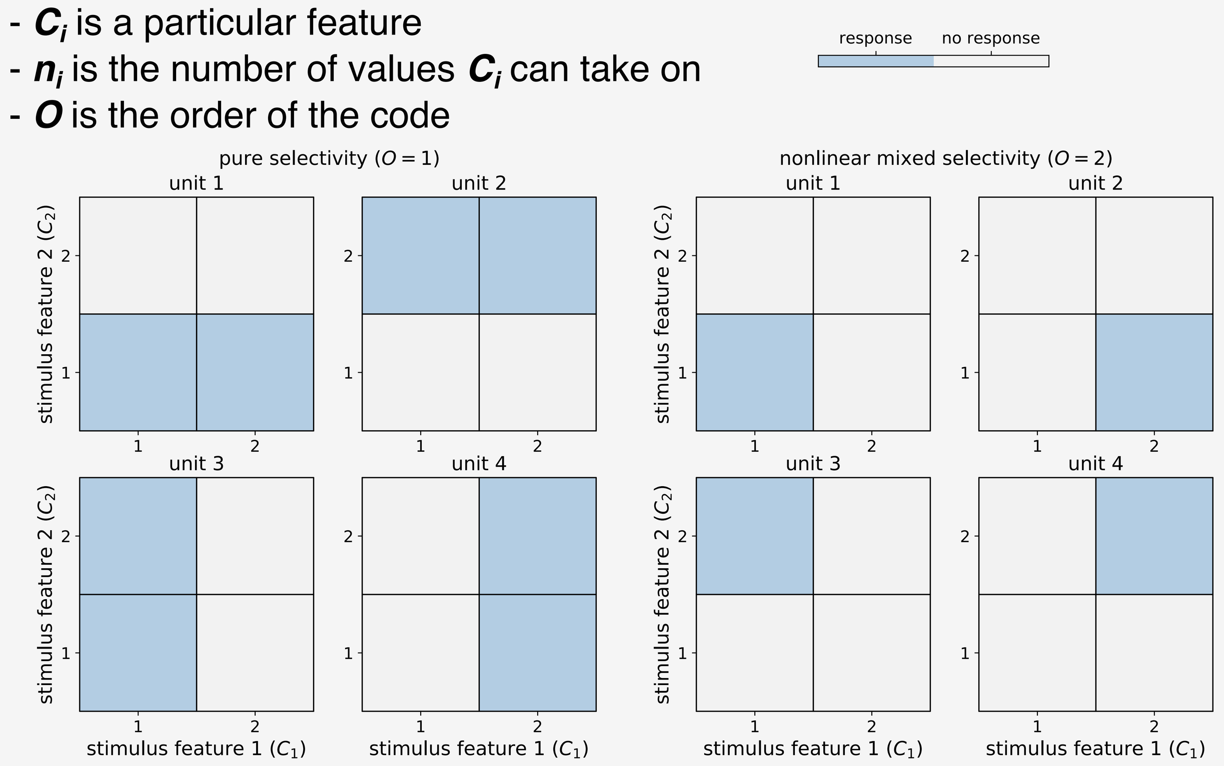
Here, we use ***a channel coding perspective*** to study a family of neural codes that vary along one crucial axis: *the extent to which different representation (or stimulus) features are nonlinearly mixed together.*

We hypothesize that mixing features produces a more noise-tolerant representation than keeping them separate.

What is nonlinear mixed selectivity (NMS)?

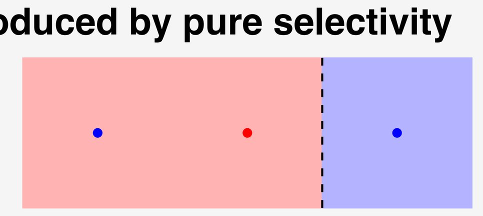
Nonlinear mixed selectivity (NMS) performs dimensionality expansion through code dimensions that respond only to particular combinations of O feature values.

- N_C is the number of features
- C_i is a particular feature
- n_i is the number of values C_i can take on
- O is the order of the code



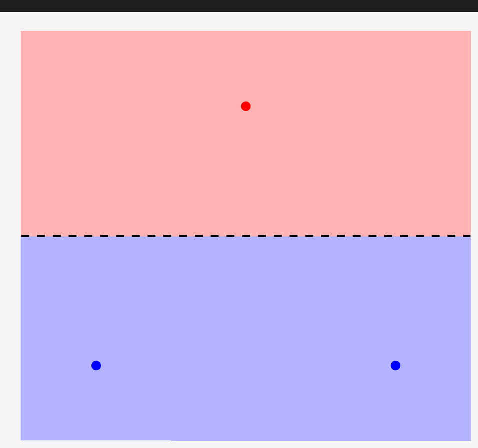
NMS for flexible computation

Low-dimensional representation, produced by pure selectivity



linear decoder fails

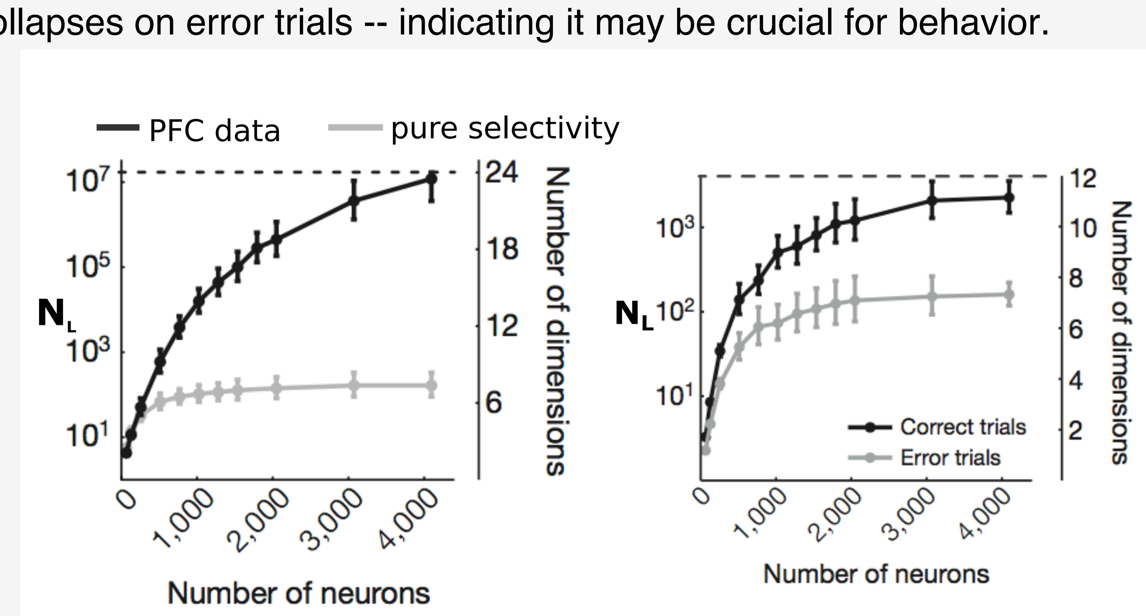
High-dimensional representation, produced by NMS



linear decoder succeeds

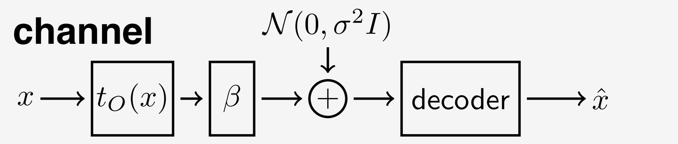
In particular, fully mixed ($O = N_C$) NMS allows a linear decoder to implement all possible partitions of any stimulus set.

NMS exists in the brain and may be involved in computation (Rigotti et al., 2013)
Work by others has shown that the neural population in prefrontal cortex has representations of the maximum possible dimension given the experimental context, as expected from NMS; and, that the dimensionality collapses on error trials -- indicating it may be crucial for behavior.

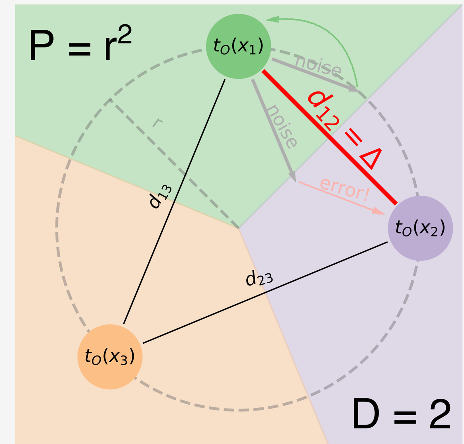


Does nonlinear mixed selectivity (NMS) produce robust representations?

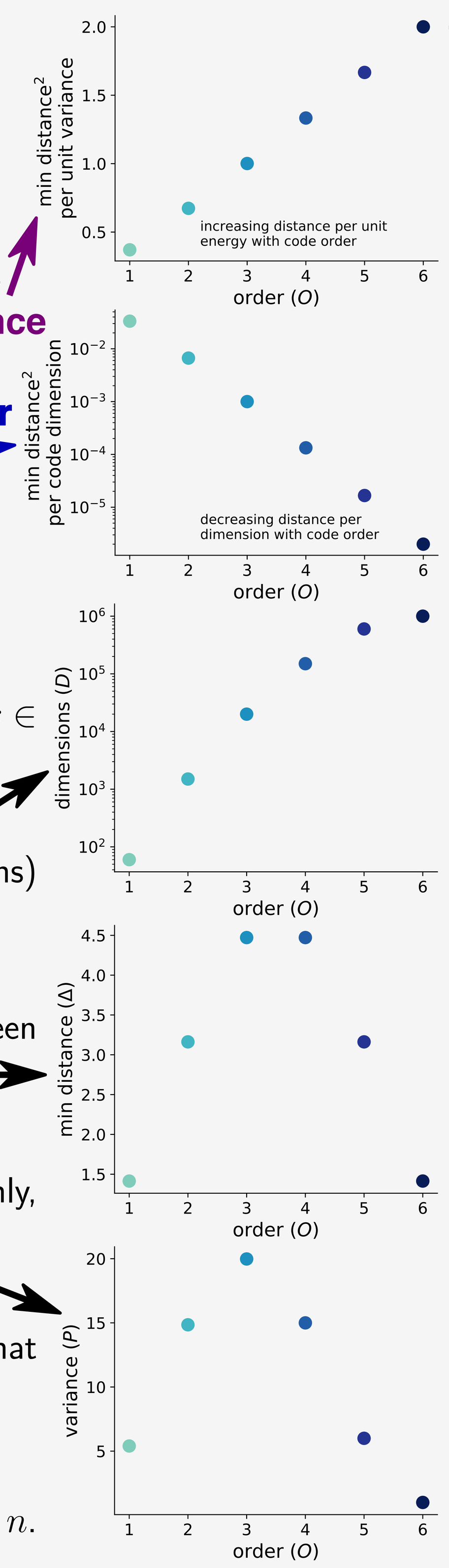
Information transmission down a noisy channel



Code properties



high order codes have **good distance per energy** but **poor distance per dimension**



The **encoding function** $t_O(x)$ operates on stimuli, $x \in X$, and produces codewords.

We will quantify our codes with three values:

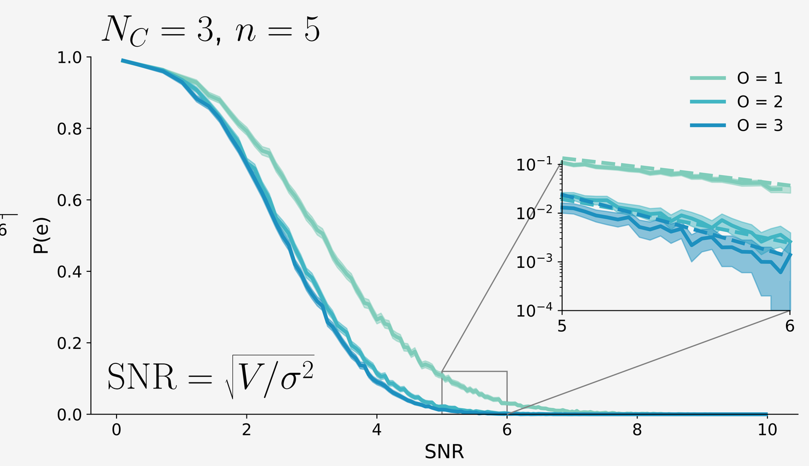
- **dimension**, the number of units (roughly, neurons) used by our code
  - $D = \binom{N_C}{O} n^O$
- **minimum distance**, the smallest distance between any two codewords in our code
  - $\Delta = \left[2\binom{N_C - 1}{O-1}\right]^{\frac{1}{2}}$
- **variance/energy**, the amount of energy (roughly, spikes) used by our code
  - $P = \binom{N_C}{O}$

The linear transform β amplifies our code, such that $P \rightarrow V$, $D \rightarrow M$, and

$$\delta = \sqrt{V/P}\Delta$$

Note: Without affecting our results, we assume $n_i = n$.

For equivalent signal-to-noise ratio, higher order codes make fewer errors than lower order codes

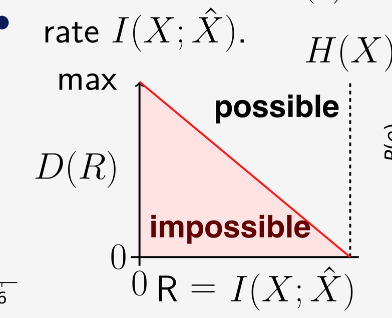


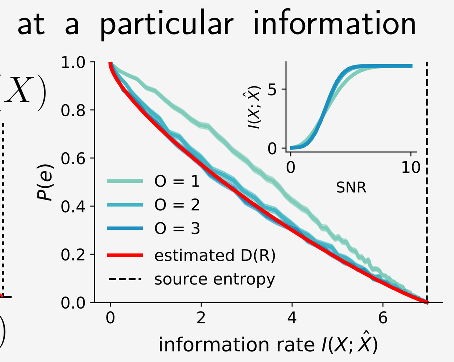
Code performance can be approximated (for relatively high SNR) using the nearest neighbor union bound (NNUB).

$$P(e) \leq N_{\Delta,O} P\left(\mathcal{N}\left(0, \frac{P_O}{V}\sigma^2\right) > \frac{\Delta_O}{2}\right)$$
$$\leq N_{\Delta,O} Q\left[\frac{\Delta_O}{2\sqrt{\frac{P_O}{V}}\sigma}\right]$$

where $Q(x)$ is the ccdf at x of $\mathcal{N}(0,1)$ and $N_{\Delta,O}$ is the number of nearest neighbors.

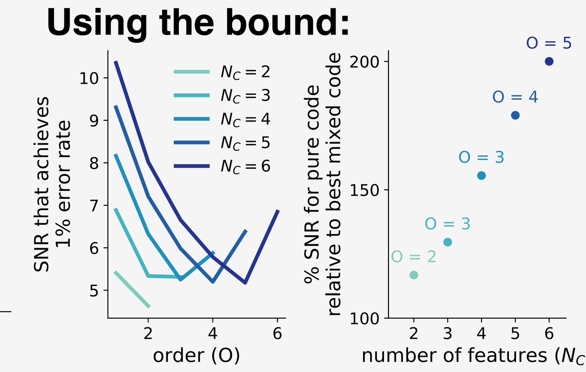
Rate-distortion theory gives the minimum achievable distortion $P(e)$ at a particular information rate $I(X; \hat{X})$.





higher order codes appear to saturate the R-D bound while the *O = 1* code does not

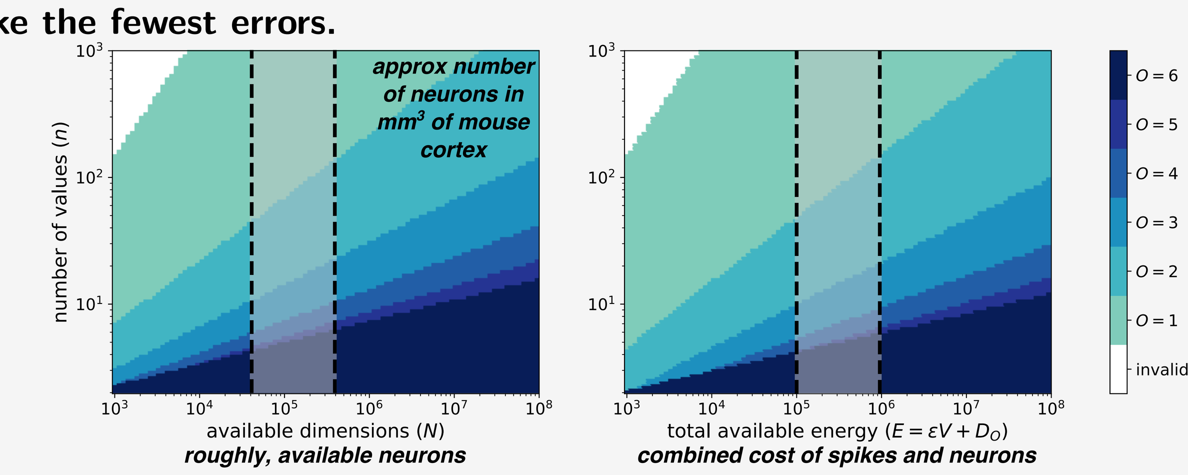
Using the bound:



pure codes use up to twice as much energy as mixed codes

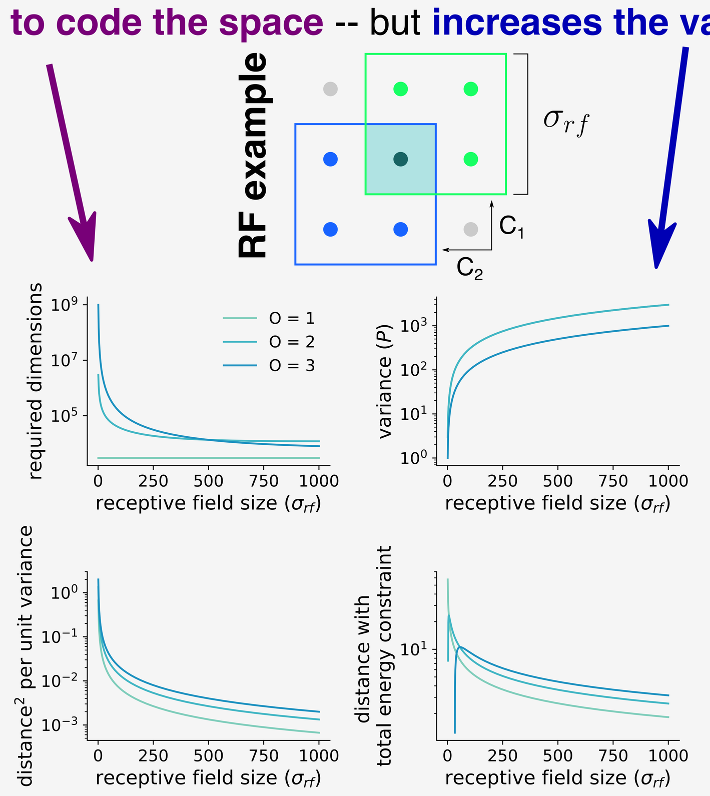
Given N available neurons, **the highest order code such that $D_O \leq N$ will make the fewest errors.**

What about the maintenance cost of silent neurons?



What about receptive fields?

While **holding the size of the coding space constant**, increasing receptive field (RF) size **reduces the number of dimensions necessary to code the space** -- but **increases the variance of each codeword**.

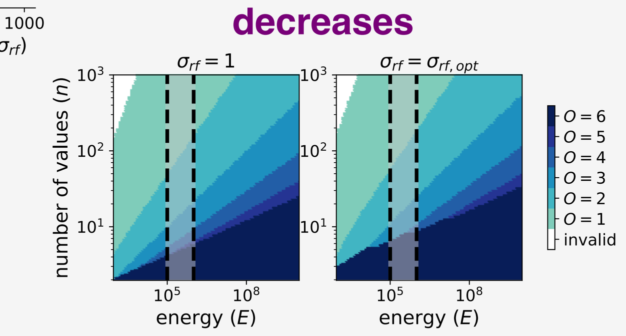


How does this affect our codes?

- **dimension**, for $O > 1$,
  $D = \binom{N_C}{O}\sigma_{rf}\left(\frac{n}{\sigma_{rf}} + 1\right)^O$
- **minimum distance**,
  $\Delta = \left[2\binom{N_C-1}{O-1}\right]^{\frac{1}{2}}$
- **variance/energy**,
  $P = \binom{N_C}{O}\sigma_{rf}$

Now we can examine how this affects the NNUB.

$$P(e) \leq \underbrace{N_{\Delta,O}(\sigma_{rf})}_{\text{decreases}} Q\left[\overbrace{\frac{\sqrt{\frac{V}{P\sigma_{rf}}}\Delta_O}{2\sigma}}^{\text{increases}}\right]$$



We can derive the optimal receptive field size given the total energy constraint

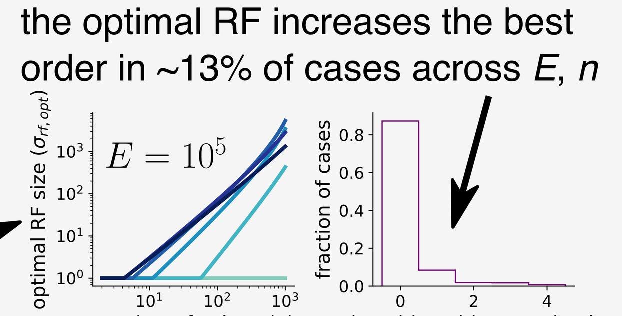
We need to maximize minimum distance for fixed N_C, O, n, and E with respect to receptive field size σ_{rf}.

$$L = \delta\left(N_C, O, \sigma_{rf}, \frac{E - D_O(n, \sigma_{rf})}{\epsilon}\right)^2$$

setting $\frac{\partial L}{\partial \sigma_{rf}} = 0$, we find

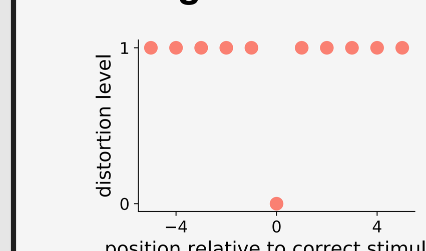
$$\sigma_{rf,\mathrm{opt}} = n\left[\left[\frac{E}{On\binom{N_C}{O}}\right]^{\frac{1}{O-1}} - 1\right]^{-1}$$

the optimal RF increases the best order in ~13% of cases across *E*, *n*



What if the nature of the error matters?

Hamming distortion

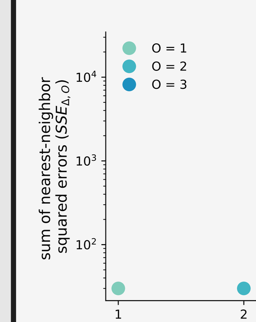


Hamming distortion is written as

$$d_{hamming}(x, \hat{x}) = \begin{cases} 1 & x \neq \hat{x} \\ 0 & x = \hat{x} \end{cases}$$

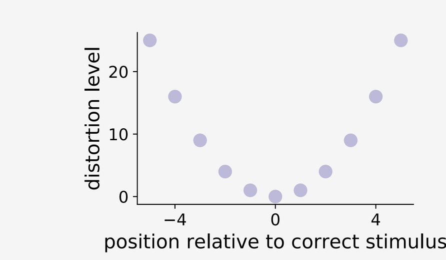
So far, we have been dealing with the **probability of error** ($P(e)$), which is **the mean of the Hamming distortion**,

$$P(e) = P(x \neq \hat{x}) = E\left[d_{hamming}(x, \hat{x})\right]_x$$



nearest neighbors have increased importance with MSE

Squared error distortion (*SE*)



Squared error distortion can be written as

$$d_{SE}(x, \hat{x}) = (x - \hat{x})^T(x - \hat{x})$$

We will use the *mean* squared error distortion (MSE),

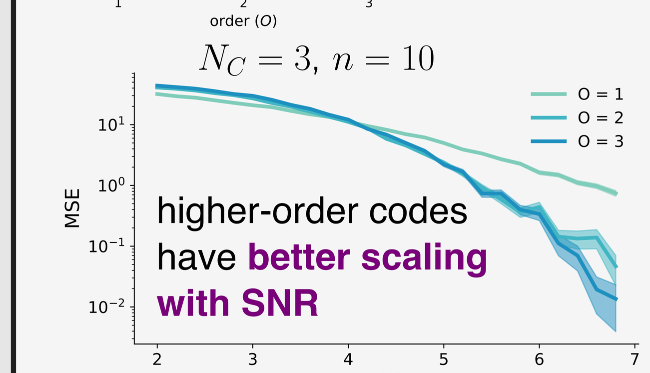
$$\mathrm{MSE} = E\left[d_{SE}(x, \hat{x})\right]_x$$

How does this shift in distortion measure change our results?

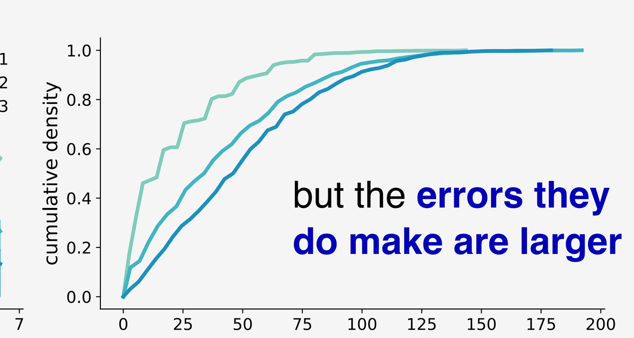
We can use,

$$MSE \approx \mathrm{SSE}_{\Delta,O} Q\left[\frac{\sqrt{\frac{V}{P}}\Delta_O}{2\sigma}\right]$$

as a weak approximation of the MSE where $SSE_{\Delta,O}$ is the sum of squared errors for all the nearest neighbors of a central codeword



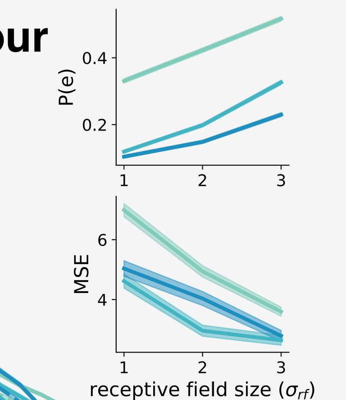
higher-order codes have **better scaling with SNR**

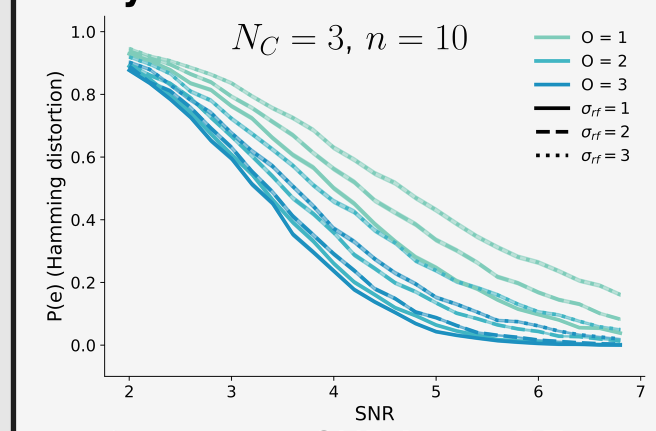


but the **errors they do make are larger**

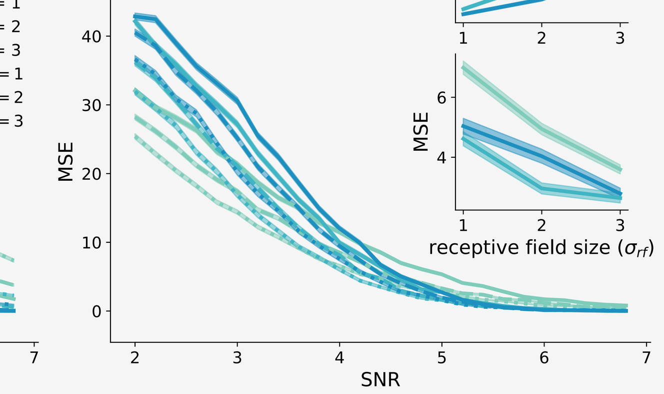
Is NMS appropriate in sensory systems?

Flexible RF sizes and MSE distortion make our analysis more relevant to sensory systems

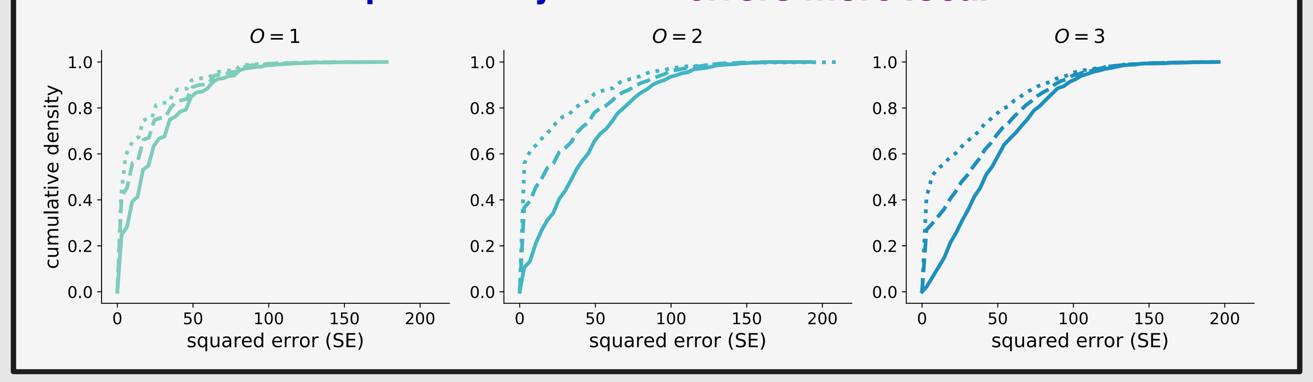




for constant SNR **larger RFs increase error probability**



but they **reduce MSE by making errors more local**

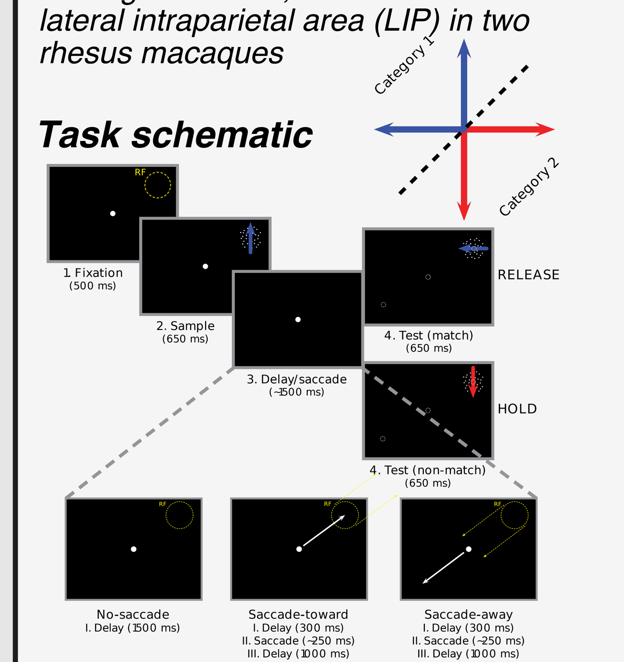


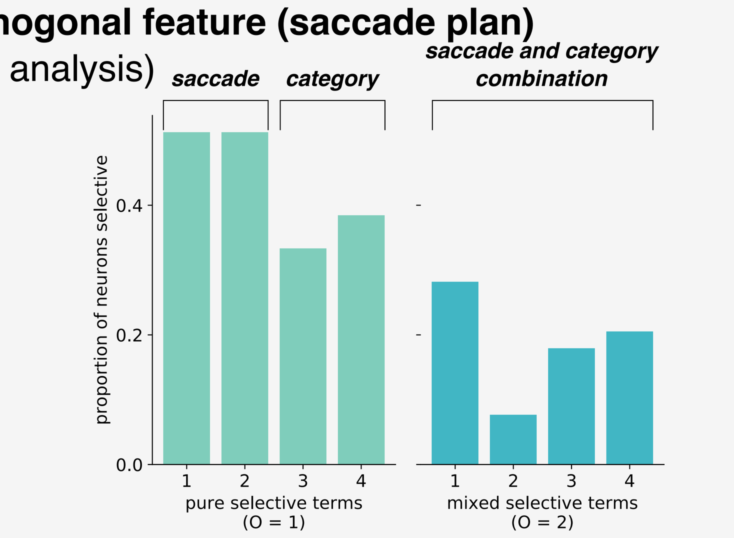
Does the brain exploit these codes?

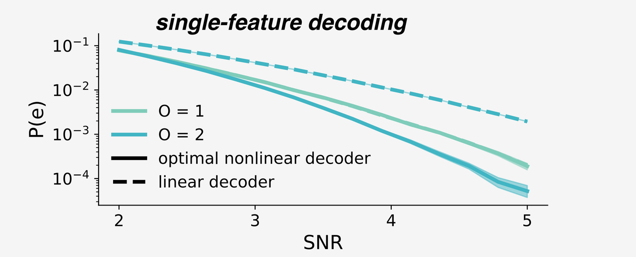
LIP has NMS that mixes a learned decision-relevant feature (category) and a decision-orthogonal feature (saccade plan) (data in Rishel et al., 2013; new analysis)

70 single neurons, recorded from the lateral intraparietal area (LIP) in two rhesus macaques

Task schematic







this is **not explained by the NMS for flexible computation framework**, but it is **expected under our error-correction framework**

Conclusions

We show that nonlinearly mixing stimulus features leads to more-robust representations.

In particular,

- **NMS produces codes with** more minimum distance per unit of energy, and therefore **lower overall error**
- allowing flexible receptive fields can make NMS advantageous in a wider range of energy and stimulus-space contexts
- together, **mean squared error (MSE) and flexible receptive fields make NMS applicable to sensory contexts** -- in which it continues to outperform pure selectivity
- study of NMS for task-relevant and task-orthogonal features indicates that NMS may be exploited for error-correction in the brain

NMS is suitable for, and appears to be used by the brain for, error-correction.

References and acknowledgments

We are grateful for funding from:
- NIH F31EY029155
- NIH R01EY019041
- NSF-NCS
- NSF CAREER award

contact: wjeffreyjohnston@gmail.com

Rigotti et al. (2013) *Nature*
Cover & Thomas (2006) *Elements of Information Theory*
Levy & Baxter (1996) *Neural Computation*