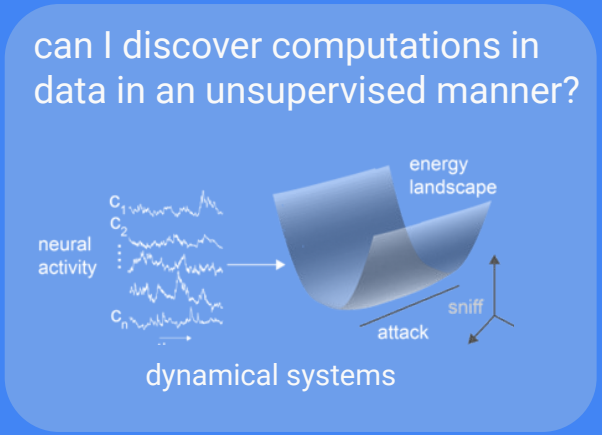
Caltech

datasai 2023 vol. 2



dynamical systems & neural population dynamics

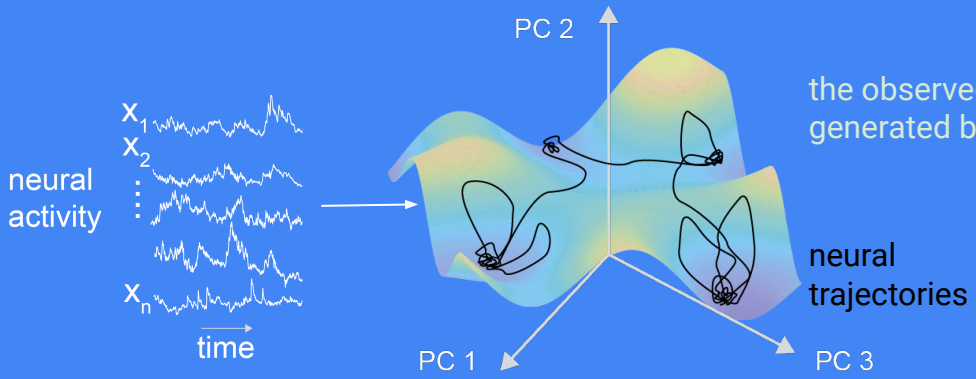
aditya nair

CHEN INSTITUTE TIANQIAO & CHRISSY

recap:

strong principle :

dimensionality reduction is a hypothesis for how neural circuits compute



the observed low dimensional activity is generated by an underlying manifold

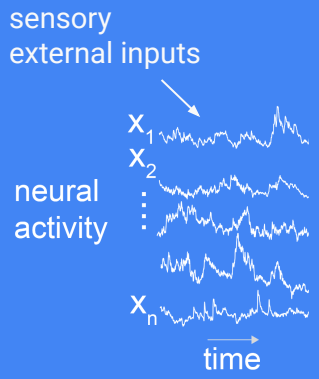
dynamical systems is the *language* that allows us to understand the underlying manifold and discover computations

Caltech

datasai_2023

the anatomy of a dynamical system

a dynamical system is a mathematical model that describes how a system changes over time

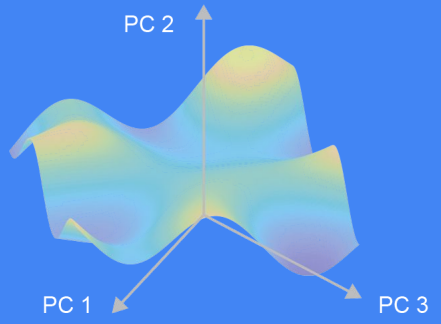


$$\frac{dx}{dt} = f(x(t), u(t))$$

non-linearity

$x(t)$: firing rate of N neurons

$u(t)$: external inputs to a circuit



a recurrent neural network is a non-linear dynamical system

the non-linearity makes interpretation of the dynamical systems difficult and can rise to complex dynamics

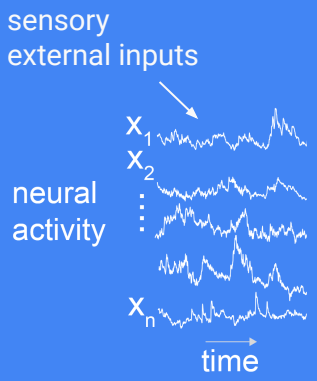
Caltech

datasai_2023

the anatomy of a dynamical system

a dynamical system is a mathematical model that describes how a system changes over time

to build some intuition, let's consider linear systems



$$\frac{dx}{dt} = Ax(t) + Bu(t)$$

$Ax(t)$: firing rate of N neurons

$Bu(t)$: external inputs to a circuit

in practice, even RNNs are understood through linearization

understanding A is crucial for understanding the computations of the dynamical system

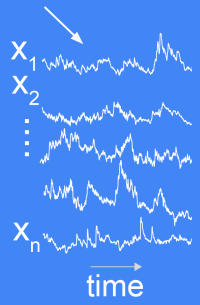
Caltech

datasai_2023

the anatomy of a dynamical system

sensory
external inputs

neural
activity

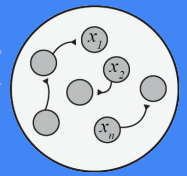


$$\frac{dx}{dt} = \underbrace{Ax(t)}_{(n \times t)} + \underbrace{Bu(t)}_{(m \times t)}$$

firing rate of N neurons

external inputs to a circuit

external input



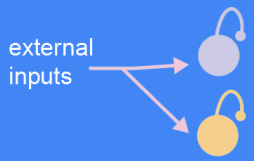
what is the meaning of A? - connectivity matrix $(n \times n)$

an eigendecomposition of A can tell us a lot about the computations performed by this circuit

Caltech
datasai_2023

the anatomy of a dynamical system $\frac{dx}{dt} = Ax(t) + Bu(t)$

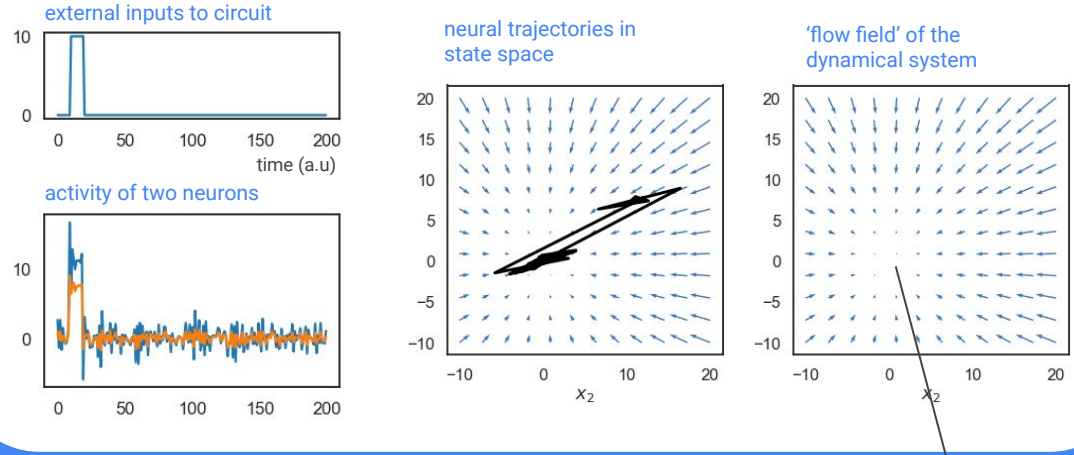
consider a simple two neuron system



let A =

$$\begin{bmatrix} -0.2 & 0 \\ 0 & -0.5 \end{bmatrix}$$

point attractors in the brain
Inagaki et al., 2019 (ALM)

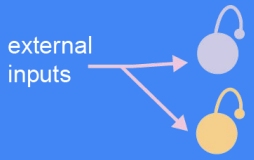


the flow field tell us how activity evolves in the absence of inputs and can reveal computations

point attractor or
stable fixed point

the anatomy of a dynamical system $\frac{dx}{dt} = Ax(t) + Bu(t)$

consider a simple two neuron system



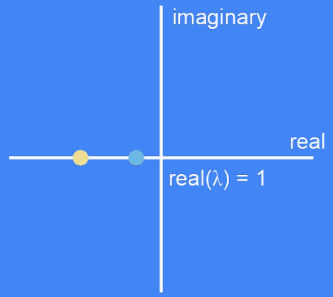
let A =

$$\begin{bmatrix} -0.2 & 0 \\ 0 & -0.5 \end{bmatrix}$$

the presence and nature of fixed points can be inferred from the eigenvalues of A

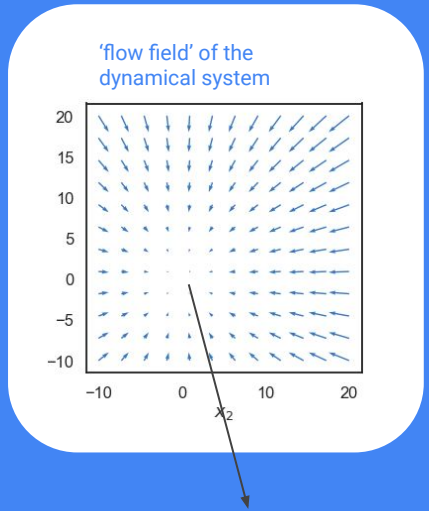
λ_1= -0.2, λ_2= -0.5

the eigenspectrum of A



a dynamical where all eigenvalues have real parts less than one are stable and contain stable fixed points

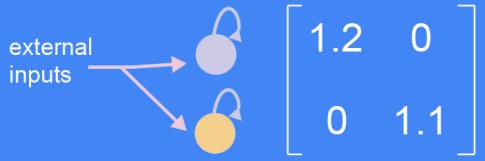
‘stability analysis’



point attractor or stable fixed point

the anatomy of a dynamical system $\frac{dx}{dt} = Ax(t) + Bu(t)$

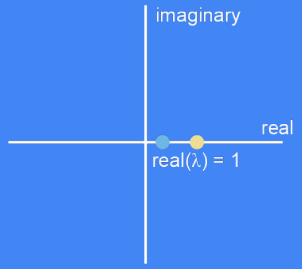
consider a simple two neuron system

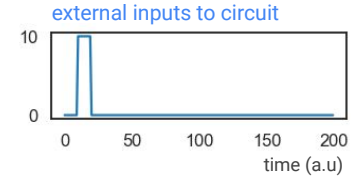


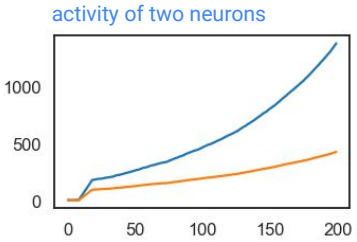
let A =

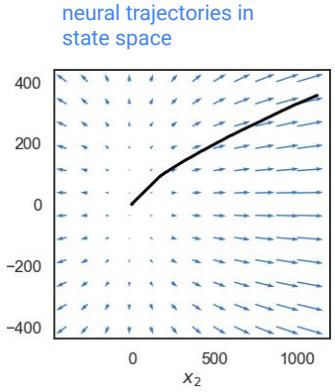
$$\begin{bmatrix} 1.2 & 0 \\ 0 & 1.1 \end{bmatrix}$$

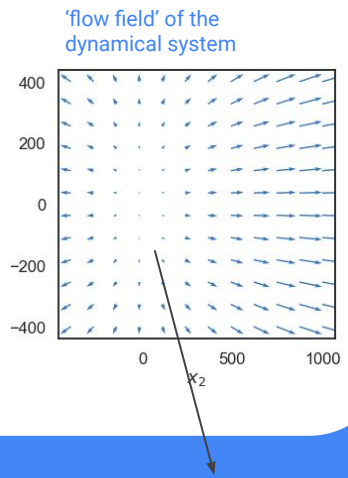
the eigenspectrum of A









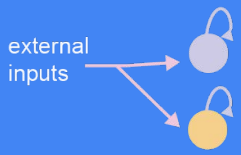


unstable fixed point

a dynamical system where all eigenvalues have real parts greater than one are unstable

the anatomy of a dynamical system $\frac{dx}{dt} = Ax(t) + Bu(t)$

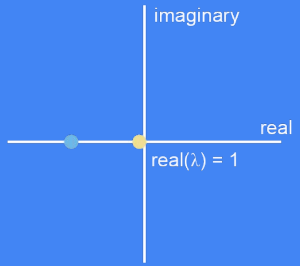
consider a simple two neuron system

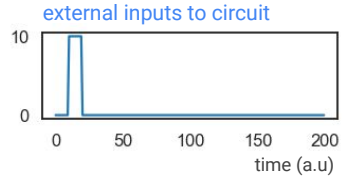


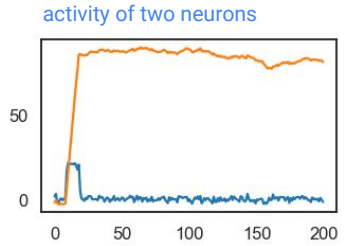
let A =

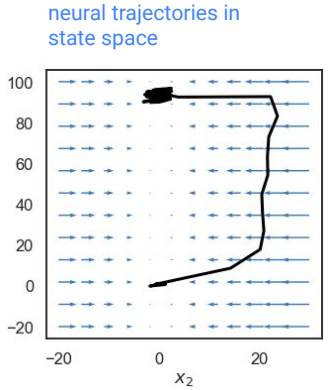
$$\begin{bmatrix} 1.01 & 0 \\ 0 & -0.2 \end{bmatrix}$$

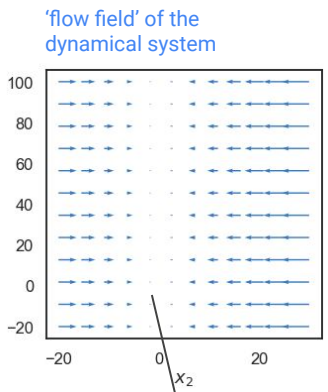
the eigenspectrum of A











a dynamical system where one eigenvalue is much larger than others and close to 1 possesses a line attractor

line attractors

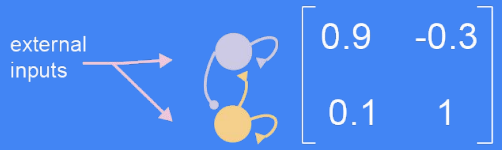
line attractors in the brain
Mante, Sussilo et al., 2019 (cortex), Nair et al., 2023 (hypothalamus)

the anatomy of a dynamical system $\frac{dx}{dt} = Ax(t) + Bu(t)$

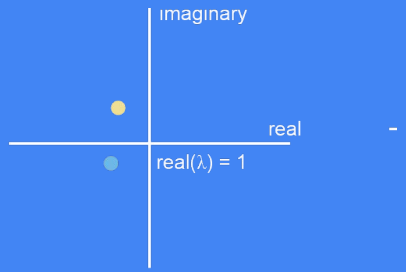
consider a simple two neuron system

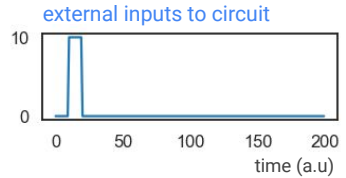
what happens when we connect the two neurons?

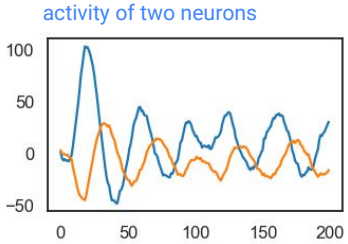
let A =

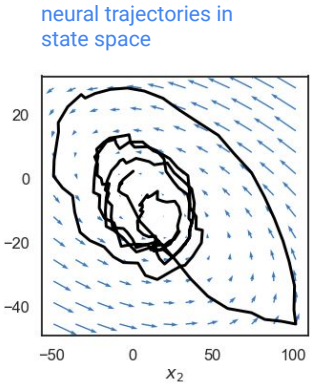


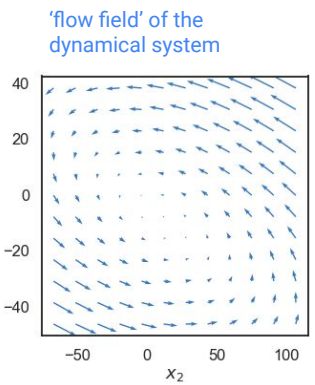
the eigenspectrum of A











dynamical systems with imaginary parts of eigenvalues will show oscillations

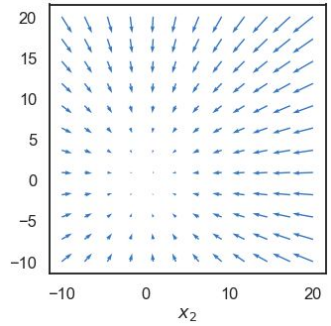
the anatomy of a dynamical system $\frac{dx}{dt} = Ax(t) + Bu(t)$

the eigenspectrum of A can reveal two important features of dynamical systems

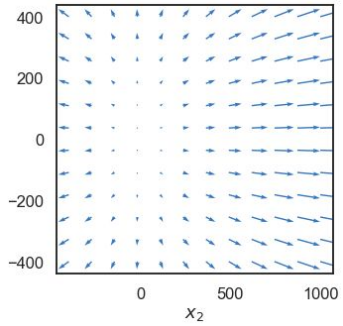
the real part determines whether activity will expand or decay

the imaginary part determines the frequency of oscillations

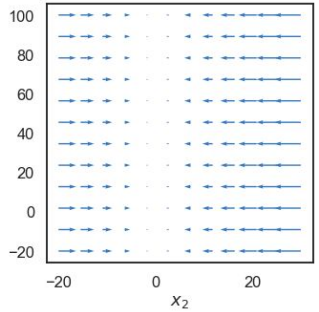
'flow field' with stable fixed pt.



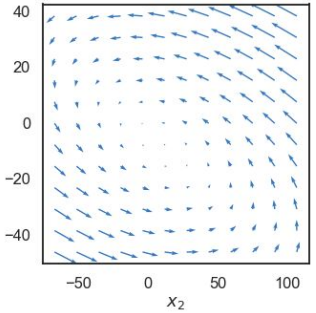
'flow field' with unstable fixed pt.



'flow field' with line attractor



'flow field' with oscillations



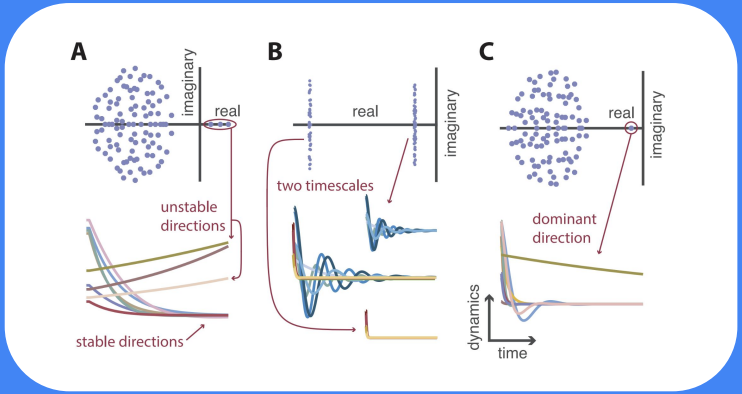
the anatomy of a dynamical system $\frac{dx}{dt} = Ax(t) + Bu(t)$

the eigenspectrum of A can reveal two important features of dynamical systems

the real part determines whether activity will expand or decay

the imaginary part determines the frequency of oscillations

these concepts readily generalize to high dimensional data



when in doubt, examine your eigenspectra

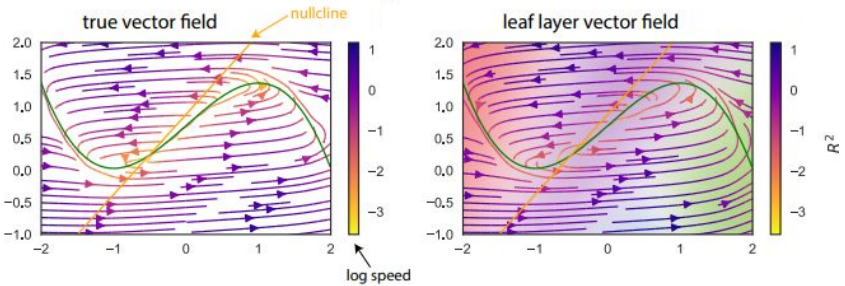
Christodoulou & Vogel (2022), the eigenvalue in neuroscience

Caltech
datasai_2023

are linear dynamical systems even useful?

true flow field from a non-linear oscillator

linear approximation using trSLDS



non-linear systems can often be approximated as a series of interacting linear systems

even RNNs are dissected by ‘linearization’, i.e by assuming those some portions of state space can be modelled by linear dynamical systems

motifs like point attractors, line attractors, unstable fixed points etc. are the building blocks of complex non-linear systems

from Nassar, J., Linderman, S. W., Bugallo, M., & Park, I. M. (2018). Tree-structured recurrent switching linear dynamical systems for multi-scale modeling.

why should we care about dynamical systems?

earliest work comes from modelling single neurons as dynamical systems

Summary of equations and parameters

We may first collect the equations which give the total membrane current I as a function of time and voltage. These are:

$$I = C_M \frac{dV}{dt} + \bar{g}_K n^4 (V - V_K) + \bar{g}_{Na} m^3 h (V - V_{Na}) + \bar{g}_l (V - V_l), \quad (26)$$

where

$$dn/dt = \alpha_n (1 - n) - \beta_n n, \quad (7)$$

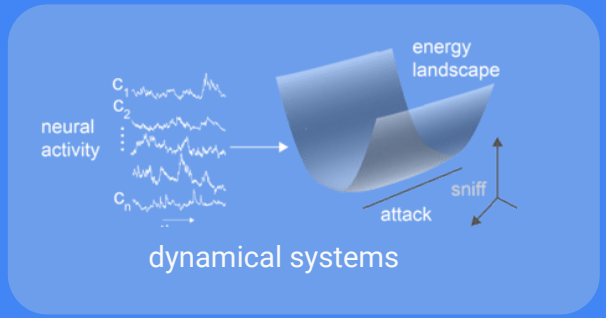
$$dm/dt = \alpha_m (1 - m) - \beta_m m, \quad (15)$$

$$dh/dt = \alpha_h (1 - h) - \beta_h h, \quad (16)$$

hodgkin-huxley equation

dynamical systems in neuroscience by Eugene Izhikevich

modelling neural population dynamics as a dynamical system



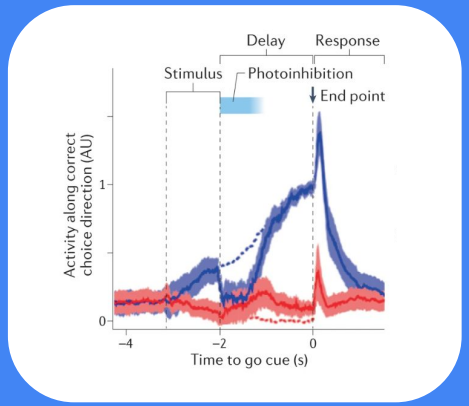
computation through neural population dynamics by Krishna Shenoy, David Sussillo et al.,

Caltech
datasai_2023

why should we care about dynamical systems?

emerging evidence for attractor dynamics in the brain

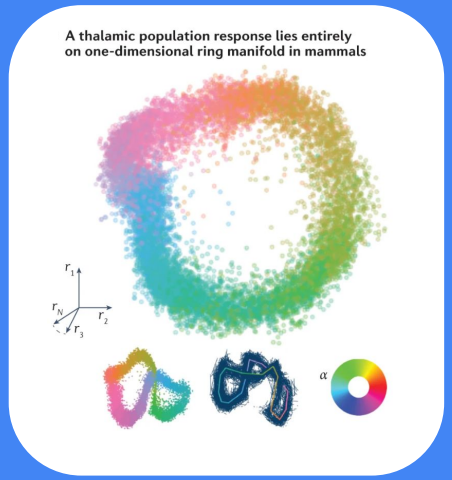
point attractors



discrete point attractors for encoding working memory, Inagaki et al., 2019

discrete multistability in the hippocampus, Josselyn & Tonegawa., 2020

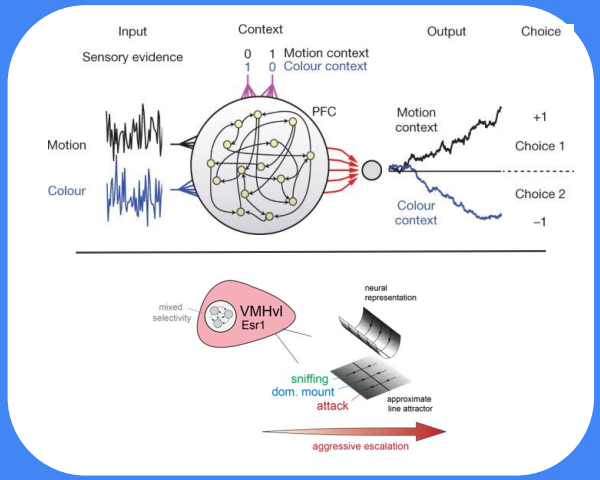
ring attractors



ring attractors for head direction in flies, Kim et al., 2017

ring attractors for head direction in the mammalian thalamus, Chaudhuri et al., 2019

line attractors



line attractors for context dependent integration Mante, Sussillo et al., 2013

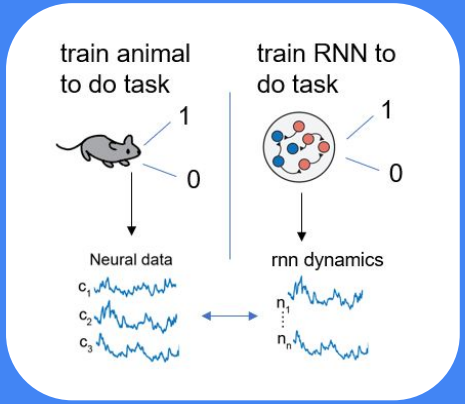
line attractors for encoding aggressive states, Nair et al., 2023

many more, see Khona & Fiete, 2022 for review

how do we find dynamical motifs from data?

two broad approaches

task based modelling

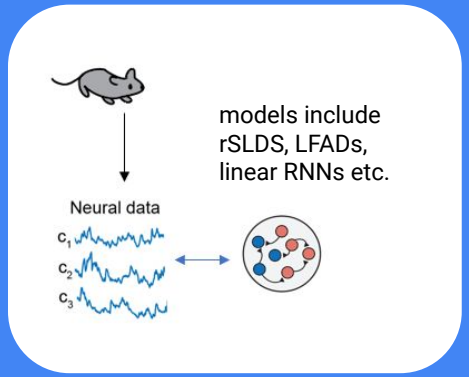


line attractors for context dependent integration
Mante, Sussillo et al., 2013

A new theoretical framework jointly explains behavioral and neural variability across subjects performing flexible decision-making
Pagan,.., Pillow, Mante, Sussillo, Brody et al., 2022

check out Jonathon Kao's lectures from datasai22 to implement Mante, Sussillo yourself!

neural population modelling



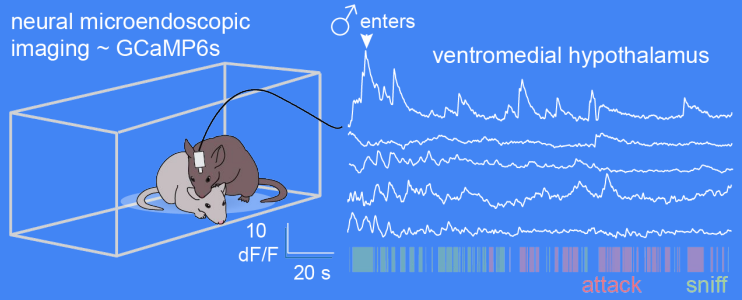
line attractors for encoding aggressive states,
Nair et al., 2023

line attractors for encoding past reward,
Sylwestrak, ..., Sussillo, Deisseroth et al., 2022

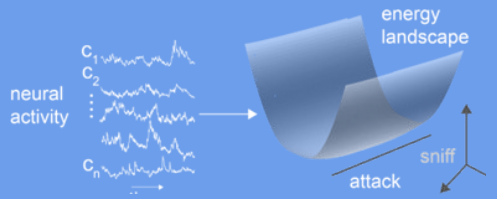
attractor dynamics during delay periods,
Rajan, Harvey, Tank, 2016

Caltech
datasai_2023

an example of neural population modelling as dynamical systems



can I discover computations in data in an unsupervised manner?

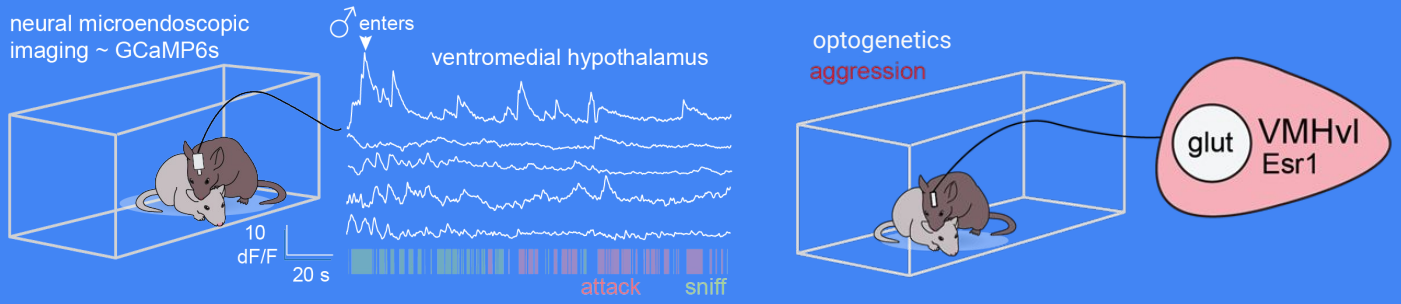


dynamical systems

Caltech

datasai_2023

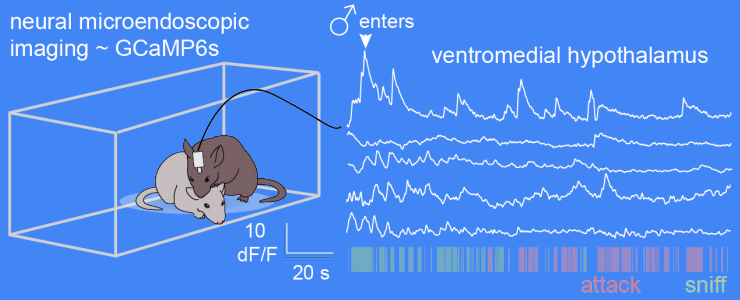
a paradox in the ventromedial hypothalamus

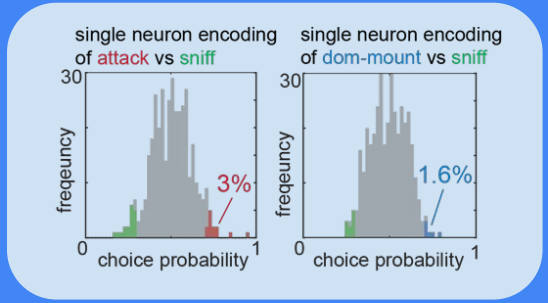


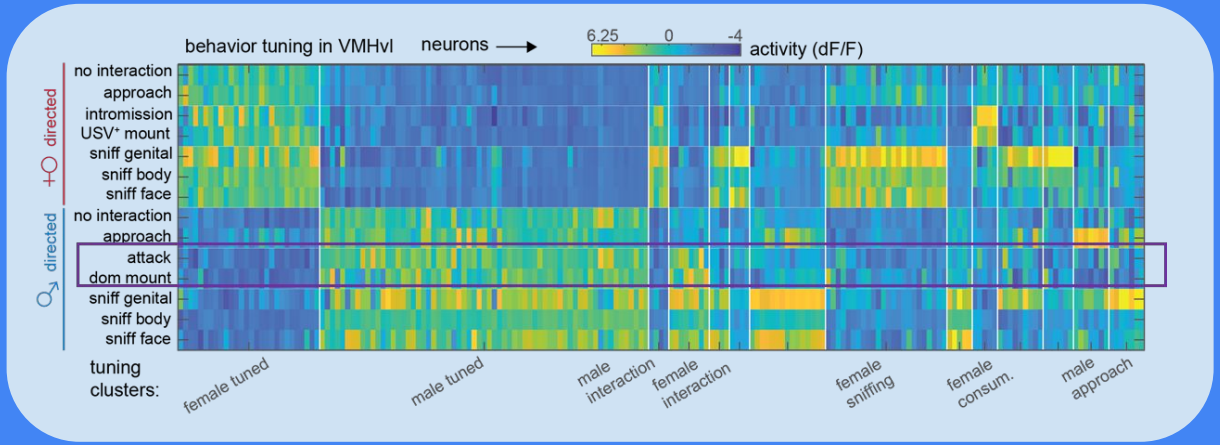
the principle components of VMHvl activity are dominated by intruder sex, not attack behavior

does this structure have anything to do with aggression?

a paradox in the ventromedial hypothalamus





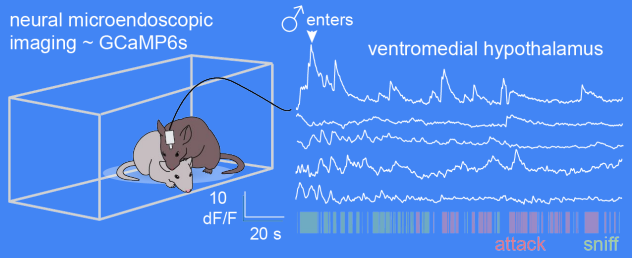


single neuron tuning in VMHvl is also dominated by intruder sex & shows mixed selectivity

is aggression encoded in the population dynamics of activity?

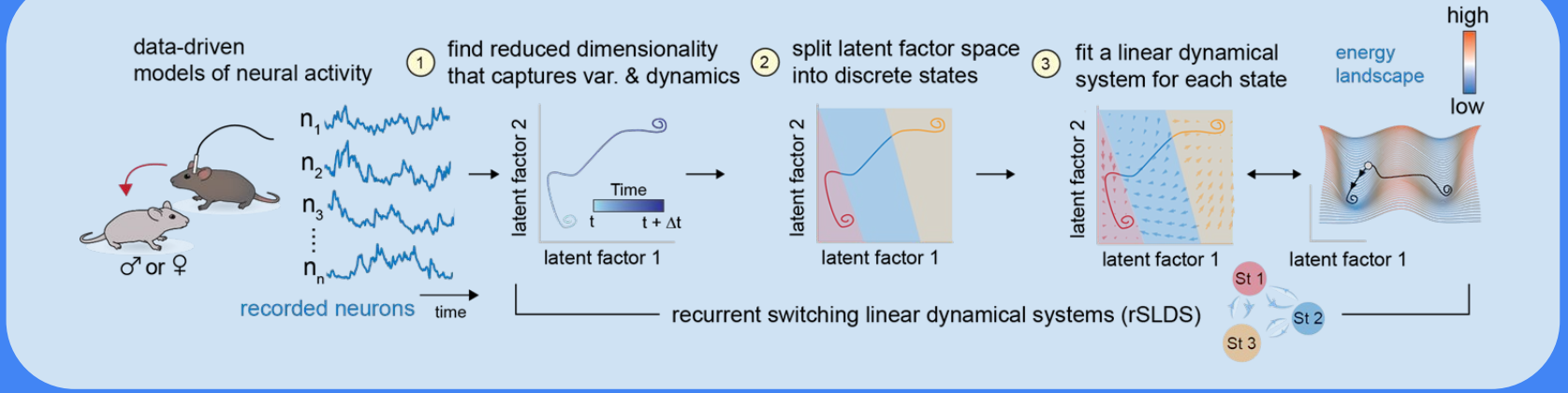
Caltech
datasai_2023

a dynamical systems approach in the hypothalamus



features from pose estimation such as velocity and distance between mice used as external inputs (u)

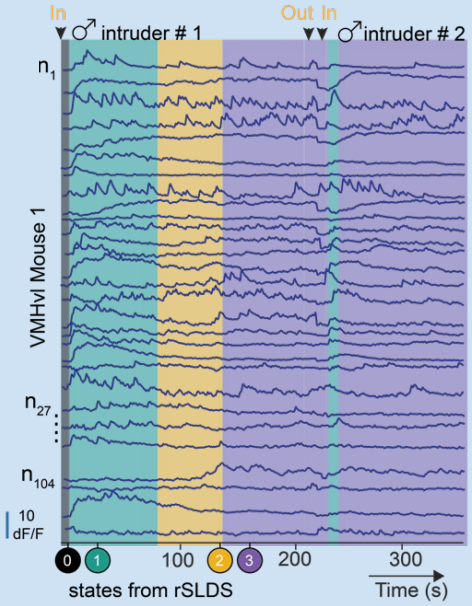
$$\frac{dx}{dt} = Ax(t) + Bu(t)$$

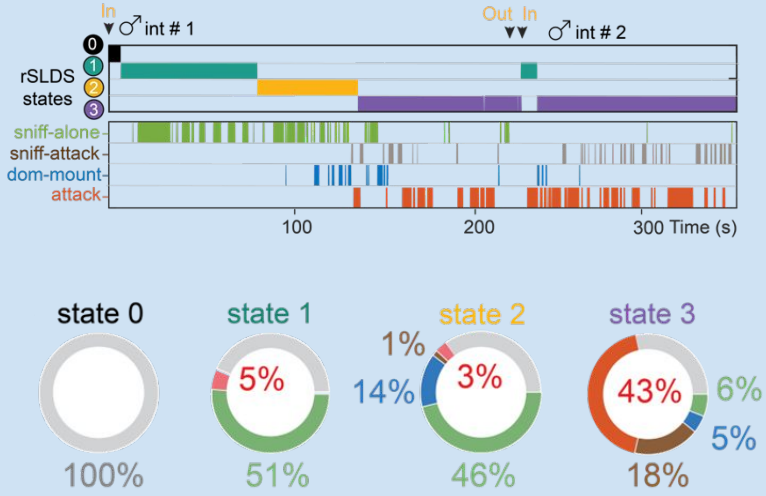


Caltech

datasai_2023

unsupervised dynamical models of the VMHvl reveal aggression enriched states





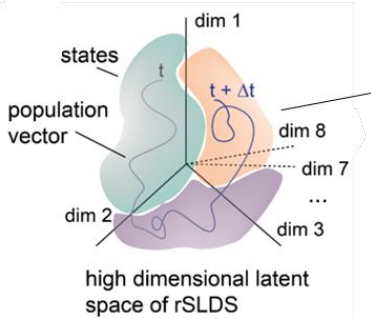
rSLDS discovers a representation of aggression without any reference to behavior annotations, *even with mixed selectivity*

Caltech
datasai_2023

how do we make sense of trained dynamical system models?

eigendecomposition!

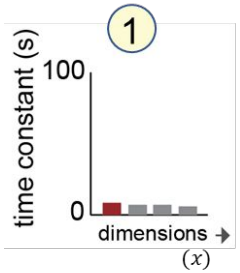
eigenvalues of A (λ) were converted into a time constant

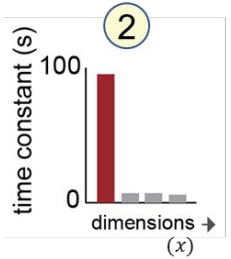


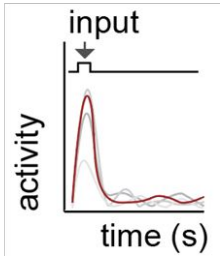
$$\frac{dx}{dt} = Ax + b$$

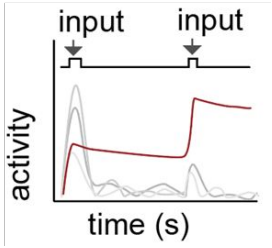
$$A \rightarrow \tau_1, .., \tau_n$$

systems with a single large τ are integrators or line attractors





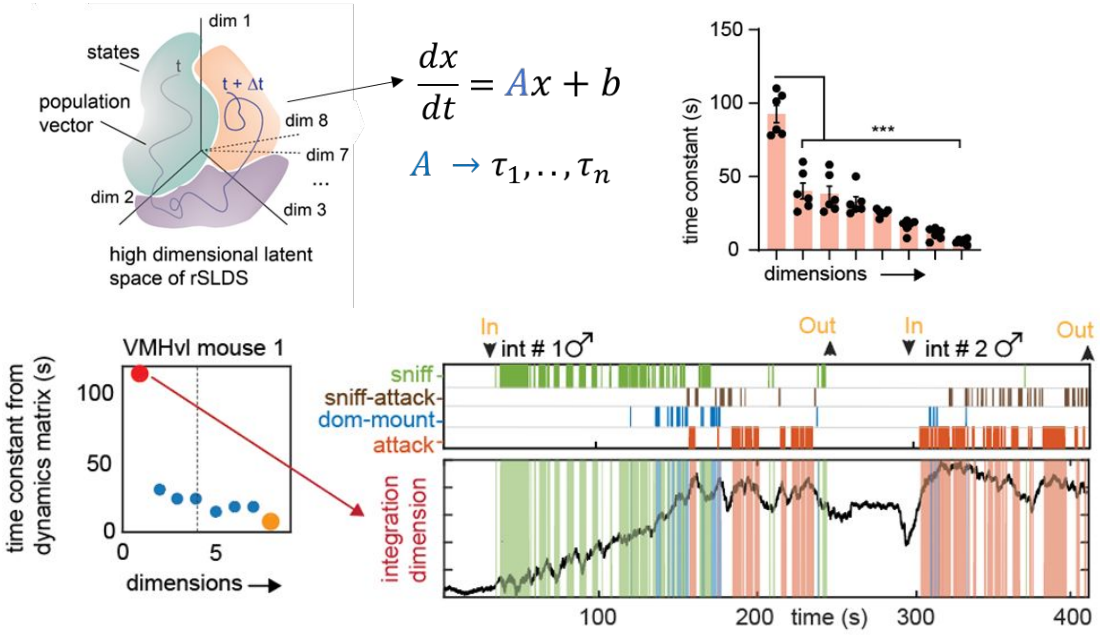




Caltech

datasai_2023

slow dynamics and integration of aggression in the VMHvl

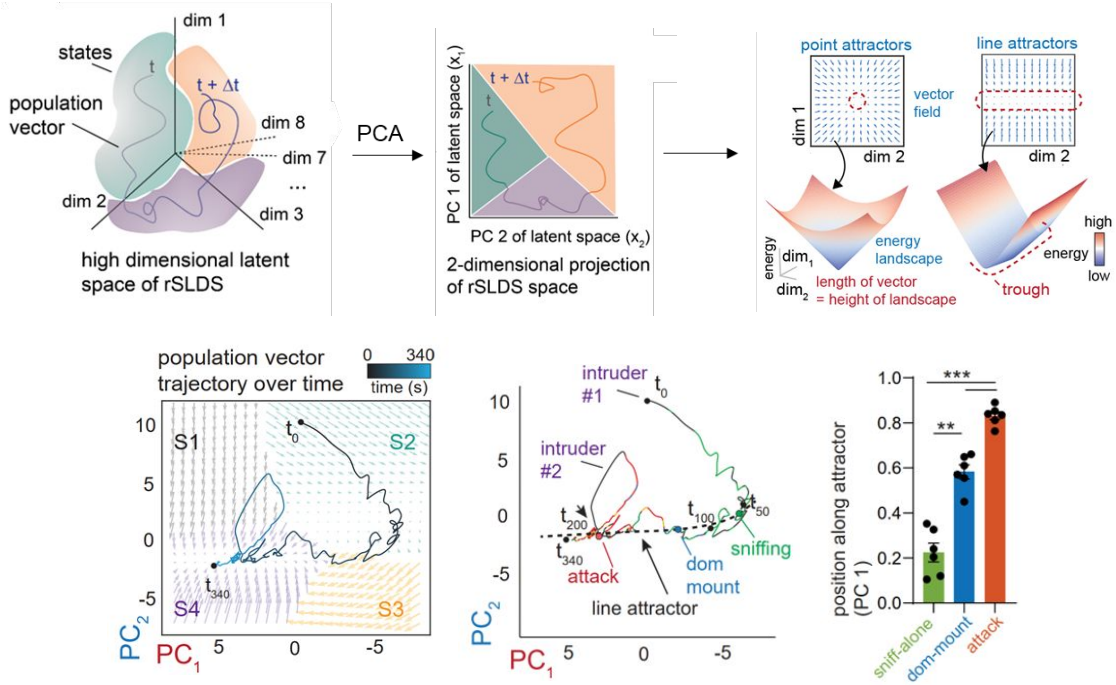


dynamical analysis of VMHvl reveals an integration dimension

Caltech

datasai_2023

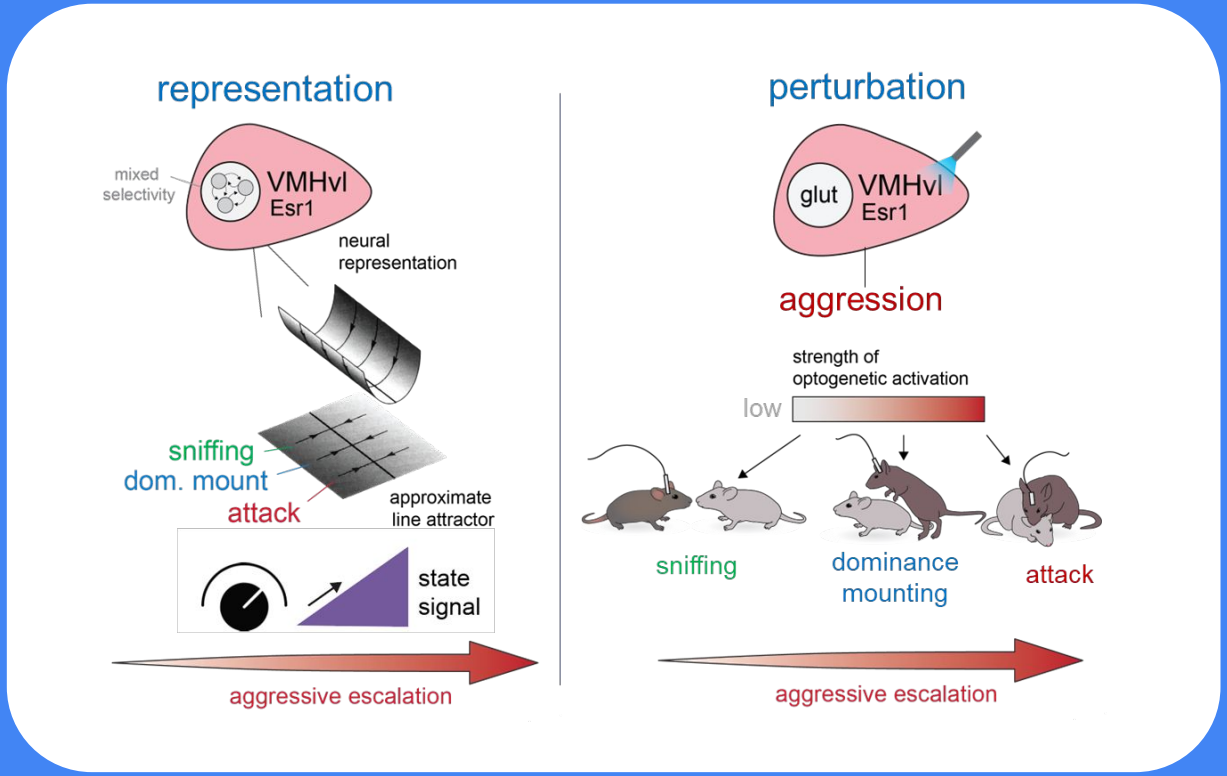
dynamics landscape of VMHvl reveals a line attractor



the underlying flow field in VMHvl shows line attractor dynamics

Caltech

datasai_2023

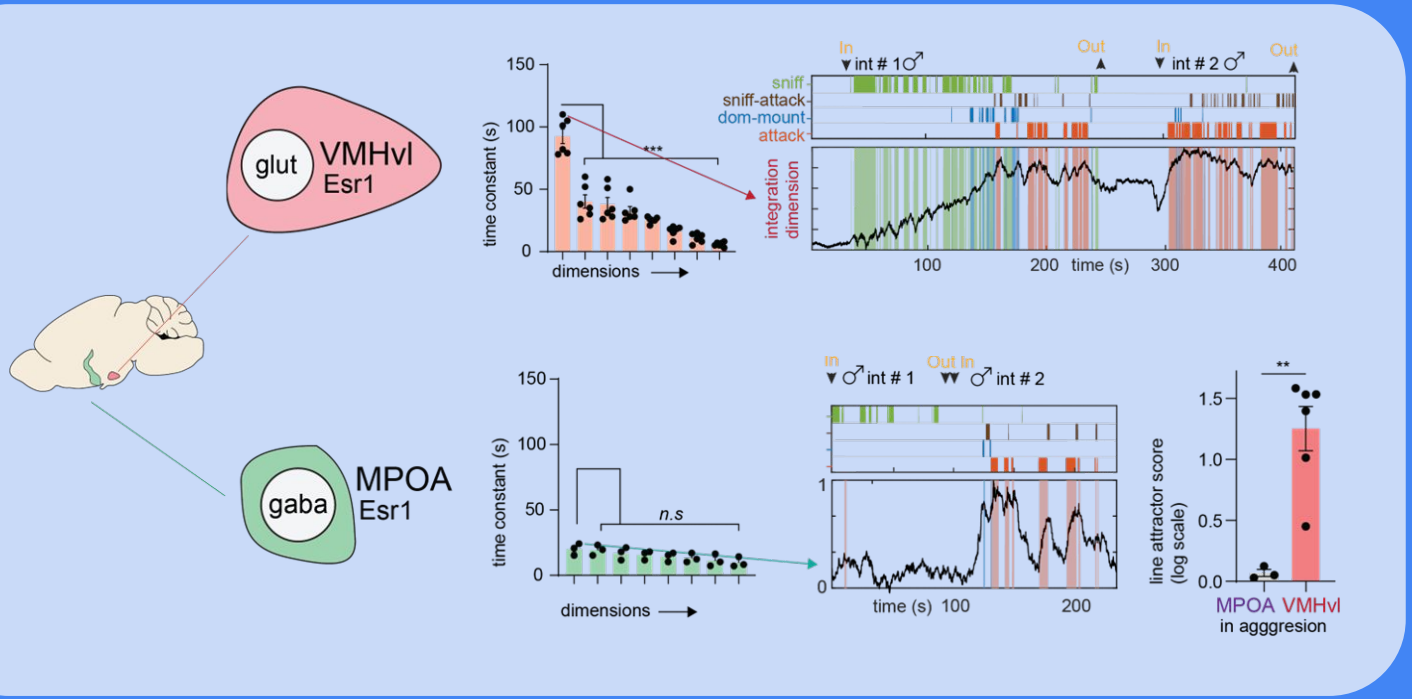


dynamics landscape of VMHvl reveals a line attractor

Caltech

datasai_2023

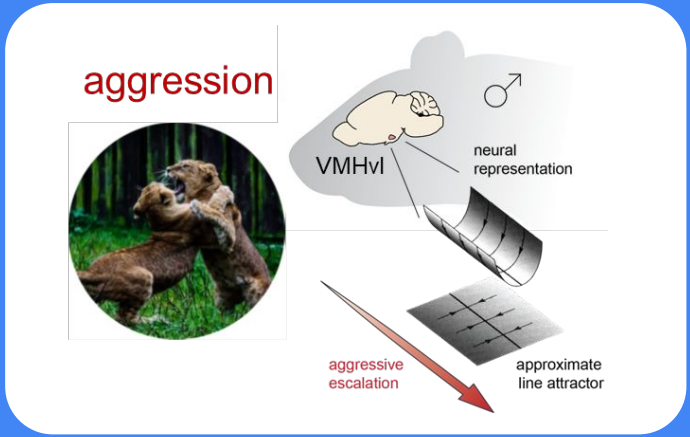
was the line attractor a trivial consequence of fitting a dynamical system to neural data?



not all regions possess a representation of state, some are more tuned to motor actions

Caltech
datasai_2023

are we observing a real line attractor?



line attractor dynamics are intrinsic properties of neural circuits

classically formed by recurrent connectivity

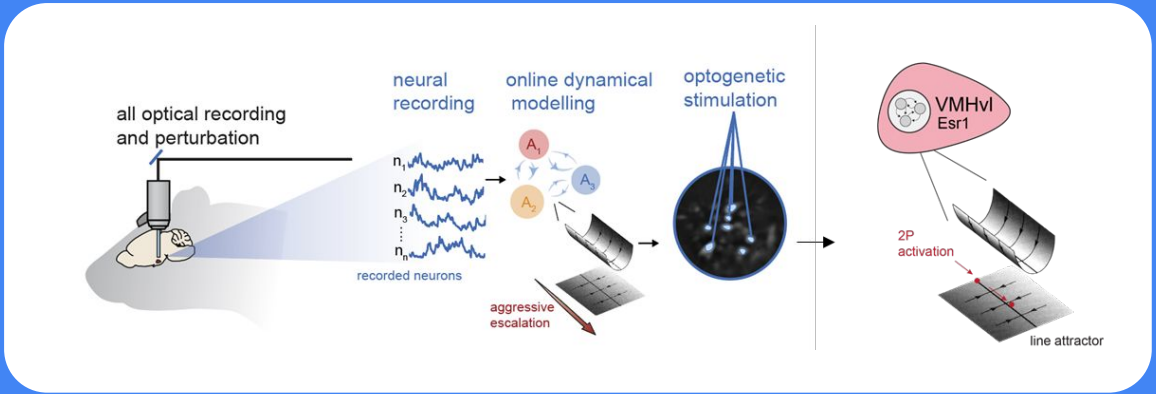
are observed line attractors in VMHvl simply ramping input, inherited dynamics or do they reflect dynamics truly intrinsic to VMHvl?

no line attractor in mammals has been able to formally rule out these possibilities

Caltech

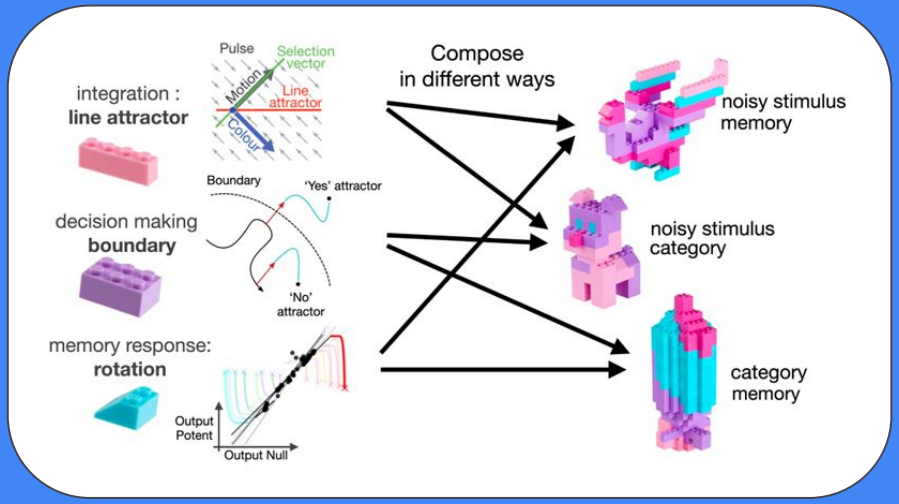
datasai_2023

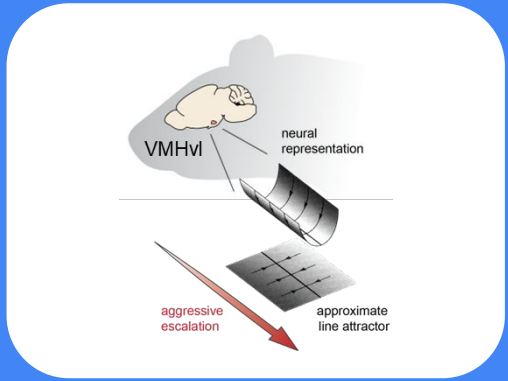
observation data alone is not sufficient, we need neural perturbations combined with data-driven modelling



the future…

are dynamical motifs the units of computation?





flexible multitask computation in recurrent networks
utilizes shared dynamical motifs
Driscoll, Shenoy, Sussillo, 2023

Caltech
datasai_2023