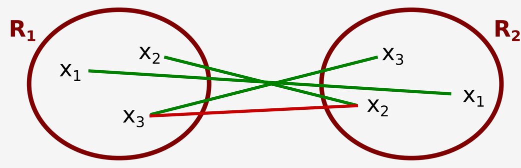
Two tradeoffs between local accuracy and catastrophic errors in a solution to the representation assignment problem

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

THE UNIVERSITY OF CHICAGO

Introduction

The assignment problem arises when the brain has multiple distinct neural representations of the same stimuli.

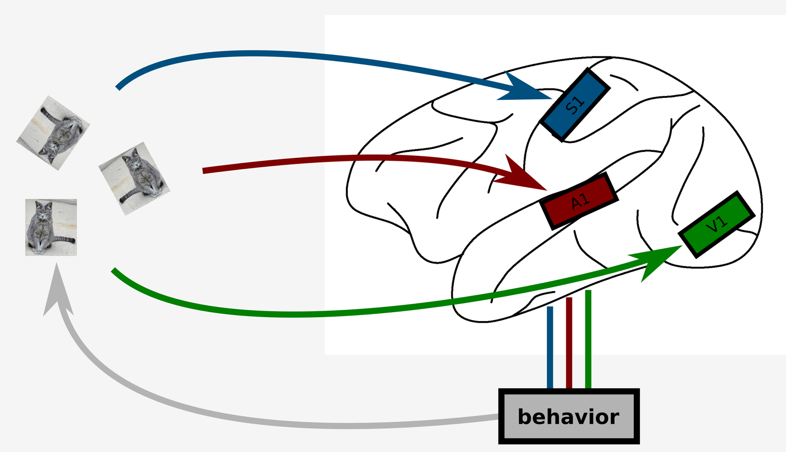


where x_i are stimuli and R_j are brain regions

These distinct representations arise due to different sensory systems and distributed neural representations in cortex and represent a particular facet of the more general binding problem[1]. Failures of assignment lead to the creation of fictitious percepts that would likely be particularly catastrophic for behavior.

In this work, we ask:
- How can the assignment problem be solved?
- What are the tradeoffs that the brain must navigate?
- What kinds of assignment errors does this solution predict?

Distinct senses produce distinct representations

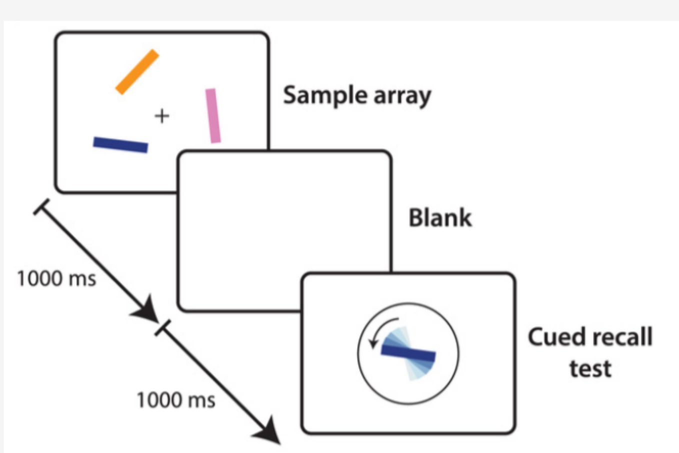


Distinct sensory systems produce multiple distinct representations of the same objects.
- How can these distinct representations be combined?
- How can the brain perform this combination efficiently?

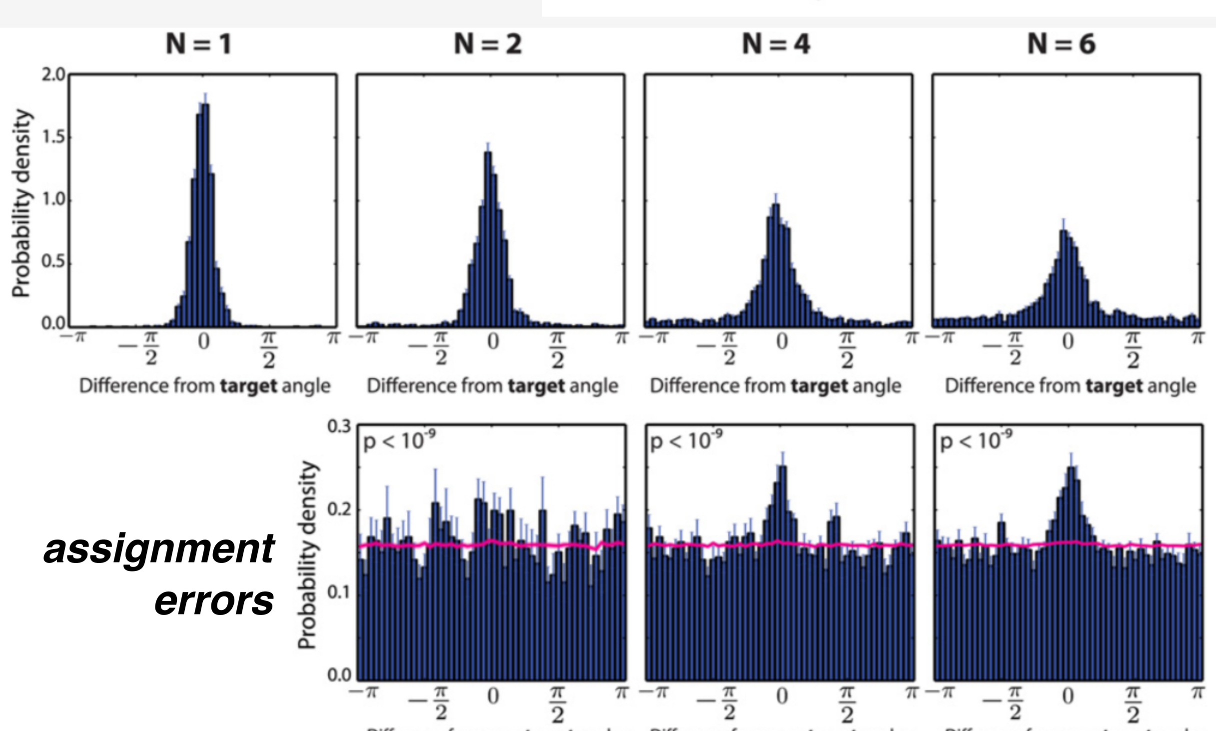
Humans make assignment errors

Assignment errors have been observed even for representations within a single sensory modality, indicating that the distributed nature of sensory representations poses a problem for neural computation.

In this task, human subjects had to recall the bar angle that corresponds to a cued bar color after a memory delay.



Color and orientation had to be correctly assigned.

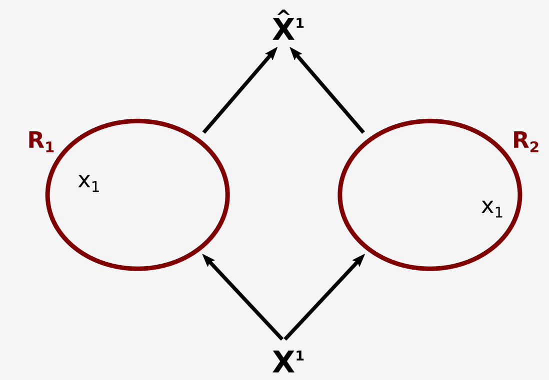


Matthey, Bays, Dayan (2015), PLoS CB

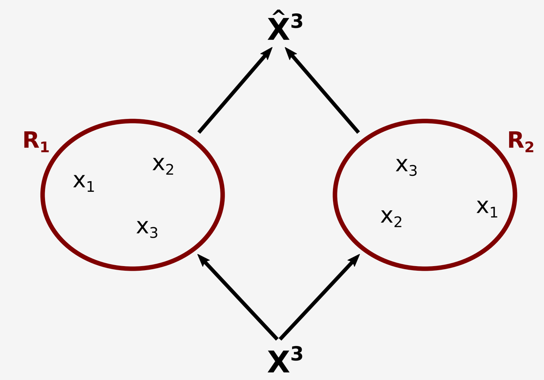
A solution to the representation assignment problem

The problem

For a single stimulus, there is only one possible mapping.



For N stimuli, there are N! possible mappings, only one is correct.



If I(R₁; R₂) = 0, all of those N! possible mappings are equally good explanations of the neural activity, and **the probability of an assignment error** is 1 - 1/N!, which **is near 1** for even moderate N.

This can be seen from the maximum likelihood solution: $\hat{X}^N = \underset{X^N}{\operatorname{argmax}}\, p(R_1|X^N)\, p(R_2|X^N)\, p(X^N)$

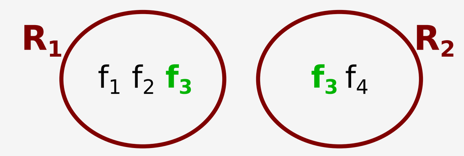
The setup

To formalize this problem, we assume:
- **N** stimuli
- **K** features each
- stimuli are **uniformly distributed** in feature space
- information about the stimuli is split across **two neural populations**

The solution

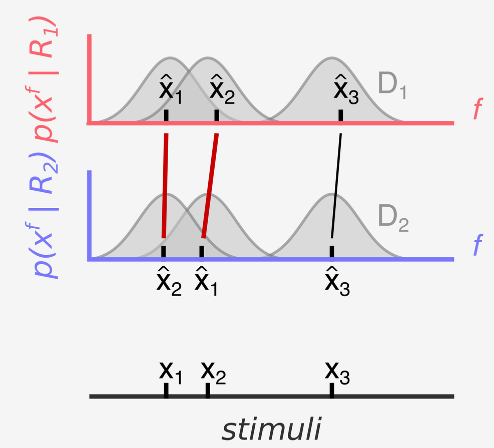
Both regions represent the same stimulus feature.

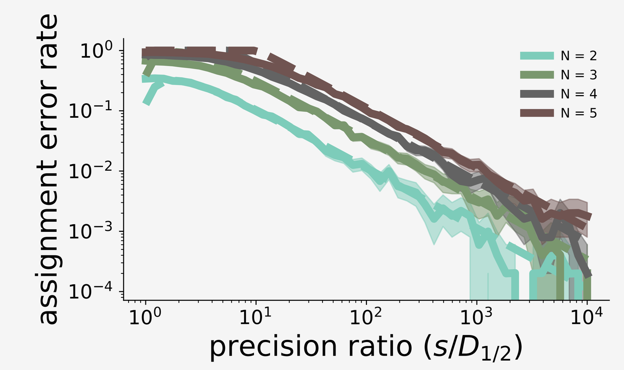
For K = 4,



e.g., ***stimulus position***

For a representation of this overlapping feature with optimal estimator variance D, we can quantify the assignment error rate.

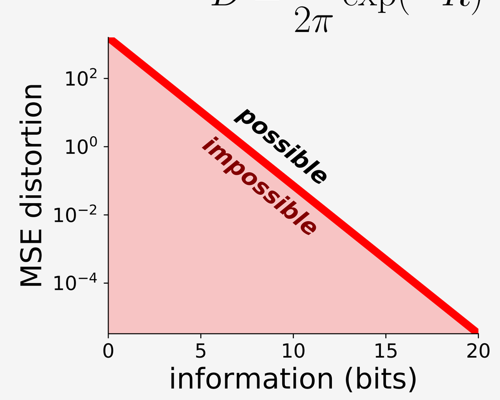




$$P(E) \approx \binom{N}{2}\left[\frac{\sqrt{2}\sqrt{D_1+D_2}}{s\sqrt{\pi}} - \frac{D_1+D_2}{2s^2}\right]$$

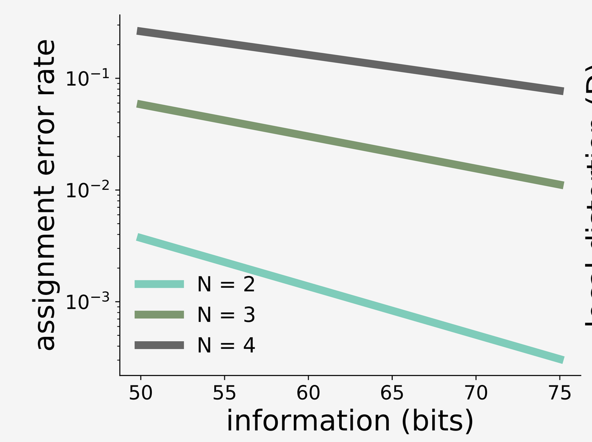
How can we quantify solution efficiency?

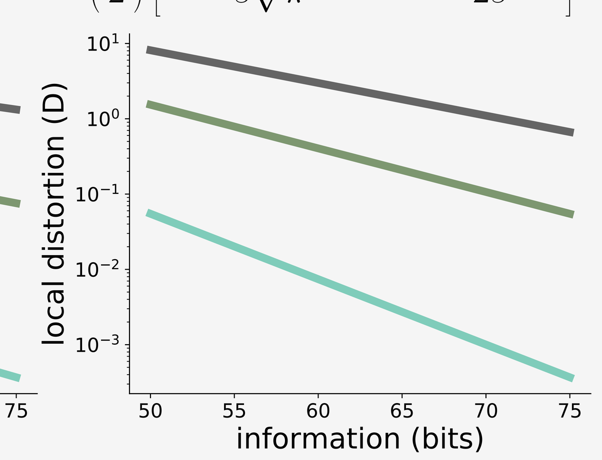
$$D = \frac{s^2}{2\pi}\exp(-R)$$



The rate-distortion bound[2] tells us the **minimum number of bits of information (R)** necessary to **achieve** a given level of **distortion (D)**.

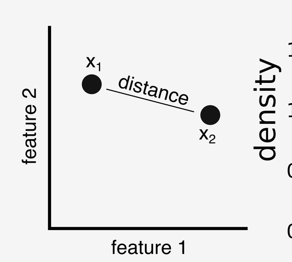
- We assume that the brain achieves this bound.
- We investigate the tradeoffs between different particular solutions.

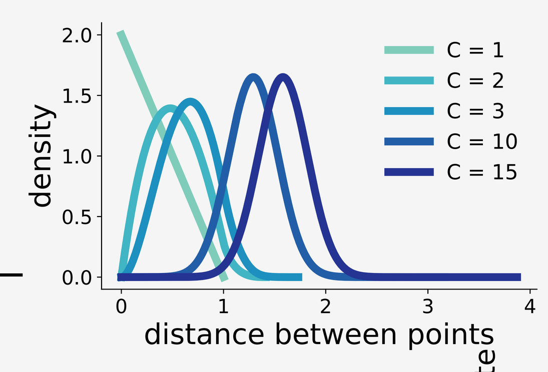




Feature overlap decreases assignment error rate

We want the **distribution of distances between two uniformly distributed points** in a hypercube of particular dimension[3,4].

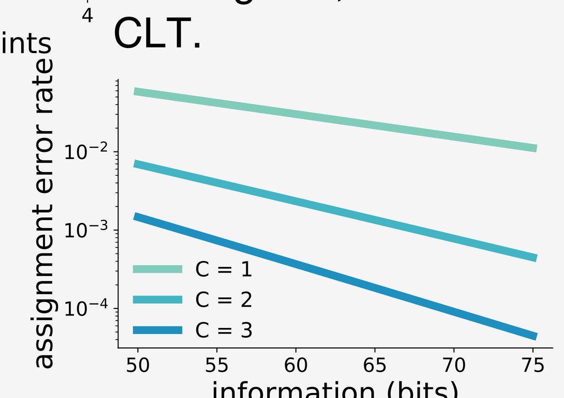




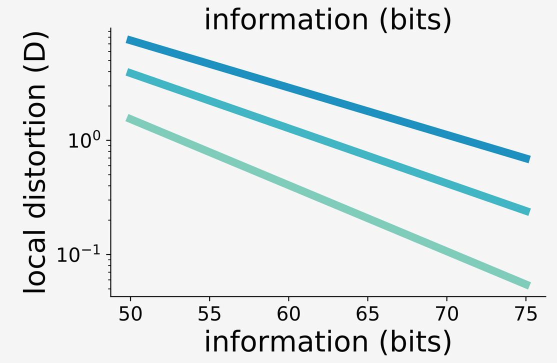
For features assumed to be of size *s = 1*.

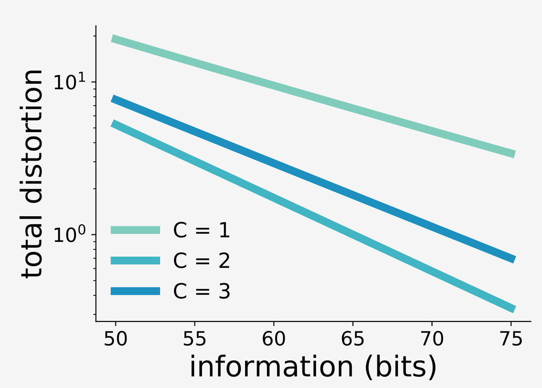
There are no known exact solutions for $C > 3$. For larger C, we use the CLT.

Increasing the number of overlapping features decreases the assignment error rate, due to the change in distance distribution, as shown above.



but it also increases the distortion of each feature given a fixed number of bits, due to the increased redundancy.



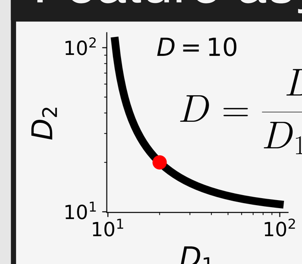


How does the brain weight these different forms of error?

$$D_{\text{total}} = D_{\text{local}} + \lambda P(AE)$$

$$\lambda = (K - C)\, s$$

Feature asymmetry decreases local distortion



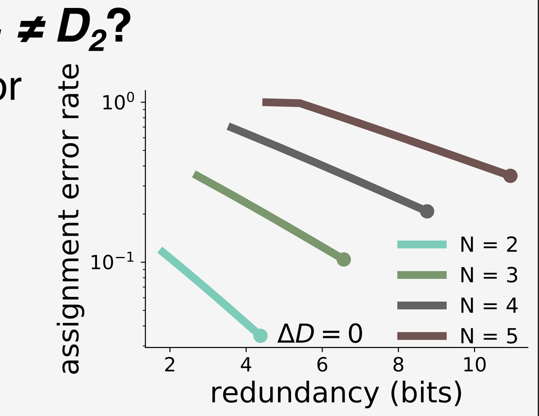
$$D = \frac{D_1 D_2}{D_1 + D_2}$$

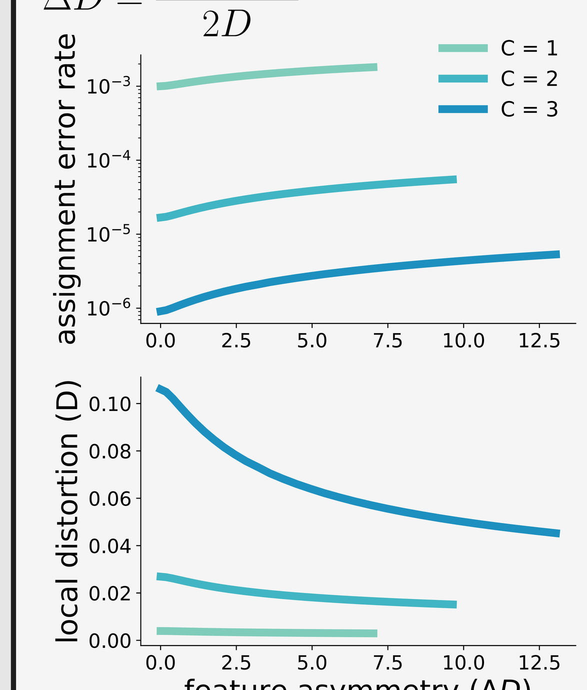
The two representations of the same features with D_1 and D_2 are optimally combined to reduce the local distortion to D.

Is there a benefit from setting $D_1 \neq D_2$?

$D_1 = D_2$ has the lowest assignment error rate, but the most wasted information.

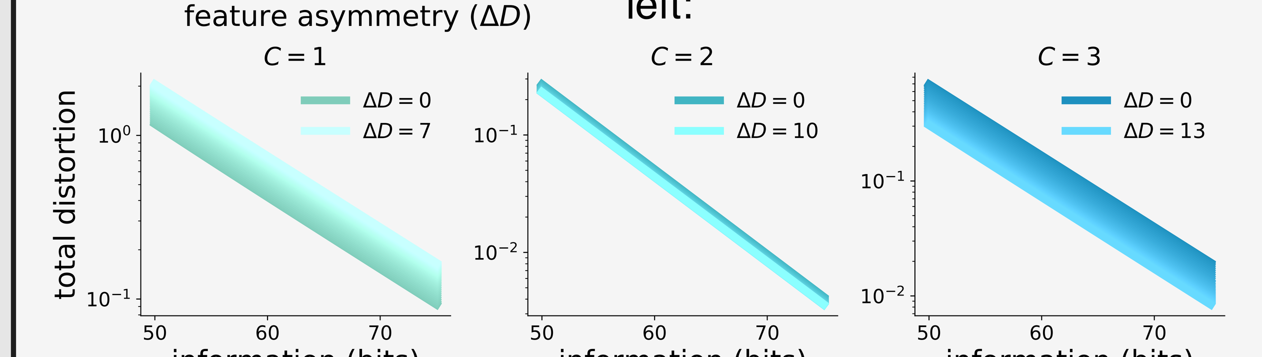
$$\Delta D = \frac{D_1 - 2D}{2D}$$





Asymmetric feature representations increase the assignment error rate due to decreased redundancy, **but decrease the local distortion** due to increased efficiency.

Using the same weighting as on the left:

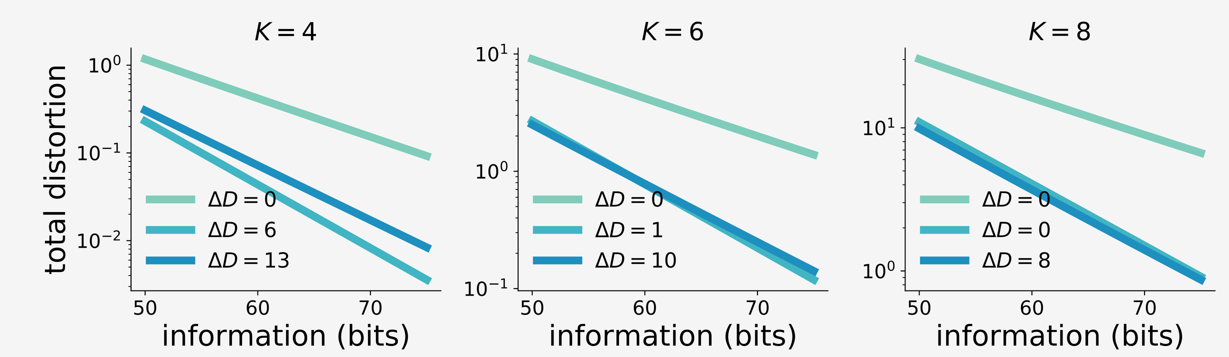


Conclusions

We demonstrate and then analyze a straightforward solution to the representation assignment problem. **The assignment problem can only be solved if the two representations have some overlapping information.** However, too much overlapping information will make the entire representation less accurate. The brain must navigate a tradeoff between **redundancy for solving the assignment problem** and **efficiency for representing stimulus information.**
- More overlapping features increase redundancy at the cost of efficiency.
- Asymmetric feature representations increase efficiency at the cost of redundancy.
- The optimal solution typically leverages both of these tradeoffs.

How is local distortion minimized?

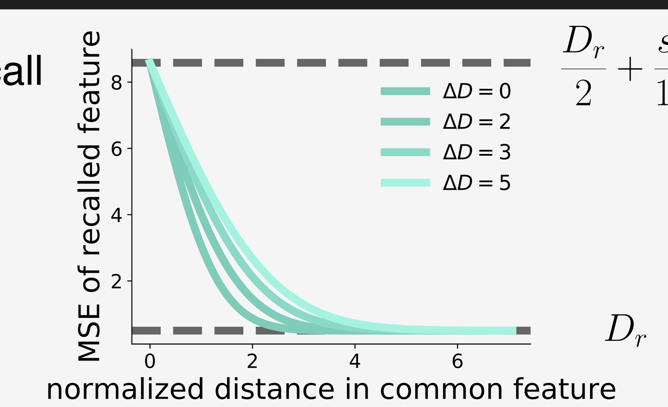
In many cases, **the optimal solution leverages both tradeoffs between redundancy and efficiency.**



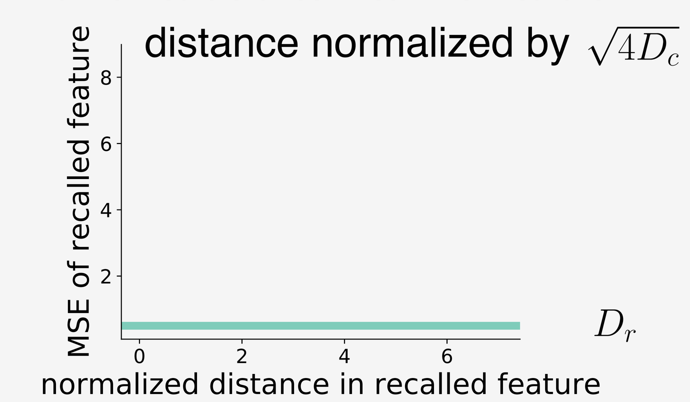
The solution that minimizes total distortion also depends on the number of features represented across the two populations: **more features overall increases the optimal level of feature overlap.**

Predictions for experimental data

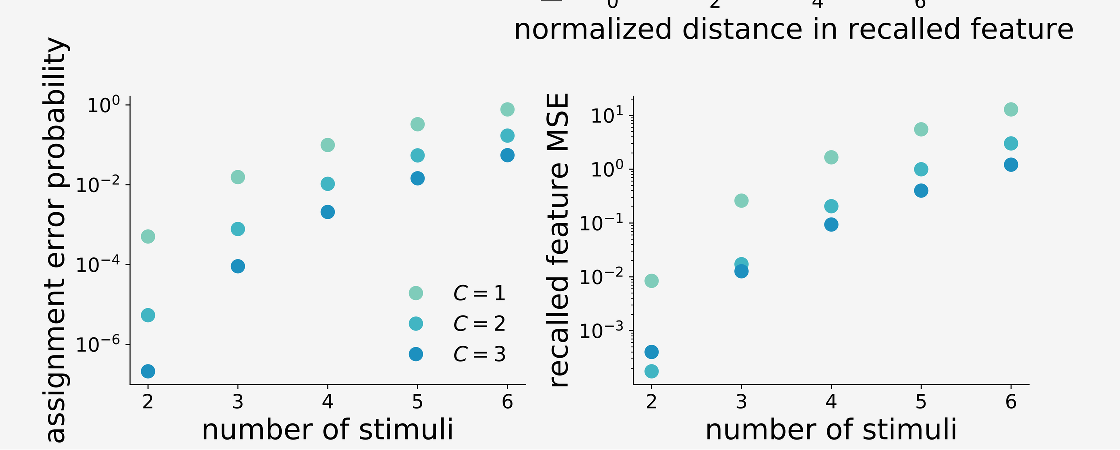
For a cross-modality cued recall task, we predict a particular pattern of errors.



distance normalized by $\sqrt{4D_c}$



The distortion of the common feature (D_c) and the recalled feature (D_r) can be estimated from experimental data.



References and acknowledgments

We are grateful for funding from:
- NIH F31EY029155
- NIH R01EY019041
- Pritzker Fellowship

[1] Treisman (1996) *Curr. Op. in Neurobiol.*
[2] Cover & Thomas (2012)
[3] Weisstein. "Square Line Picking." *MathWorld*
[4] Weisstein. "Cube Line Picking." *MathWorld*

contact: wjeffreyjohnston@gmail.com