Back to the Continuous Attractor

Ábel Ságodi

In collaboration with
Guillermo Martín-Sánchez, Piotr Sokół, Il Memming Park
Champalimaud Centre for the Unknown

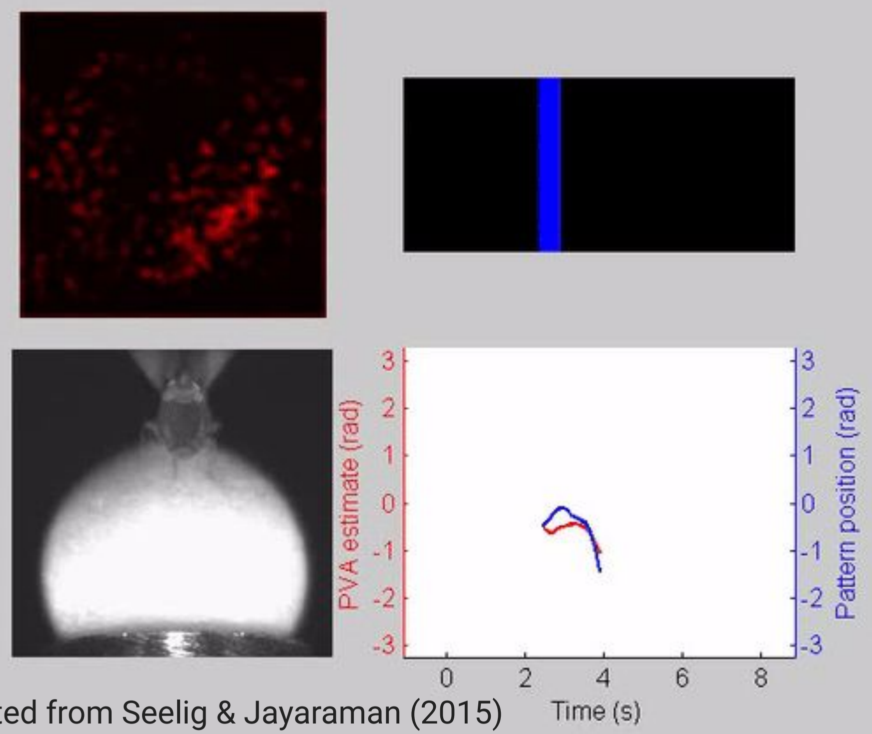
NeurIPS 2024

CATNIP LAB

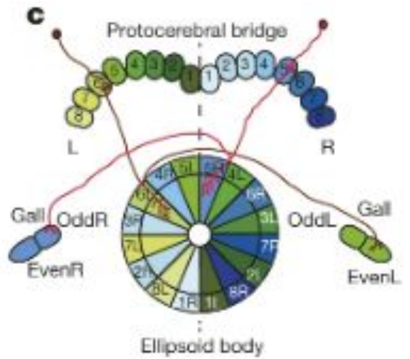
Internal compass representation in *Drosophila*

Neural activity in ellipsoid body

Decoded orientation



Adapted from Seelig & Jayaraman (2015)



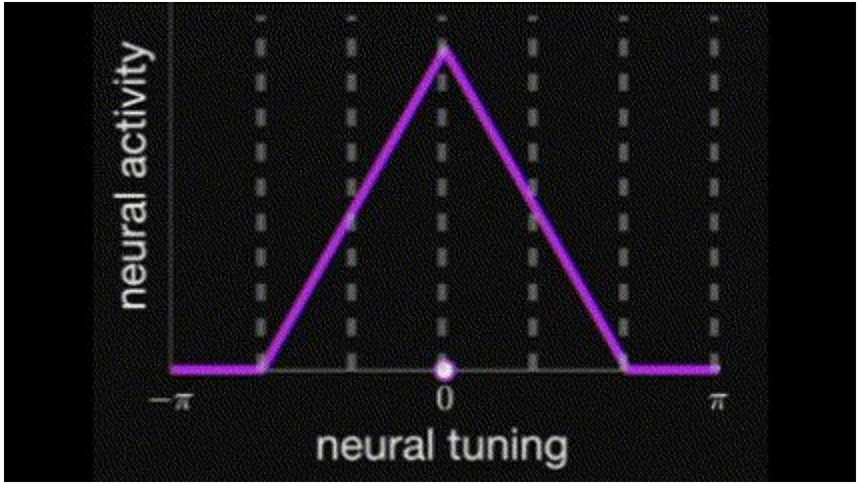
Central complex

- Stable bump of activity on a ring
- Activity bump is maintained in dark
- Angular velocity integration

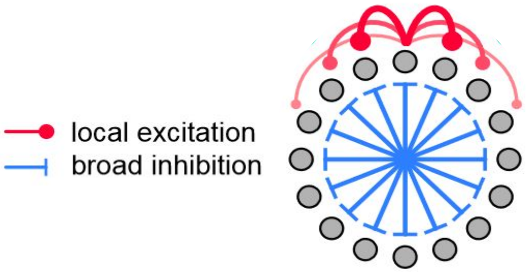
Kim *et al.* (2017)
Green *et al.* (2017)
Hulse *et al.* (2021)

Ring attractor: an internal compass model

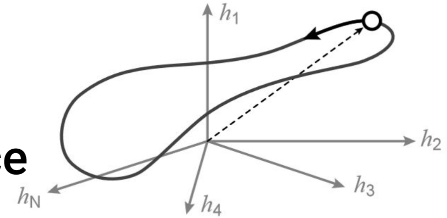
- Bump is stable in absence of input (fixed point)
- Bump shifts with angular velocity input
- All bumps together form a ring
- **Symmetric connectivity matrix**



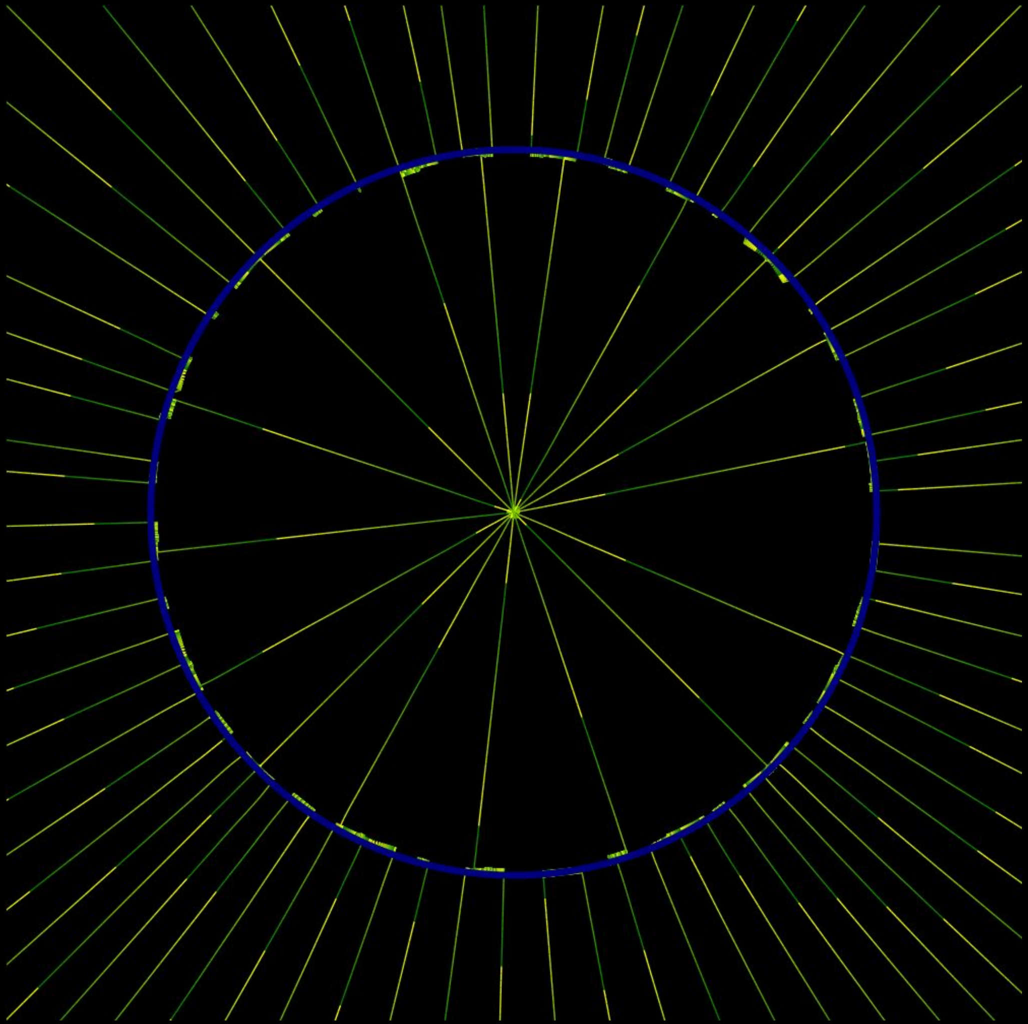
Adapted from Noorman *et al.* (2024)
N=6 neurons, ReLU



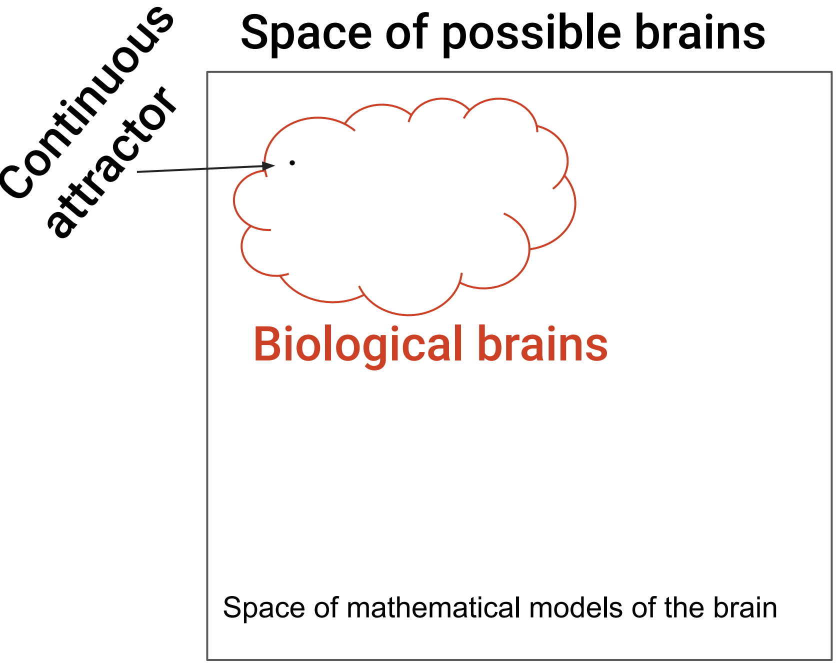
***N* dimensional neural state space**



Skaggs *et al.* (1994)
Zhang *et al.* (1996)
Ermentrout (1998)
Wang (2001)
Compte (2006)



The fine-tuning problem

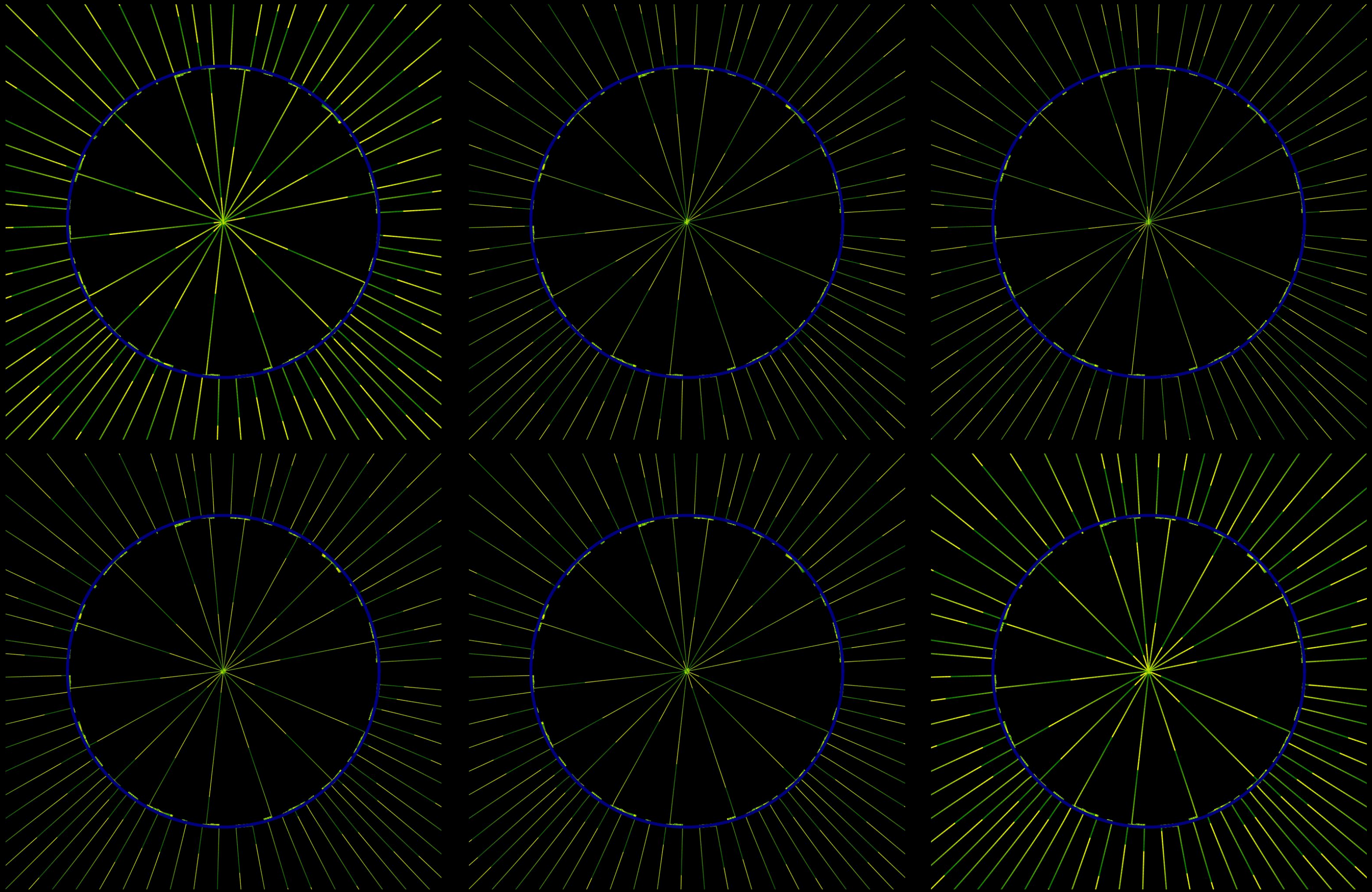


Noise is omnipresent in biological systems

- **Factors that affect brain dynamics:**
  - Temperature, neuromodulators: e.g. alcohol
  - Constantly fluctuating synaptic weights
- **Function in animals is not affected**
- **Continuous attractors are brittle**

How can our mathematical models be so out of touch with biology?

Averbeck, Latham & Pouget (2006)
Shimizu *et al*. (2021)
Fauth & Van Rossum (2019)
Park, Ságodi & Sokół (2023)



Observation 1: Approximate ring attractors retain ring-like activity

- **Ring attractors bifurcate into ring-like activity with different stability structures**
    - **Ring attractor with few neurons** Noorman *et al.* (2024)

- **Finite size approximations of ring attractors have the same bifurcation structure**
    - **Gaussian** Skaggs et al. (1994), Zhang et al. (1996)
    - **Low-rank** Mastrogiuseppe & Ostojic (2018)
    - **Embedding Manifolds with Population-level Jacobians** Pollock & Jazayeri (2020)

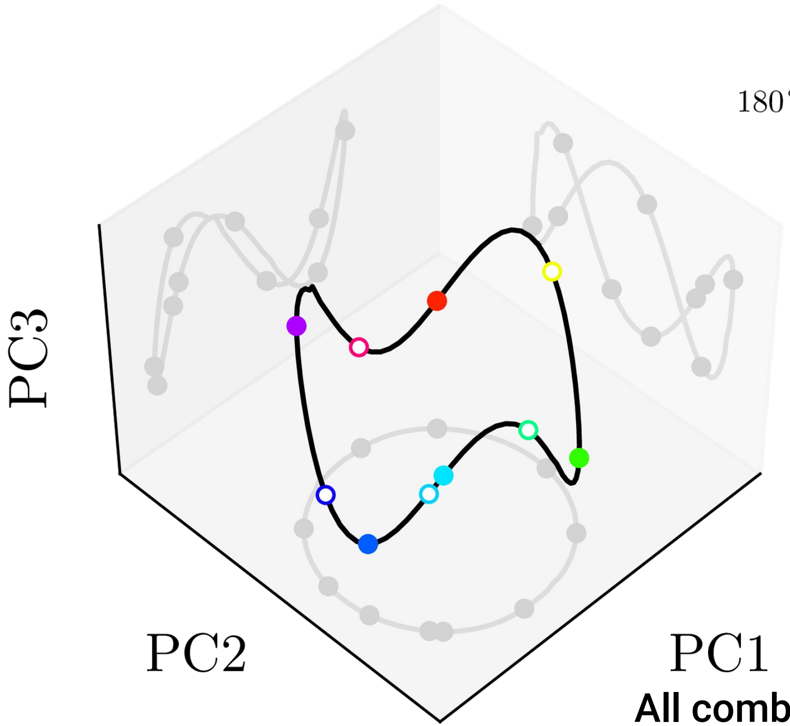
Observation 2

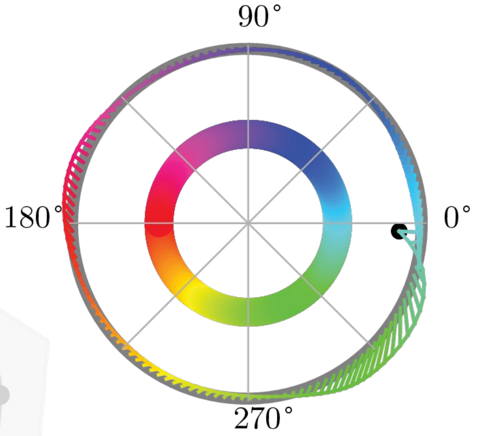
Trained RNNs on the angular velocity integration task

— attractive ring

○ saddle

● stable fixed point





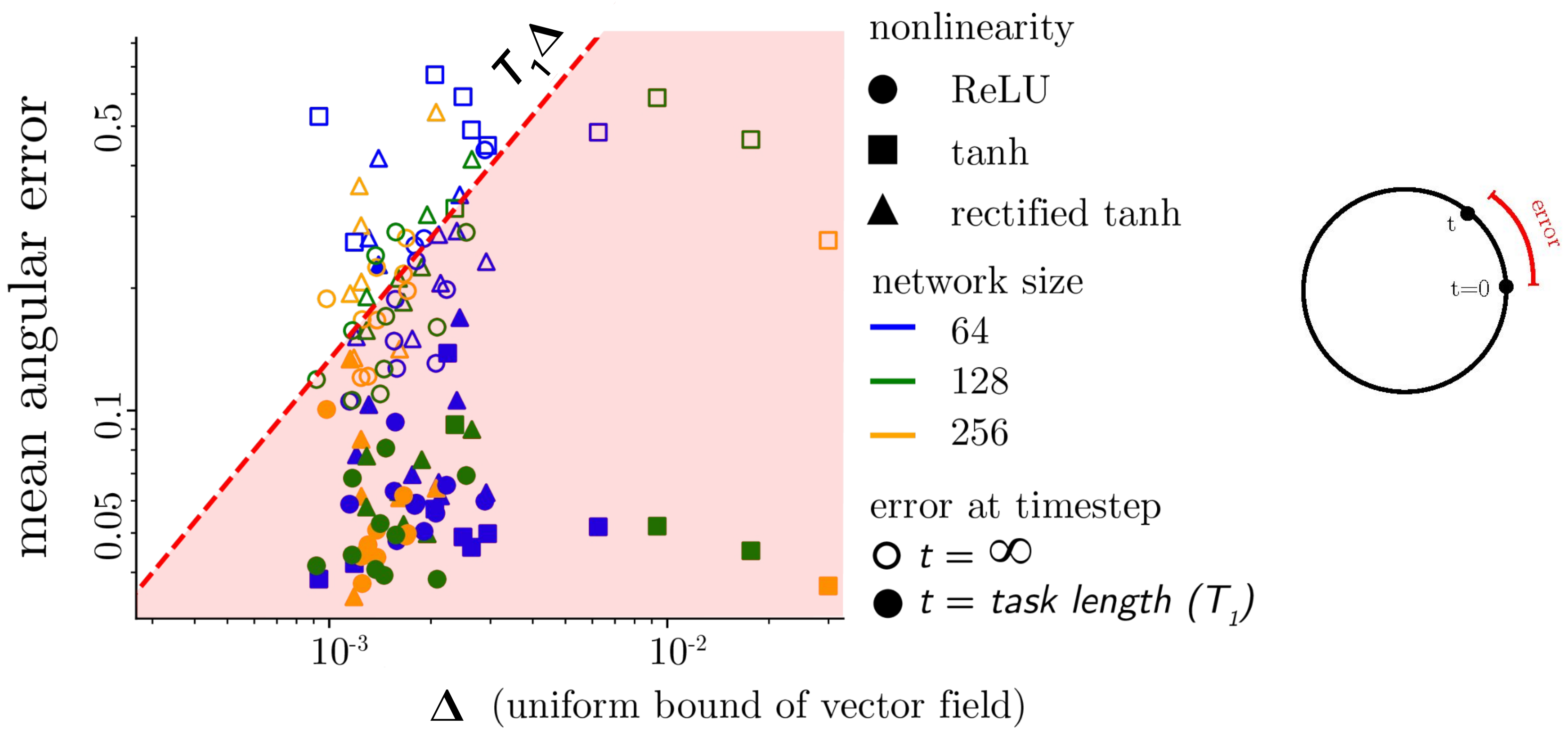
All networks have:

- **Different stability structures (different number of fixed points)**
- **The neural activity is attractive onto a ring**

All combinations:

- **ReLU, tanh, rectified tanh**
- **Sizes: 64, 128, 256**

Uniform norm of vector field bounds error



Approximate continuous attractor theory

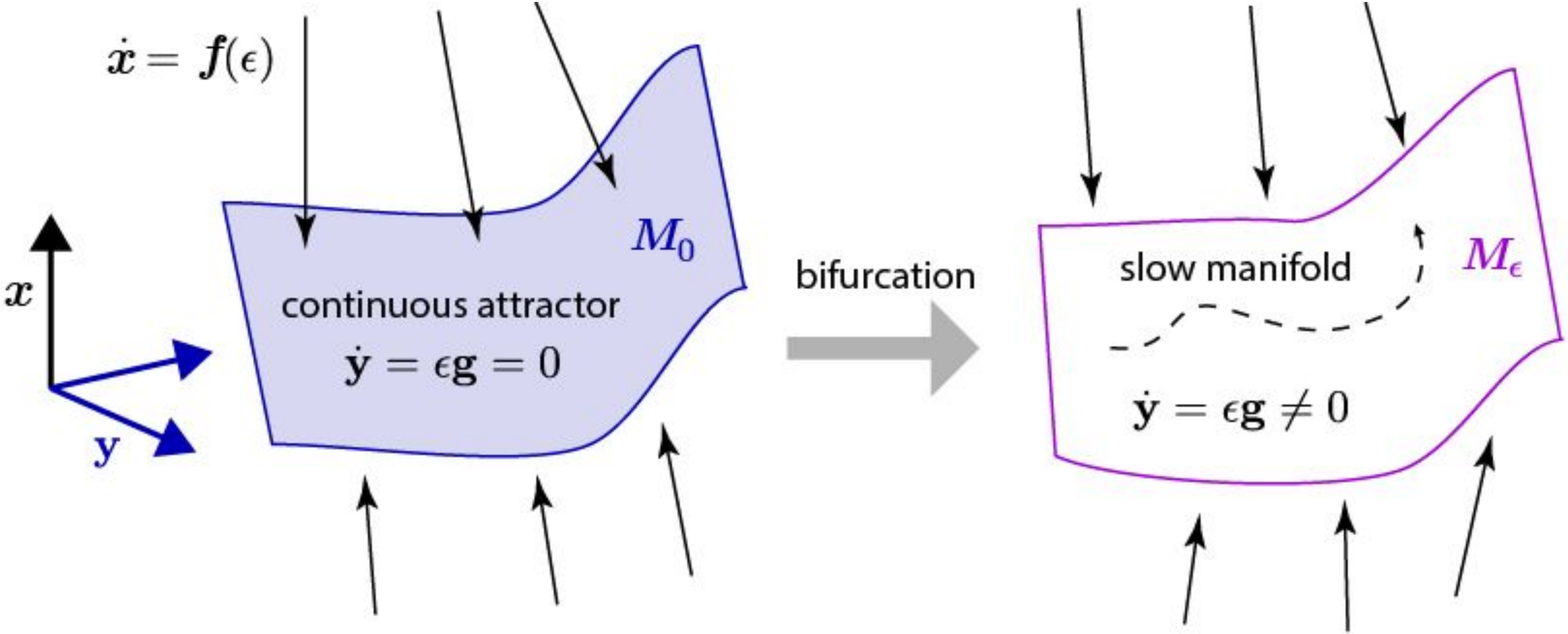
Why are all these systems so similar?

1. Continuous attractors **persist**

2. Behavioral similarity ⇔ Configuration similarity

1. Persistence of continuous attractors

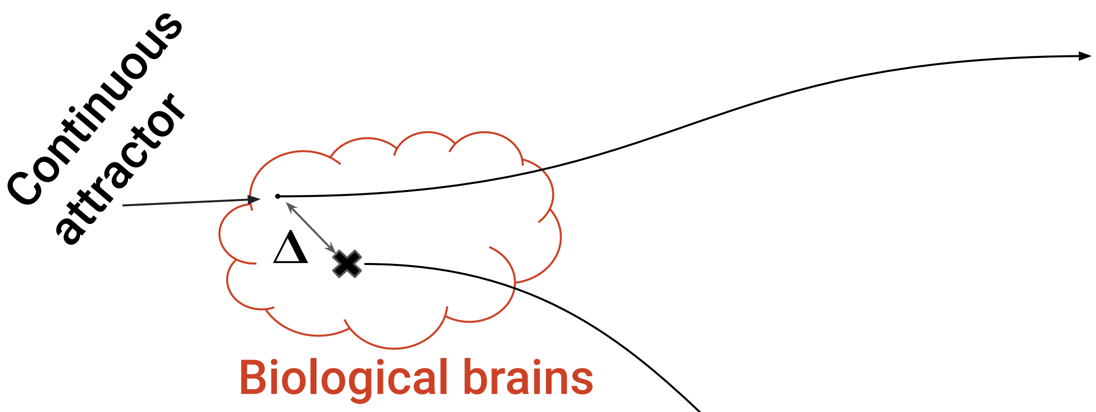
Fenichel's Invariant Manifold Theorem

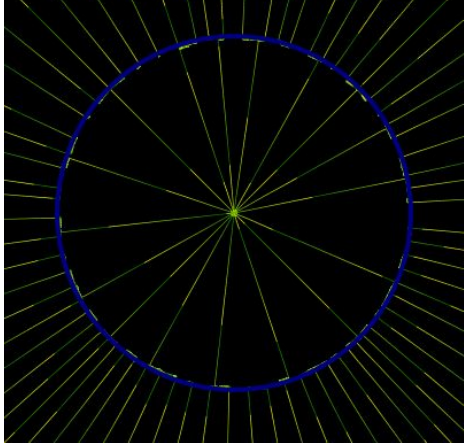


Explains why the perturbations to the ring attractor resulted in such similar dynamics

Fenichel (1971)
Mañé (1978)
Jones (1995)
Simpson (2018)

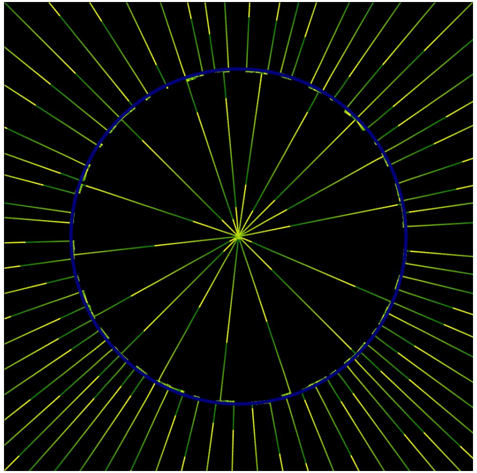
2. Systems close to a ring attractor have small behavioral error





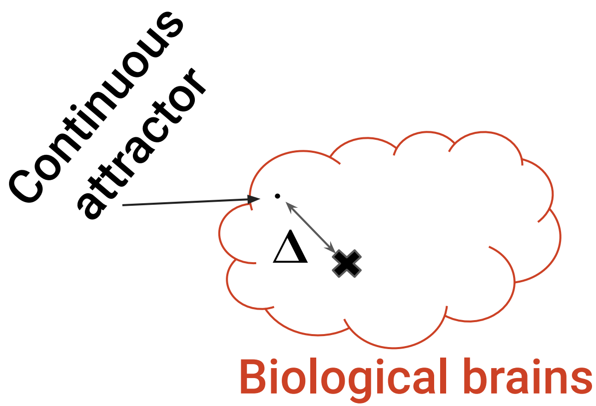
Ring attractor

$$\Delta \coloneqq |vf_{ca} - vf_{pert}|$$



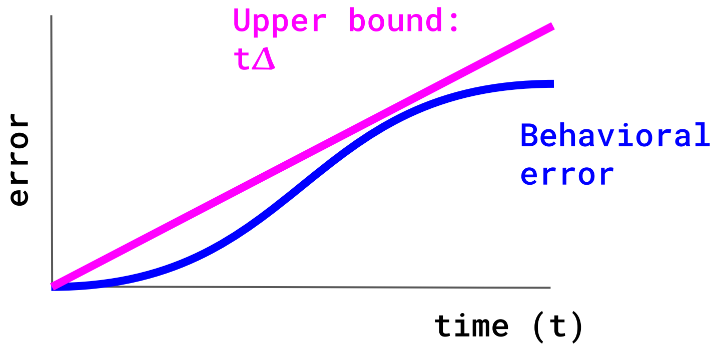
Approximate ring attractor

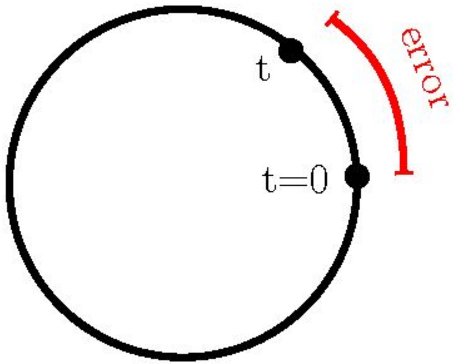
2. Systems close to a ring attractor have small behavioral error



Explains why all task trained RNNs have attractive dynamics onto a ring

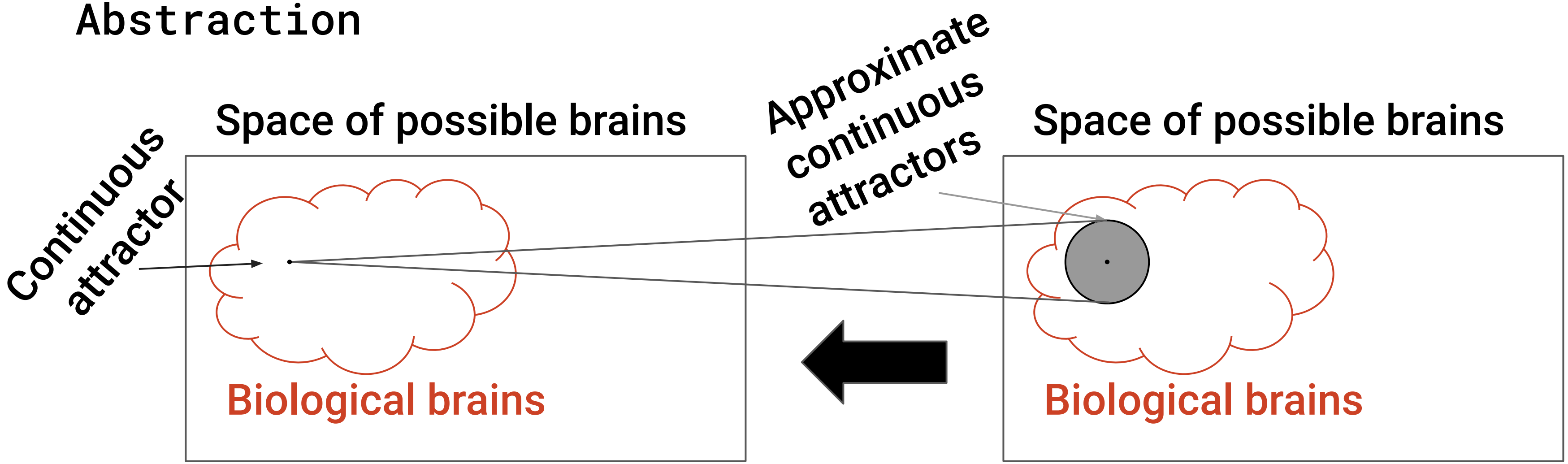
$$\Delta \coloneqq |vf_{ca} - vf_{pert}|$$





Back to the continuous attractor

Abstraction



Abstract out the details of approximations of continuous attractors

- Zebrafish
  - heading direction
  - self-location
- Aggression in mice

Potochnik (2018)
Chirimuuta (2024)
Nair *et al.* (2022)
Petrucco *et al.* (2023)
Yang *et al.* (2022)

Guillermo Martín-Sánchez

Piotr Sokół

Il Memming Park

Camera-ready: https://arxiv.org/abs/2408.00109

NeurIPS 2024 Poster
Thursday December 12, 11:00-14:00