南方科技大学 SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

生物医学工程系 Department of Biomedical Engineering

NEURAL COMPUTING CONTROL 神经计算与控制实验室 NCC lab

Best Paper of 2022

Disentangling with Biological Constraints: A Theory of Functional Cell Types

Author: James C.R. Whittington, Will Dorrell, Surya Ganguli, Tim Behrens

Presenter: Ziyuan Ye

Content

- Introduction to Tim Behrens
- Background & previous puzzles
- Disentangling in machines
- Disentangling in brain
- Take-home message

Content

- Introduction to Tim Behrens
- Background & previous puzzles
- Disentangling in machines
- Disentangling in brain
- Take-home message

Information about Tim Behrens

Tim Behrens

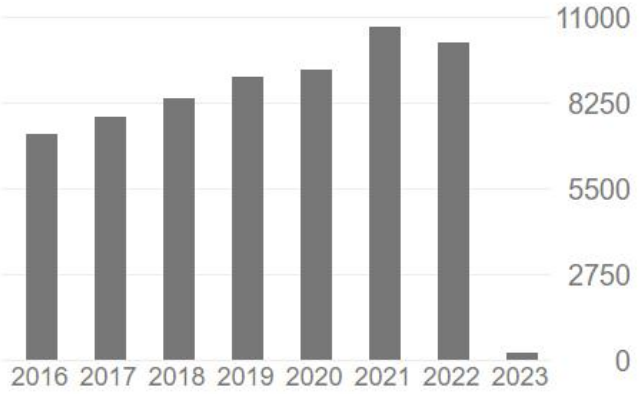
Professor of Computational Neuroscience, University of Oxford. Honorary Prof, UCL
在 fmrib.ox.ac.uk 的电子邮件经过验证

Computational Neuroscience Behavioral Neuroscience Decision Making Learning
Brain connectivity

标题	引用次数	年份
Advances in functional and structural MR image analysis and implementation as FSL SM Smith, M Jenkinson, MW Woolrich, CF Beckmann, TEJ Behrens, ... Neuroimage 23, S208-S219	12847	2004
Fsl M Jenkinson, CF Beckmann, TEJ Behrens, MW Woolrich, SM Smith Neuroimage 62 (2), 782-790	8338	2012
Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data SM Smith, M Jenkinson, H Johansen-Berg, D Rueckert, TE Nichols, ... Neuroimage 31 (4), 1487-1505	6578	2006
The WU-Minn human connectome project: an overview DC Van Essen, SM Smith, DM Barch, TEJ Behrens, E Yacoub, K Ugurbil, ... Neuroimage 80, 62-79	4012	2013
Probabilistic diffusion tractography with multiple fibre orientations: What can we gain? TEJ Behrens, HJ Berg, S Jbabdi, MFS Rushworth, MW Woolrich neuroimage 34 (1), 144-155	3595	2007

引用次数 查看全部

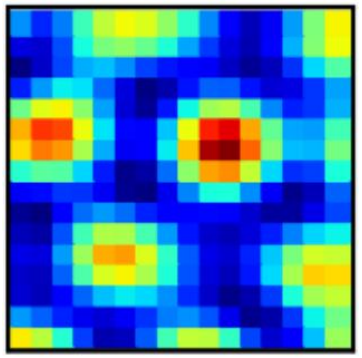
	总计	2018 年至今
引用	96588	47953
h 指数	115	86
i10 指数	191	180



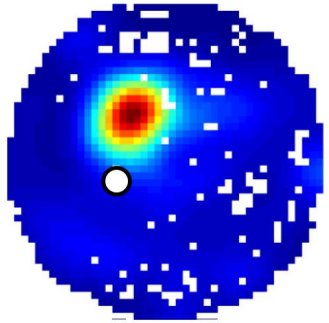
Content

- Introduction to Tim Behrens
- Background & previous puzzles
- Disentangling in machines
- Disentangling in brain
- Take-home message

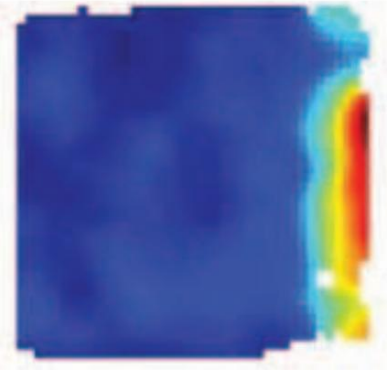
Why many different bespoke cellular responses exist for physical space?



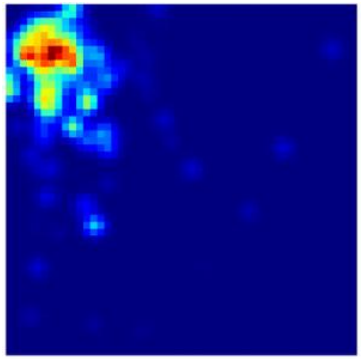
Grid cell



Object-vector cell

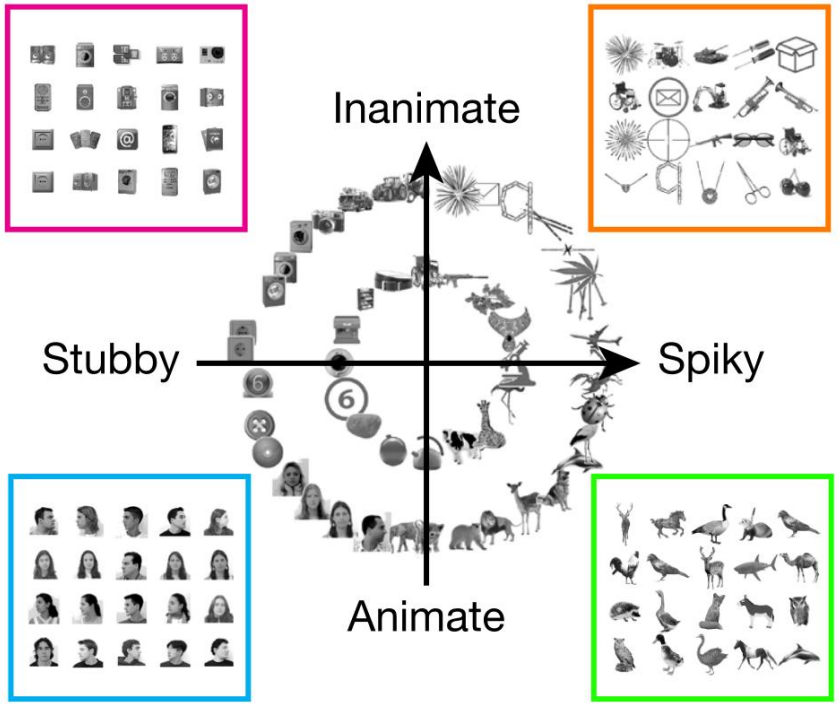


Border cell

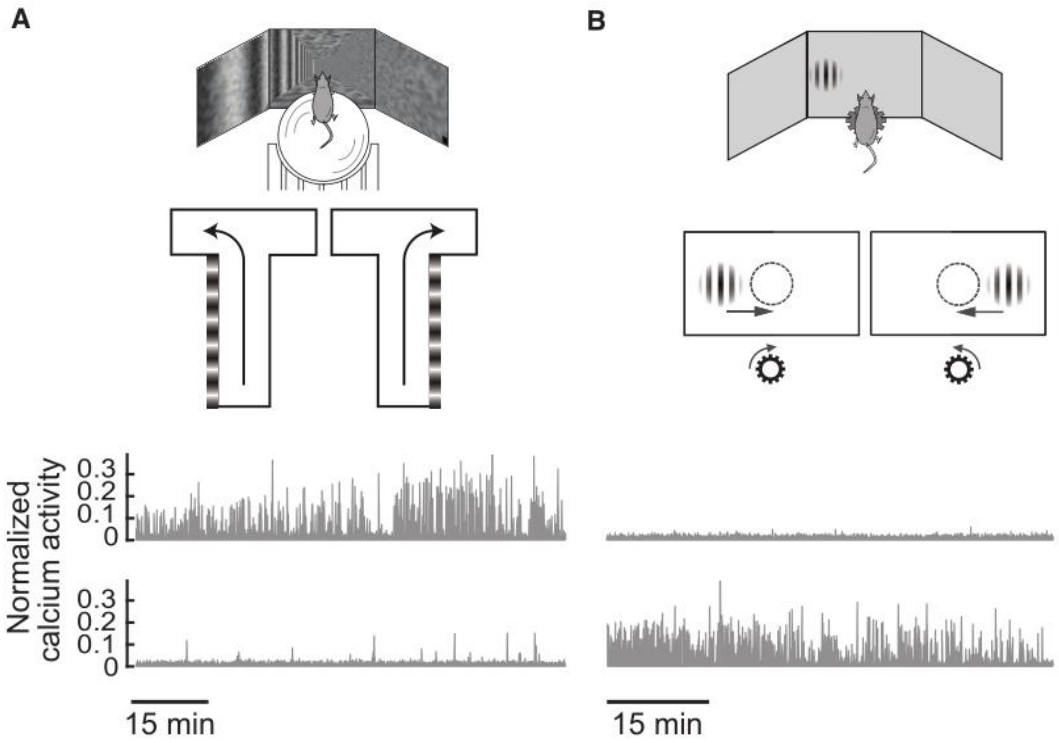


Place cell

Why many different bespoke cellular responses exist in different tasks?



Bao et al. (2020) Nature



Lee et al. (2022) Neuron

Why are some neural representations entangled and others not?

Main contributions:

- **From biological aspect**: This paper shows the most efficient biological representation puts different factors in different neurons.
- **From machine aspect**: This paper builds machines that learn disentangling representations with simple biological constraints of nonnegativity and minimising neural activity energy.

Content

- Introduction to Tim Behrens
- Background & previous puzzles
- Disentangling in machines
- Disentangling in brain
- Take-home message

Linear disentangling with biological constraints

Linear model

$$z = Me + b_z$$

Stimuli: $e \in \mathbb{R}^k$ with k independent components;
Mean(e_i)= 0, Var(e_i)=σ^2
Neural representation: z
Weights: $M \in \mathbb{R}^{n \times k}$
Bias:$b_z \in \mathbb{R}^n$

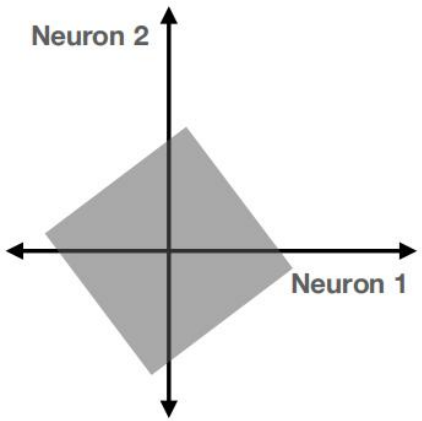
Biological constraints

- ✓ The neural representation is **nonnegative** with $z_i \geq 0$ for all $i = 1, \ldots, n$
- ✓ **Minimising neural activity energy**: min $(E||z||^2)$

$$z = e_1 \begin{bmatrix} 0.8 \\ 0.6 \end{bmatrix} + e_2 \begin{bmatrix} -0.6 \\ 0.8 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\sum_j Var(z_j) = 2\sigma^2$$

$$\mathbb{E}||z||^2 = 2\sigma^2$$

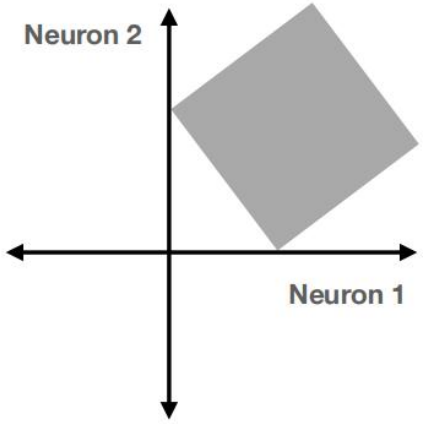


Make non-negative

$$z = e_1 \begin{bmatrix} 0.8 \\ 0.6 \end{bmatrix} + e_2 \begin{bmatrix} -0.6 \\ 0.8 \end{bmatrix} + a \begin{bmatrix} 1.4 \\ 1.4 \end{bmatrix}$$

$$\sum_j Var(z_j) = 2\sigma^2$$

$$\mathbb{E}||z||^2 = 2\sigma^2 + 2a^2 1.4^2$$

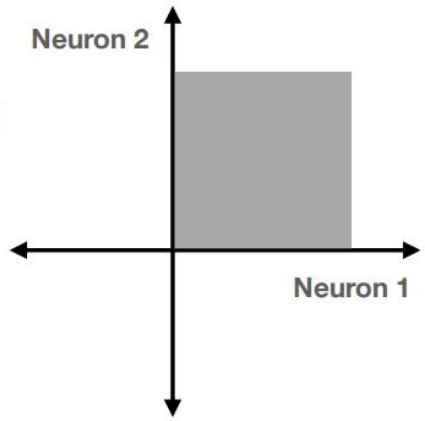


Minimise activity energy

$$z = e_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + e_2 \begin{bmatrix} 0 \\ 1 \end{bmatrix} + a \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$\sum_j Var(z_j) = 2\sigma^2$$

$$\mathbb{E}||z||^2 = 2\sigma^2 + 2a^2$$



Disentangling in machines

Regularizers as constraints

$$\mathcal{L} = \underbrace{\mathcal{L}_{\text{nonneg}} + \mathcal{L}_{\text{activity}} + \mathcal{L}_{\text{weight}}}_{\text{Biological constraints}} + \underbrace{\mathcal{L}_{\text{prediction}}}_{\text{Functional constraints}}$$

Disentangling in machines

Regularizers as constraints

$$\mathcal{L} = \mathcal{L}_{\text{nonneg}} + \mathcal{L}_{\text{activity}} + \mathcal{L}_{\text{weight}} + \mathcal{L}_{\text{prediction}}$$

Biological constraints

Functional constraints

$$\mathcal{L}_{\text{nonneg}} = \beta_{\text{nonneg}} \sum_i \max(-a_i, 0)$$

a_i denotes single neural activity

Disentangling in machines

Regularizers as constraints

$$\mathcal{L} = \underbrace{\mathcal{L}_{\text{nonneg}} + \mathcal{L}_{\text{activity}} + \mathcal{L}_{\text{weight}}}_{\text{Biological constraints}} + \underbrace{\mathcal{L}_{\text{prediction}}}_{\text{Functional constraints}}$$

$$\mathcal{L}_{\text{nonneg}} = \beta_{\text{nonneg}} \sum_i \max(-a_i, 0)$$

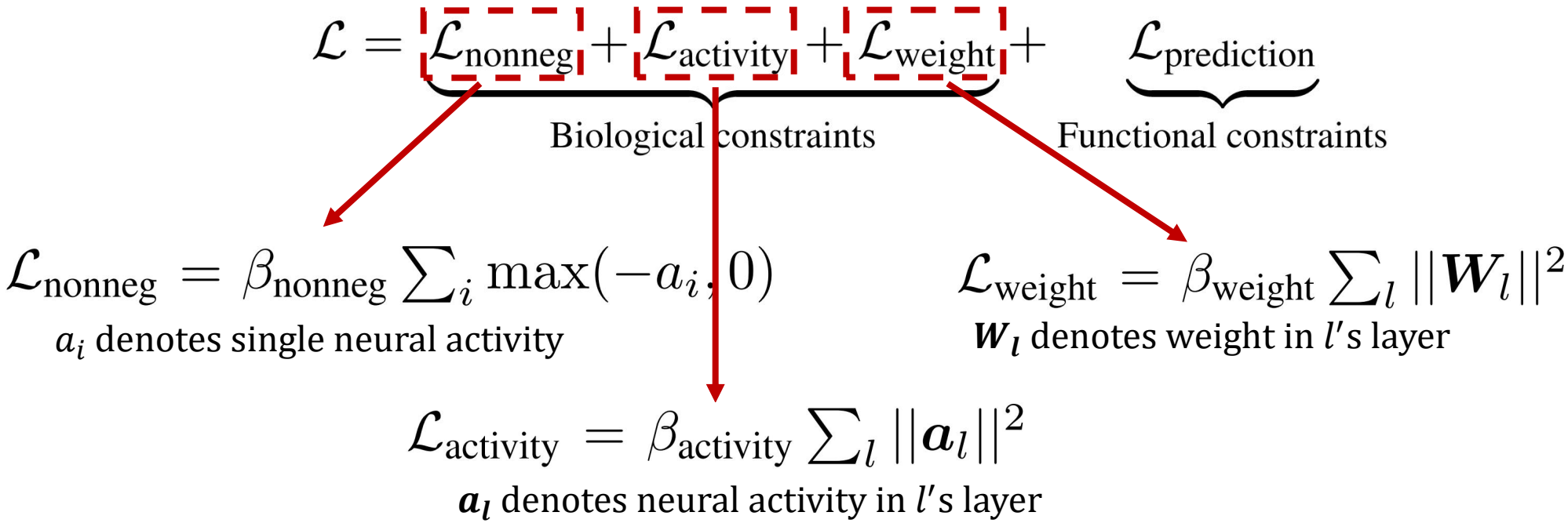
a_i denotes single neural activity

$$\mathcal{L}_{\text{activity}} = \beta_{\text{activity}} \sum_l ||\boldsymbol{a}_l||^2$$

$\boldsymbol{a_l}$ denotes neural activity in l's layer

Disentangling in machines

Regularizers as constraints



Disentangling in machines

Disentangling metric

Mutual Information Ratio (MIR)

$$r_n = \frac{\max_f(\mathrm{I}_{n,f})}{\sum_f \mathrm{I}_{n,f}}$$

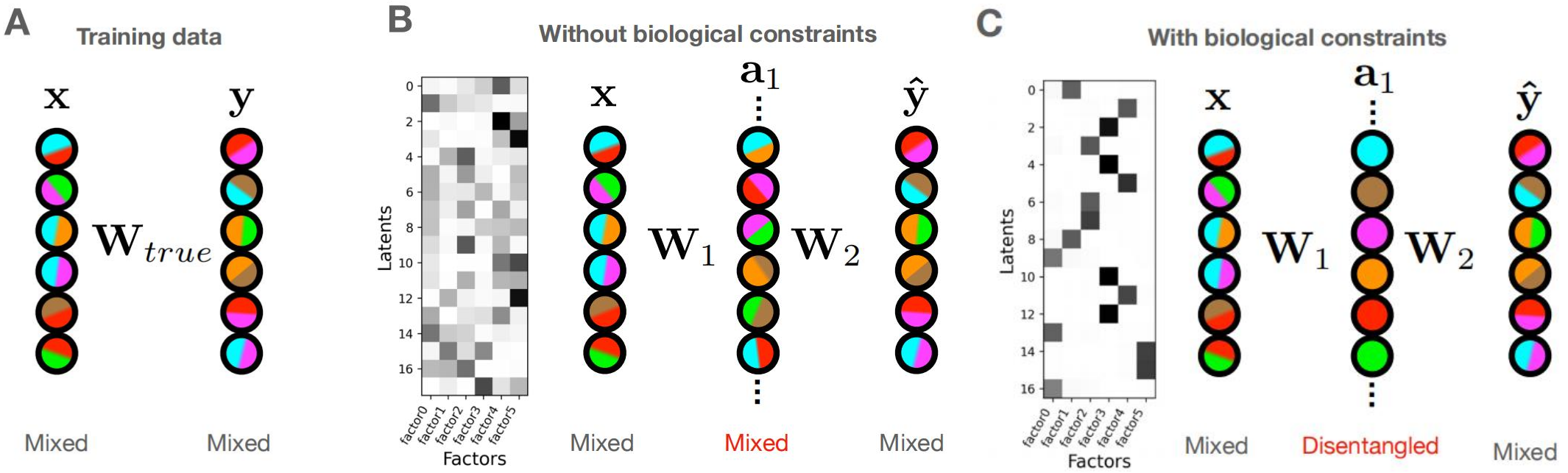
$$MIR = \frac{\frac{\sum_n r_n}{n_n} - \frac{1}{n_f}}{1 - \frac{1}{n_f}}$$

$I_{n,f}$:measures the mutual information between neurons and factors

n_n:the number of (active) neurons
n_f:the number of factors

Disentangling in machines

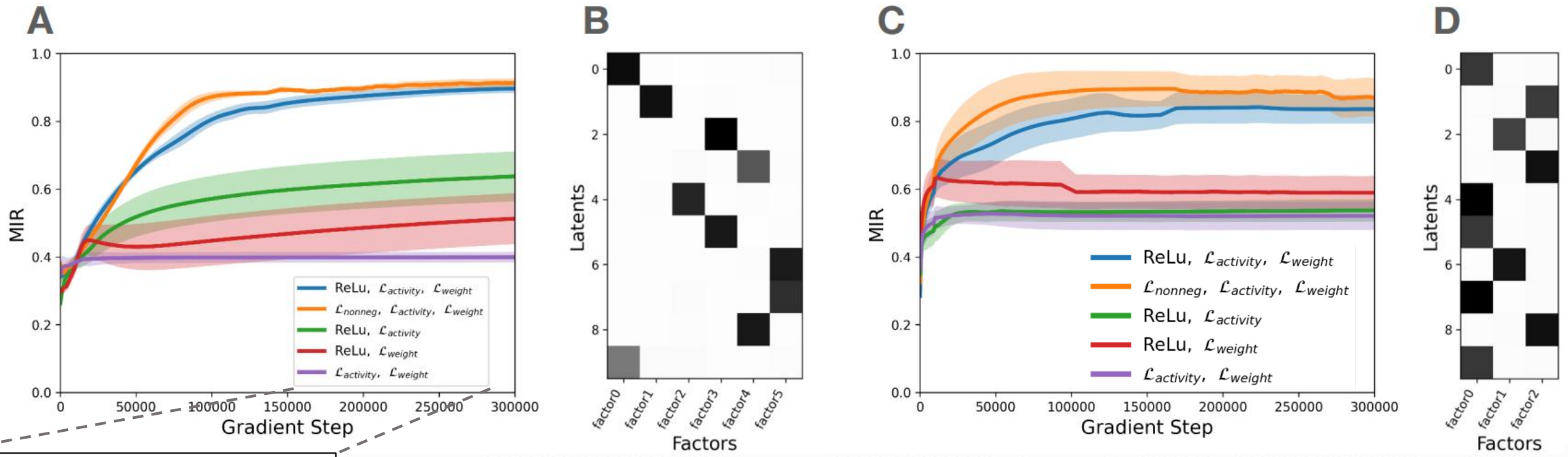
$$\mathcal{L} = \underbrace{\mathcal{L}_{\text{nonneg}} + \mathcal{L}_{\text{activity}} + \mathcal{L}_{\text{weight}}}_{\text{Biological constraints}} + \underbrace{\mathcal{L}_{\text{prediction}}}_{\text{Functional constraints}}$$

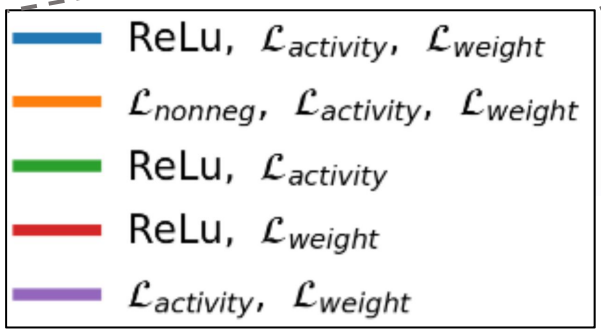


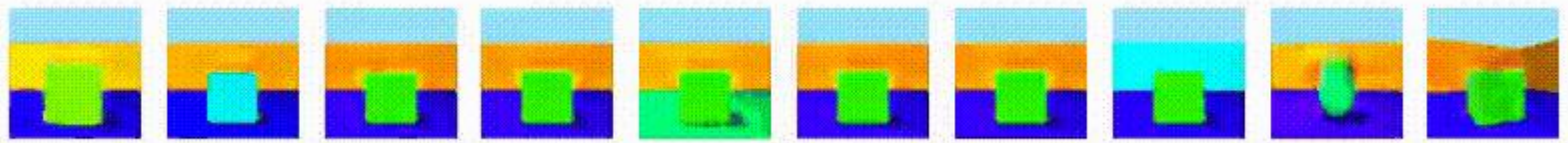
Disentangling results on shallow linear networks

Disentangling in machines

$$\mathcal{L} = \underbrace{\mathcal{L}_{\text{nonneg}} + \mathcal{L}_{\text{activity}} + \mathcal{L}_{\text{weight}}}_{\text{Biological constraints}} + \underbrace{\mathcal{L}_{\text{prediction}}}_{\text{Functional constraints}}$$

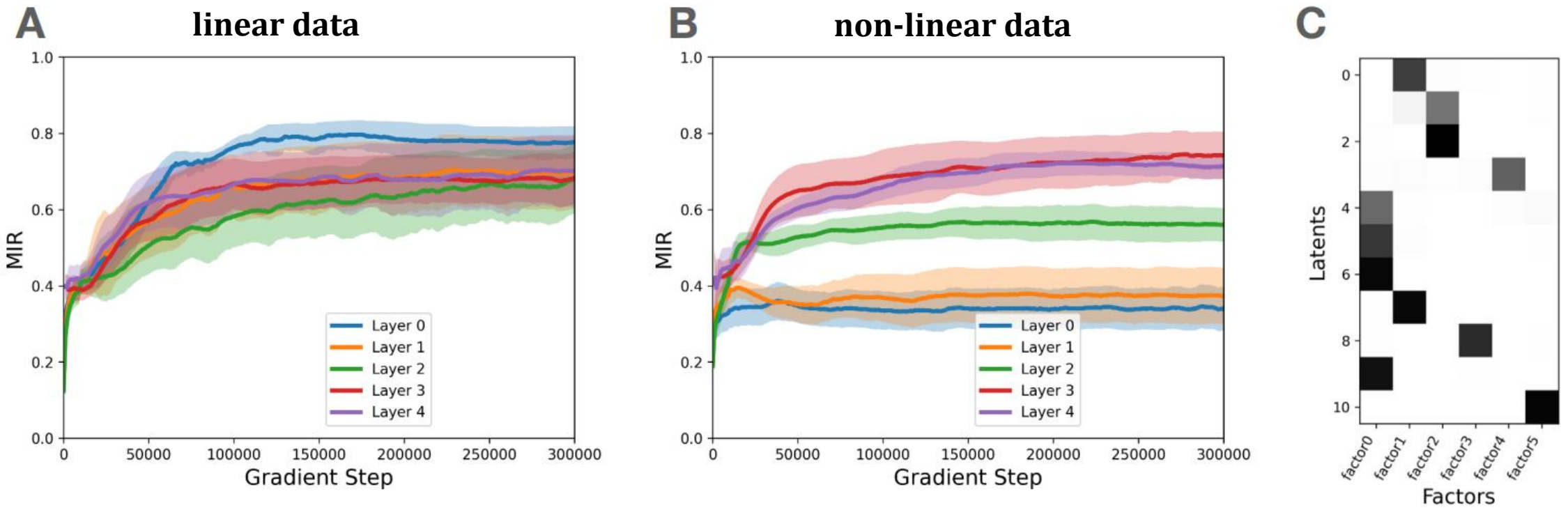






Disentangling results on autoencoders

Disentangling in machines

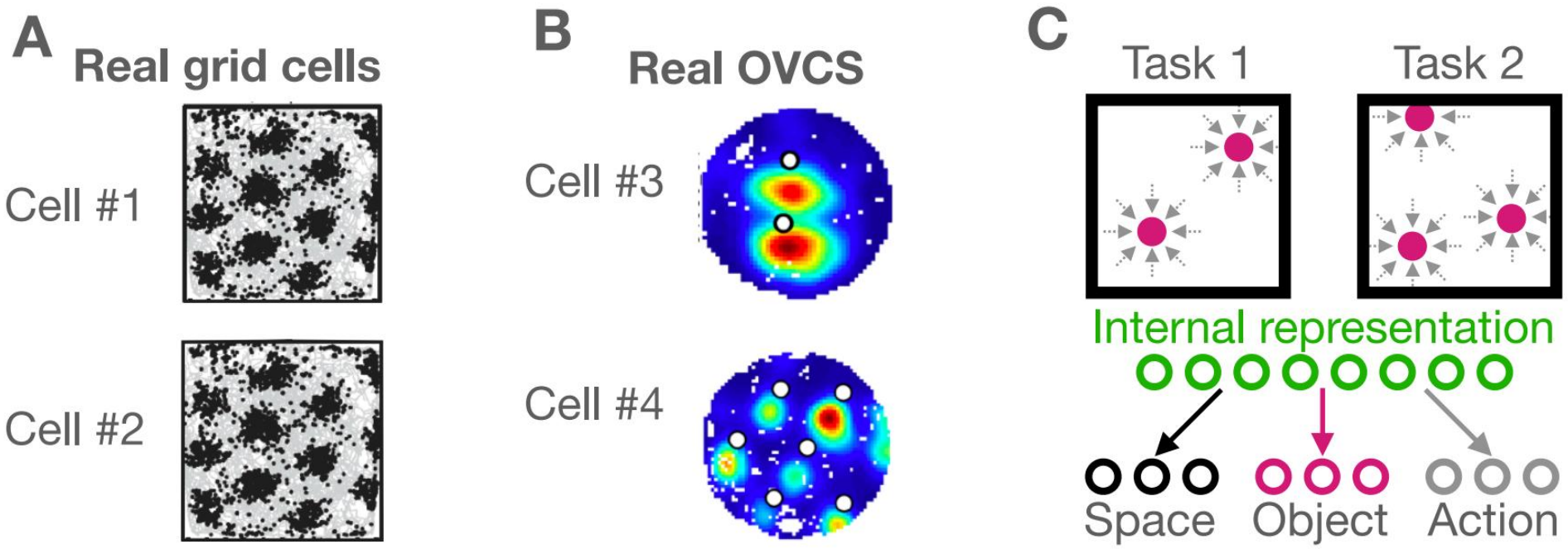


Disentangling results on deep non-linear networks

Content

- Introduction to Tim Behrens
- Background & previous puzzles
- Disentangling in machines
- Disentangling in brain
- Take-home message

Disentangling in Brain

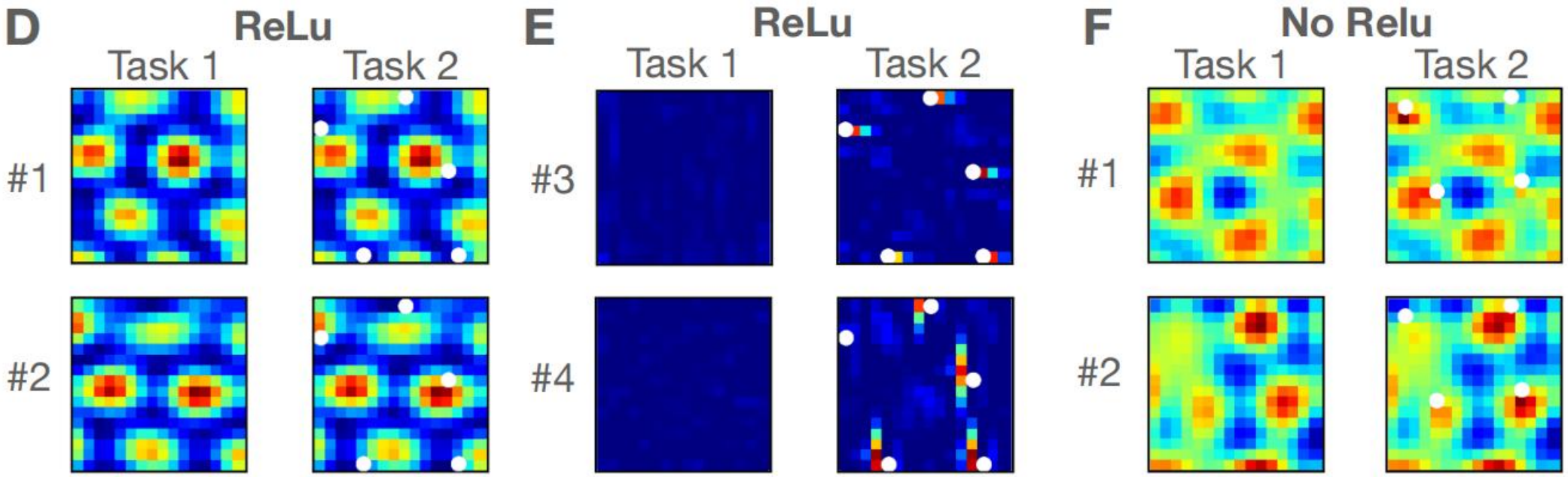


Modules of distinct cell types form with nonnegativity and factorised tasks

Tasks setting:
1. Rodents must know where they are in space
2. Rodents must also approach one of multiple objects

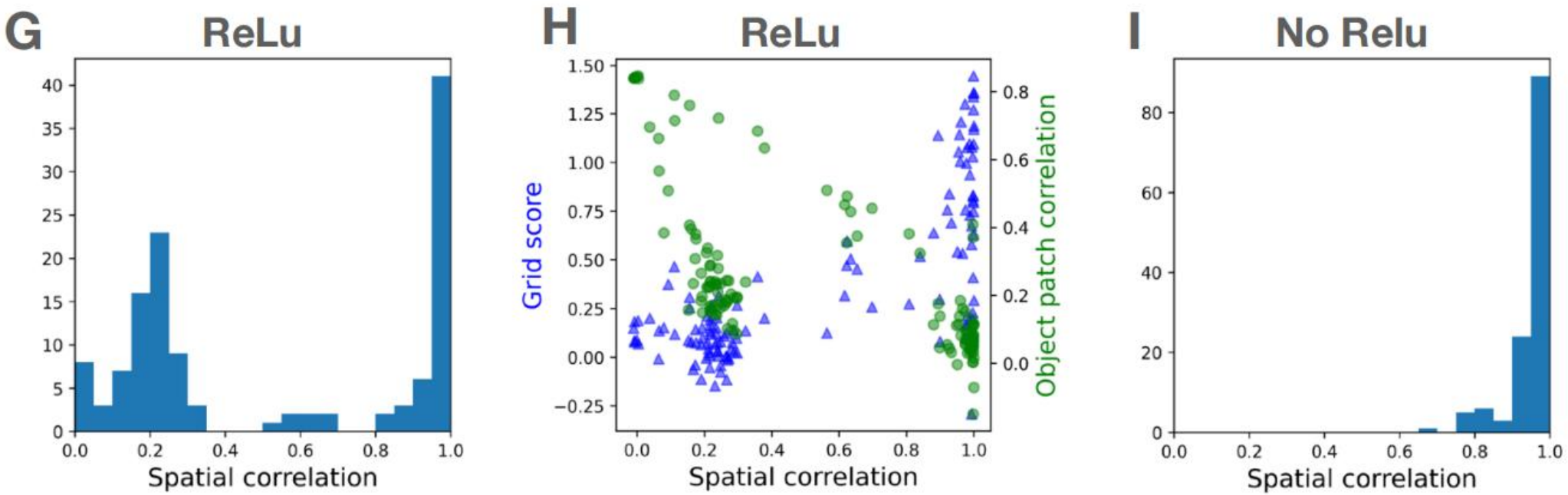
If objects appear in different places in different contexts, tasks can be factorized into:
1. Where am I in allocentric spatial coordinates?
2. Where am I in object-centric coordinates?

Disentangling in Brain



Modules of distinct cell types form with nonnegativity and factorised tasks

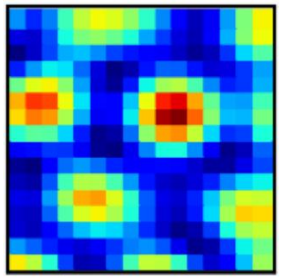
Disentangling in Brain



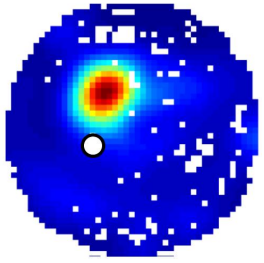
Modules of distinct cell types form with nonnegativity and factorised tasks

Answer to the questions

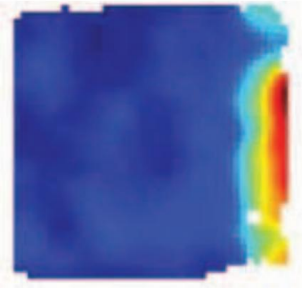
Why many different bespoke cellular responses exist for physical space?



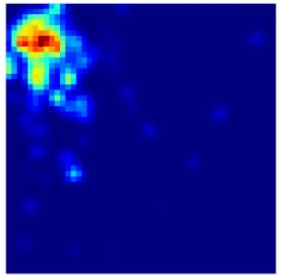
Grid cell



Object-vector cell

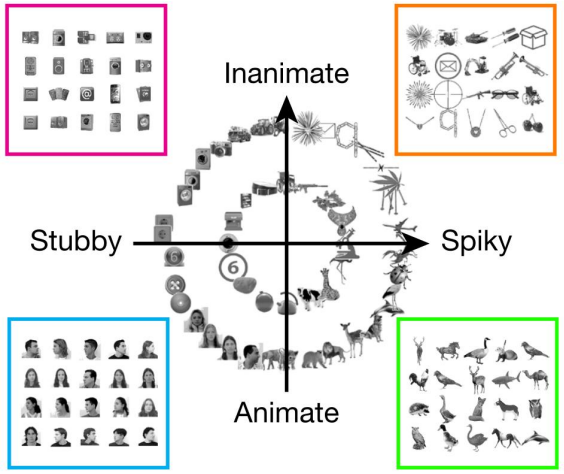


Border cell

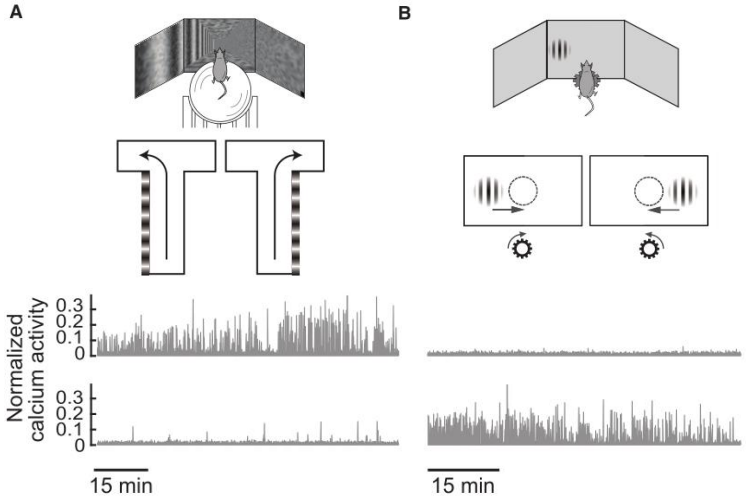


Place cell

Why many different bespoke cellular responses exist in different tasks?



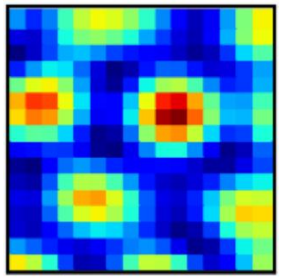
Bao et al. (2020) Nature



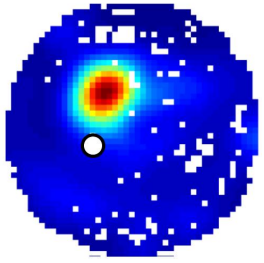
Lee et al. (2022) Neuron

Answer to the questions

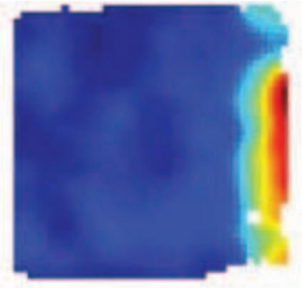
Why many different bespoke cellular responses exist for physical space?



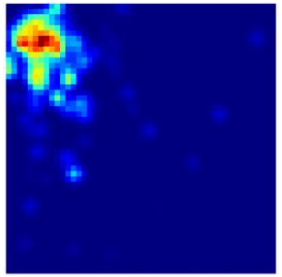
Grid cell



Object-vector cell



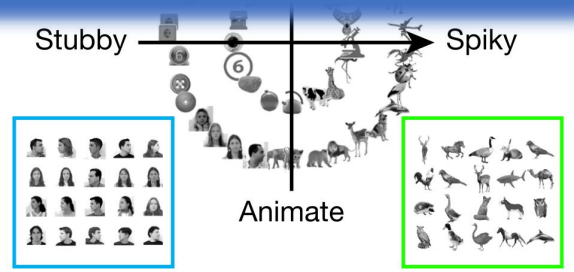
Border cell



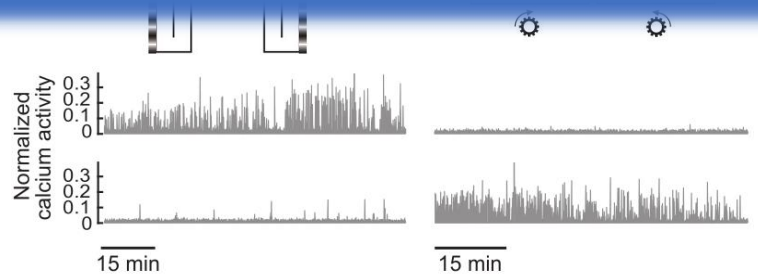
Place cell

Why many different bespoke cellular responses exist in different tasks?

Since space, boundaries, and objects appear in a factorised form, and so are optimally represented by different neural populations for each factor.



Bao et al. (2020) Nature



Lee et al. (2022) Neuron

Content

- Introduction to Tim Behrens
- Background & previous puzzles
- Disentangling in machines
- Disentangling in brain
- Take-home message

Take-home message

Scientific question:

- Why are some neural representations entangled and others not?

Technical question:

- How can we build an AI model that is able to learn disentangled representations?

Main contributions:

- **From biological aspect**: This paper shows the most efficient biological representation puts different factors in different neurons.
- **From machine aspect**: This paper builds machines that learn disentangling representations with simple biological constraints of nonnegativity and minimising activity energy.

Acknowledgement

Many thanks to Prof. Liu and all members in NCCLab

Thanks for your attention!