Neuron

CellPress
OPEN ACCESS

Perspective

The neural computation of affective internal states in the hypothalamus: A dynamical systems perspective

Aditya Nair,[1,2,3,4,5,*] Amit Vinograd,[1,2,3,8] Mengyu Liu,[1,2,3,9] George Mountoufaris,[1,2,3,10] Scott Linderman,[6,7] and David J. Anderson[1,2,3,*]

[1]Division of Biology and Biological Engineering 140-18, California Institute of Technology, Pasadena, CA, USA
[2]Tianqiao and Chrissy Chen Institute for Neuroscience, California Institute of Technology, Pasadena, CA, USA
[3]Howard Hughes Medical Institute, California Institute of Technology, Pasadena, CA, USA
[4]Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, Singapore
[5]Institute of Molecular and Cellular Biology, Agency of Science, Technology and Research, Singapore, Singapore
[6]Department of Statistics, Stanford University, Stanford, CA, USA
[7]Wu Tsai Neurosciences Institute, Stanford University, Stanford, CA, USA
[8]Present address: Gray School of Medicine, Tel Aviv University, Tel Aviv-Yafo, Israel
[9]Present address: Department of Bioengineering, Stanford University, Stanford, CA, USA
[10]Present address: Department of Physiology, The University of Texas Southwestern Medical Center, Dallas, TX, USA
*Correspondence: aditya.nair@ntu.edu.sg (A.N.), wuwei@caltech.edu (D.J.A.)
https://doi.org/10.1016/j.neuron.2025.11.003

SUMMARY

Internal affective states accompany evolutionarily ancient survival behaviors such as mating, aggression, and predator defense and may contribute to emotional feelings in humans. In this perspective, we introduce a dynamical system framework for thinking about such states. We synthesize evidence from recent studies suggesting that key state features, such as their intensity and duration, may be encoded by approximate line attractor manifolds in the hypothalamus. Evidence for these attractors arises from unsupervised data-driven dynamical system modeling of high-dimensional calcium imaging data from genetically identified cell populations in freely behaving mice. Dissection of the fit dynamical models and closed-loop modeling with experimental perturbations raise new questions regarding circuit- and cellular-level mechanisms of attractor implementation. These findings challenge prevailing views of hypothalamic behavioral control and afford a new avenue to study the emergence of slow state-encoding neural dynamics across scales, from single neurons to recurrent networks and neuromodulatory signaling.

INTRODUCTION

A fundamental goal in neuroscience is to understand how animals use sensory cues to produce behavioral outputs. Essential to this process is the concept of the internal state, a low-dimensional "intervening variable,"[1] which can integrate the influence of multiple sensory cues and transform them into a low-dimensional representation that guides behavioral decisions and motor actions.[2–4] Internal states can either be evoked by specific behavior-promoting sensory cues or be a prior condition that modulates the transformation of those cues into behavior. For example, a state of aggressiveness can be induced by sensory cues from a competing conspecific male,[5–7] but aggressiveness can also be promoted by extended social isolation.[8] Thus, internal states can not only drive goal-directed behavior but also provide adaptability, allowing organisms to flexibly modify their actions based on past experiences and current needs.[2]

How such internal states are represented by neural activity remains an important unsolved problem.[2] Innate "survival" behaviors such as mating, fighting, predator defense, parenting, feeding, and drinking provide tractable experimental platforms to address this issue in different model organisms, because they are evolutionarily ancient and are accompanied by strong motivational, drive, or affective states. While such states have traditionally been measured using assays from behavioral psychology,[9] they must also have corresponding neural representations that can be measured and described mathematically. It is the identification of these representations that is the topic of this perspective.

Determining where in the brain internal affective states underlying naturalistic behaviors are encoded and identifying the neural activity patterns that represent such states are hindered by the problem of how such states can be operationalized. Many studies infer the presence of an internal affective state based on observable behavior. For example, a state of aggressiveness is typically inferred by observing fighting behavior, while a state of fear is inferred by observing freezing behavior.[10] However, this approach risks conflating neural representations of internal states with those of the behaviors that express those states (or of the sensory stimuli that evoke them). For example, neural activity in the amygdala or hypothalamus during freezing does not necessarily represent a state of fear[11]; it could instead encode a

This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

sensory stimulus that evokes fear, a pre-motor pattern for freezing, or a representation of plastic changes that occurred during fear conditioning.[12] It is difficult to disentangle these representations when such cues, states, and behaviors are highly correlated.

In the case of behaviors that maintain physiological homeostasis, like eating or drinking, motivational states can be measured independently of consummatory behaviors using incentive-guided operant conditioning assays with instrumental readouts,[13,14] such as bar-pressing or nose-poking to obtain rewards.[1,9] Here, what is measured is whether the animal will do work and, if so, how much work (i.e., how many nose pokes or bar presses it will perform) to obtain a reward, rather than consumption of the reward itself. These work-related task variables are thereby taken as an operational readout of an animal's internal motivation state ("hunger" or "thirst"). Standard supervised approaches can then be applied to identify neural representations of such task variables in concurrently recorded brain activity[13,15] and to distinguish them from representations of nose-poking or lever-pressing behavior itself.[16] Moreover, because the animals used for operant assays are highly trained to perform a repetitive artificial behavior, neural data can be averaged from hundreds of trials to increase statistical power.[17,18] This approach has provided valuable insights into the neural coding of motivational states such as thirst.[13]

In contrast to behaviors that restore physiological homeostasis, innate behaviors such as fighting, mating, or predator defense are not elicited by resource deprivation and are therefore more difficult to adapt to operant conditioning paradigms designed to measure need-based motivational drives. Furthermore, operant assays require intensive training, which defeats the purpose of studying the neural coding of internal states underlying innate behaviors. While operant assays have been adapted to measure aggressive motivation,[19] these assays do not measure internal states underlying aggressive behavior itself: a mouse cannot both attack a conspecific and press a bar for a reward at the same time. Finally, operant assays reduce the dimensionality of affective internal states to reward vs. punishment, and there are important additional dimensions to these states[4,20,21] that are not measurable by reinforcement learning. For this reason, it has been argued that neural signals encoding affective or emotion states cannot be studied in animals, because there are no state-specific task variables that can be measured and correlated with neural activity.[9] If one accepts the premise that neural signals encoding affective states cannot be studied in the absence of a behavior-driven approach, it is difficult to refute this argument.

Here, we describe an alternative approach to identifying neural representations of internal affective states that is neural data driven rather than behavior driven. Briefly, this approach is based on the idea that such representations should differ from behavior representations in their neural *dynamics*: they may exhibit features such as persistence and continuous scaling (ramp up or down)[4,20] that are uncharacteristic of behavior representations, which are usually time-locked to the discontinuous bout structure of specific motor actions. We provide examples to show how candidate internal state signals exhibiting these properties can be extracted from high-dimensional neural data through unsupervised dynamical system modeling, an approach agnostic to task variables.[22] These signals can be visualized as manifolds (i.e., low-dimensional surfaces) embedded in neural state space,[23,24] such as line attractors,[25,26] and aligned retrospectively with annotated video recordings to interpret their behavioral relevance.[27,28] Moreover, although these statistical models are agnostic to implementation, they can be used to investigate experimentally how these emergent dynamics are implemented mechanistically in neural connectivity and signaling.[29,30] In this way, this neural data-driven approach frees the study of affective internal state coding from the shackles imposed by a traditional behavior-driven approach.

UNDERSTANDING NEURAL CONTROL OF INNATE BEHAVIORS THROUGH PERTURBATIONAL VS. OBSERVATIONAL STUDIES

The approach described above has relied on technical advances that have enabled a gradual shift from perturbational to observational approaches to the neural control of innate affective behaviors. Traditionally, perturbational rather than observational approaches have driven this field for several reasons. One is historical: the field was established by the discovery that electrical stimulation of discrete brain regions could evoke certain behaviors (e.g., aggression).[31] A perturbational approach has the additional benefit of establishing causation, while observational studies only establish correlations. In addition, observational studies of brain activity during innate behaviors were initially restricted by reliance on single-unit recordings, whose low yield limits conclusions that can be drawn about distinct subpopulations.[32] The practice of approximating population activity measurements in single animals by trial-averaging single-unit recordings from different neurons, while applicable to highly repeatable learned behaviors, is not applicable to innate behaviors due to their high intrinsic variability.

This limitation has been overcome by the development of methods to record simultaneously from hundreds or thousands of neurons in a single animal during a single trial of an innate behavior. While multi-channel extracellular electrodes afford such sampling,[33] fast, sensitive genetically encoded calcium indicators (GECIs)[34] permit imaging of hundreds of genetically identified cells in populations whose causal role in behavior can be established independently by optogenetic or chemogenetic perturbation studies. At the same time, the development of miniature head-mounted microscopes fitted with gradient refractive index (GRIN) lenses[35,36] has transformed imaging (both one and two photon) in deep subcortical structures such as the hypothalamus[37,38] and amygdala,[12,39] sites where artificial stimulation can evoke specific innate behaviors.[40–44] Together, these technological advances have revolutionized the ability to obtain high-dimensional neural data from mice performing individual trials of instinctive behaviors.

In some cases, such population imaging experiments have confirmed that hypothalamic neuron subsets that promote a given behavior when artificially activated exhibit activity time-locked to natural occurrences of the same behaviors. For example, studies in the medial preoptic area (MPOA) have identified specific cell classes that can evoke parental behavior[45] or

CellPress
OPEN ACCESS

Box 1. Primer on dynamical systems theory in neuroscience

In its simplest form, a dynamical system tracks the evolution of a state, y, over time as a function of inputs, u (Figure 1A). The dynamics are specified by a differential equation, $\frac{dy}{dt} = f(y,u)$, that describes how the current state and input determine the change in state. The state y could represent the activity of a local population of interconnected neurons or that of neurons connected across various brain regions or nuclei. A dynamical system may be linear or nonlinear. *Linear* dynamical systems may be written as, $f(y,u) = Ay + Bu$ (where A and B are matrices), since y on the right-hand side of the equation appears only to the first power. If the equation involved products, powers, or other functions of y (for example, y^2, $\cos y$, etc.), then the dynamical system would be nonlinear.

The mathematical tractability of linear dynamical systems makes them attractive models. For instance, consider a simplified neural circuit with two state variables, $y = (y_1, y_2)$. For example, the variables could represent the average activity of two non-overlapping cell populations (Figure 1B), but this framework allows for more complex setups as well. In this system, the matrix A is a 2×2 connectivity matrix that describes the strength and sign of connections between and within these populations (Figure 1C, left; for simplicity, these populations are represented by single neurons in Figure 1B). In this simple example, the entries of the dynamics matrix Adictate the strength of excitatory or inhibitory influences between and within the two populations (for the mathematical formalisms describing such a system, see Figure S1).

The dynamics matrix can reveal emergent properties of the network via its eigenvectors and eigenvalues (Figures S1B and S1D). The eigenvectors specify the "modes" of the system, and the eigenvalues specify how activity evolves along each dimension in time (see the supplemental appendix for non-mathematical intuition). Eigenvalues have two numerical components: real and imaginary. Eigenspectra (the collection of eigenvalues extracted from the dynamics matrix) are often displayed in a two-dimensional plot, whose axes are the real and imaginary components of the eigenvalue (Figure 1C, middle). The real part of the eigenvalue determines the rate at which activity decays or grows along that dimension (eigenvector), following an input aligned with that dimension. Negative real eigenvalues far below 0 (Figure 1C, middle, red point) yield fast (small) time constants (Figure 1C, right, τ_2) leading to exponential decay back to the fixed point (here, the origin) and are stable (Figure 1D, "short time constant"); eigenvalues > 0 lead to an exponential increase in activity along that dimension and are unstable, and zero eigenvalues yield marginally stable systems (Figure 1C). A slightly negative real eigenvalue (i.e., just below zero) indicates the presence of a nearly stable mode with a slow (large) time constant (Figure 1C, middle, blue point and right, τ_1). Such slow modes allow the cell population to produce persistent activity after receiving an external input (Figure 1D, "long time constant"). (Note that sometimes linear dynamical systems are presented in "discrete time" via an update rule of the form $y_{t+1} = Ay_t + Bu_t$, rather than as a continuous time differential equation, and in this formulation stability is determined by the magnitude of the eigenvalues rather than the sign of their real part.[56])

Let's first consider a scenario with a circuit motif where there is no functional connectivity between y_1 and y_2 and which possesses self-inhibition (Figure 1E, blunt arrowheads). Depending on the specific entries in the dynamics matrix, the eigenspectrum for this dynamics matrix can result in two eigenvalues that are both negative and have short time constants (Figure 1F). We can visualize the activity of both modes in a two-dimensional plot of their activity against each other. This set of all possible combinations of y_1 vs. y_2 activity in a neural system is called the neural state space. The pattern of connectivity determines the "flow field": a set of arrows at each point in neural state space that together dictate the direction and speed with which activity changes at any given point in this two-dimensional state space. (Figure 1G). A trajectory through neural state space, called the "population vector," shows how activity evolves over time in response to inputs (Figure 1H). Some flow fields (conservative fields) can also be interpreted as the result of a ball rolling on a 3D "energy landscape." In this example, we see that the arrows in the flow-field map point to a single position in state space that acts as a point of stability point (Figure 1G). The presence of such a single fixed point reveals a "point attractor," which can be visualized in an energy landscape as a well (Figure 1I). Activity in other regions of the flow field tends to flow into this well (hence the name "attractor"), and once in it, the system resists other influences that try to push the state out of the well, e.g., from external inputs (Figure 1H, blue arrows). In essence, the attractor can be thought of as a "sticky" point in neural state space (for further non-mathematical intuition, see the supplemental appendix).

The pattern of connectivity can also create other strong constraints on where activity can flow in neural state space. Consider another system, with a circuit motif that features functional connectivity between y_1 and y_2 in addition to the self-inhibition present in each node (Figure 1J). The resulting eigenvalues are real, and depending on the precise entries of the dynamics matrix, one eigenvalue can approach 0 (Figure 1K, blue point; see also Figure S1C, λ_1). Linear systems with eigenvalues approaching 0 are marginally stable and display a flow field where arrows point to a line of nearly stable points (Figure 1L), which corresponds to a long, shallow trough in the energy landscape (Figure 1N)—this is called a line attractor. External inputs to the system can kick the system off the attractor, and then the intrinsic dynamics pull the neural state to another stable point along the line, allowing the state to represent integrated inputs over time (Figure 1M). The position along the line encodes a continuous variable, and since the points along the line are (nearly) stable, the encoded value persists for long periods of time. For this reason, a line attractor is often referred to as a type of "continuous" attractor. Note that biological line attractors are not infinitely stable due to sources of leaks in the system and/or energy limitations and therefore are approximations of an ideal mathematically defined line attractor (Figure 1D).

(Continued on next page)

Box 1. *Continued*

Many more types of flow fields can also be obtained depending upon the precise values of the dynamics matrix,[26] including oscillatory or rotational dynamics, another commonly observed computation related to sequential activity and decision thresholds seen in cortical and subcortical circuits,[28,57] which may be inferred also from linear systems based on the imaginary component of eigenvalues. However, many computations require *nonlinear* dynamical systems. For example, bistable systems with multiple point attractors require nonlinear dynamics.[58] Likewise, the ring attractor for maintaining heading direction in *Drosophila*[59] arises from a nonlinear dynamical system. Data-driven modeling techniques, like LFADS and rSLDS, employ various strategies to approximate nonlinear systems as linear systems and shed light on their fixed points and dynamical structure.

male mating[46,47] when optogenetically activated. Calcium imaging experiments have confirmed that these neurons are specifically active during the naturalistic performance of these behaviors in freely moving animals.[28,46,47] Such observations support the idea that certain innate behaviors may be represented by "cell identity" codes[28,48] in the hypothalamus.

In other settings, however, such studies have failed to yield clear correlations with causal function or natural behavior. For instance, optogenetic activation of estrogen receptor 1 or progesterone receptor-expressing neurons in the ventrolateral subdivision of the ventromedial hypothalamus (VMHvl$^{Esr1/PR}$) can reliably trigger aggressive behavior,[49–51] yet calcium imaging of such neurons revealed that only a small fraction of these cells was specifically tuned and time-locked to attack.[37,46,52] Instead, most neurons in this nucleus exhibited "mixed selectivity": they were activated to varying extents during multiple social behaviors, e.g., sniffing and attack. Furthermore, their dynamics were not synchronous with the onset and offset of specific motor actions.[28,37]

From a more traditional perspective, these latter results may appear paradoxical: how can a population of neurons trigger a specific behavior in a time-locked manner when optogenetically activated if the activity of these neurons is neither specifically tuned nor time-locked to natural occurrences of the same behavior? Here, we present a different way of thinking that resolves this paradox and yields new insights into the neural coding of internal states. An essential feature of this perspective is its focus on neural dynamics.

THE UTILITY OF INCORPORATING NEURAL DYNAMICS IN THE STUDY OF INTERNAL STATES

Neural signals that are not time-locked to motor behaviors are of particular interest because, as mentioned earlier, dynamics are a key feature that can distinguish neural activity encoding internal states from that encoding observable behavior. Innate behaviors are often punctate: they are typically expressed in bouts, with well-defined onsets and offsets. By contrast, introspection and intuition suggest that internal affective states should precede and outlast the behavioral bouts that express them: our subjective experience of rising anger occurs before we display that anger behaviorally (e.g., by picking a verbal or physical fight) and takes time to dissipate even after the fight is over. If one accepts this premise, it follows that neural signals representing an internal affective state should exhibit such slow dynamics as well.[4,20] An additional feature of such states is that their intensity often escalates over time.[4,53] This suggests that scaling or ramping may be an additional characteristic of internal state-coding neural signals. Thus, if patterns of persistent activity and ramping in neural dynamics are apparent during an otherwise intermittent and sporadic behavior, such signals may be candidates for encoding internal affective states. In this way, unsupervised data-driven analysis of neural population activity and dynamics may reveal features that distinguish the representation of an internal state from that of the associated motor behavior, even in the absence of a method to operationalize and measure such states behaviorally.

How can one analyze high-dimensional neural data to identify dynamic signals that may encode internal states? One approach is to model the neural data as a dynamical system. A dynamical system is a set of equations that describe how phenomena evolve in time.[54,55] These phenomena may pertain to anything from the behavior of a pendulum to that of a neural circuit. For example, the firing rates of a network of neurons can be modeled as a dynamical system (Box 1). The model captures the influence of each neuron's activity on the future firing rates of others, based on their connectivity. It also incorporates the influence of input to the network from external sources (Figure 1A). Importantly, the network's behavior may be described more concisely by a set of lower-dimensional latent variables that capture correlations in the firing rates of groups of neurons, rather than of each individual cell, in the network (Figures 2B and 2C). Whatever its dimensionality, a dynamical system model can reveal the coding of certain variables by *temporal changes* in the activity of the system rather than simply by the instantaneous or average activity in the neurons.

Formally, a dynamical system model is a set of differential equations that describe the time evolution of a mathematically defined system. The system may exist in many different states, and a dynamical system model specifies how those states evolve over time.[54] In the context of a neural circuit, the system may consist of a set of neurons, and the states may correspond to the activity of individual neurons or of groups of neurons. The equations of a dynamical system define how those states change over time and in response to external inputs, thereby characterizing the future activity of the neural circuit. Importantly, the exact pattern of observed activity can be highly variable from trial to trial. This variability could stem from differences in the "initial conditions" (the state of the system at the beginning of a given trial), from differences in the external inputs, or from intrinsic noise in the dynamical system (if the model is stochastic). In the context of neural circuits, the instantaneous activity of the system often corresponds to a point in a "neural

CellPress
OPEN ACCESS

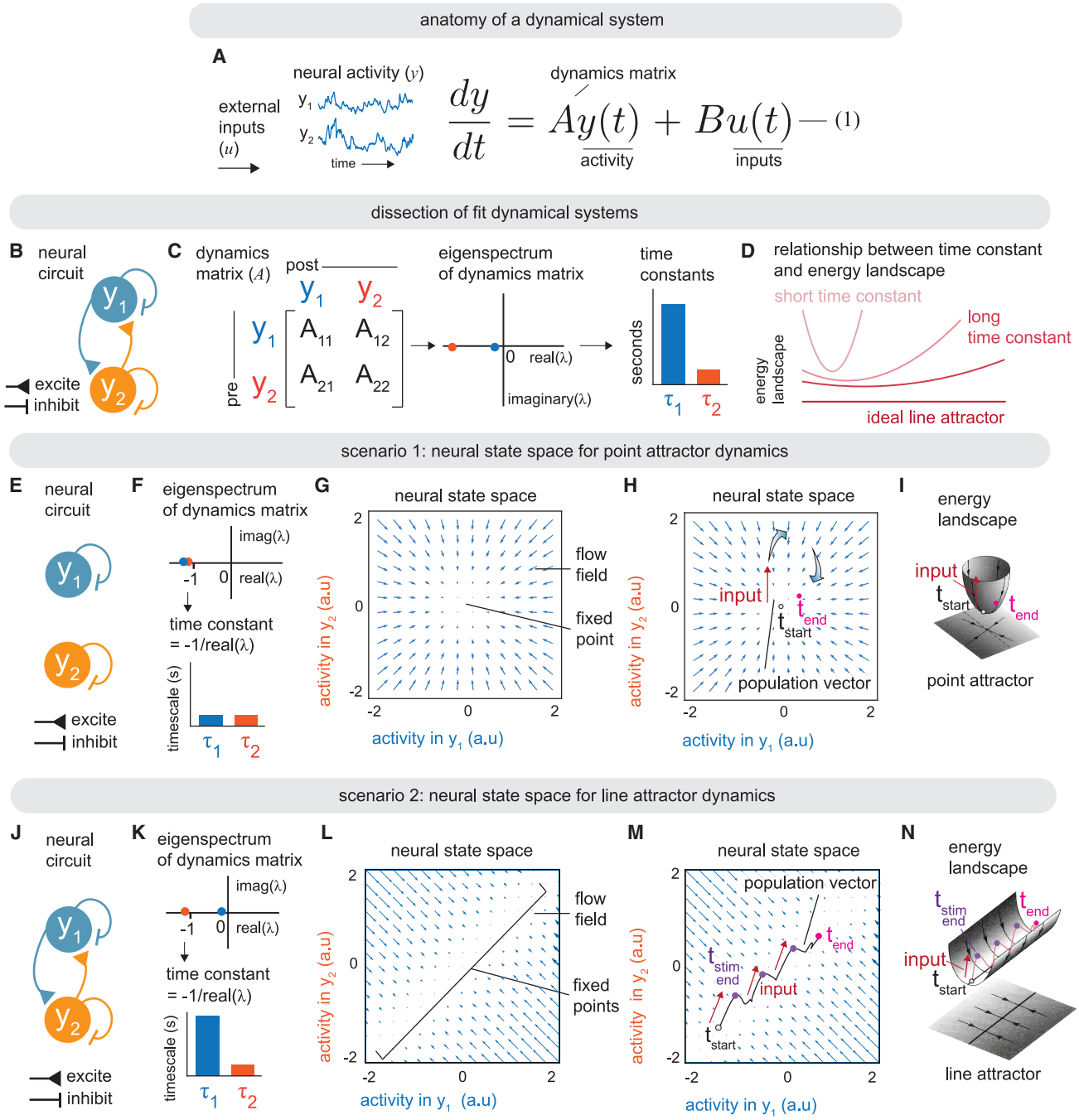


Figure 1. Anatomy of a linear dynamical system

(A) Left: cartoon depicting a neural recording with activity (y) of a two-neuron system, as well as external inputs (u) to the circuit. Right: equation for a generic continuous linear dynamical system. The dynamical system tracks the evolution of the state y. The dynamics matrix (A) accounts for interactions within the network y, and the matrix B specifies how inputs influence the state.

(B) A cartoon of a coupled neural system consisting of two populations, y_1 and y_2.

(C) Left: a dynamics matrix displaying the functional coupling between dimensions y_1 and y_2. In this case, A can be interpreted as a 2×2 connectivity matrix with presynaptic populations on the columns and postsynaptic populations on the rows; thus, A_{12} indicates the strength of the connection of y_2 onto y_1. Middle: understanding the dynamics matrix through eigendecomposition, plotted in two dimensions (called eigenspectra) to show the real and imaginary parts of the eigenvalues. See the supplemental appendix and Figure S1 for a detailed mathematical description. Right: the magnitude of the eigenvalues can be used to calculate a time constant, which shows the rate of decay along each dimension in the presence of an external input. The value of the time constant for each dimension is inversely proportional to the negative of its eigenvalue. Smaller time constants (eigenvalues of lower magnitude, e.g., −10) mean faster rates of decay, and larger time constants (eigenvalues of higher magnitude, e.g., −2) mean slower rates of decay.

(legend continued on next page)

state space," and over time, the system traces out a "trajectory" through this high-dimensional space (see Box 1 and the supplemental appendix).

These trajectories are governed by a set of differential equations that describe the rate of change of neural state at each point in state space. They can be visualized either as a "flow field" of arrows, whose direction and length at each point in space represent the direction and rate of change of the system (Figures 1D and 1H), or, in some cases, as an "energy landscape," whose features are shaped by the rules the system obeys (Figures 1F and 1J). The predicted behavior of the system on any given trial is a time-evolving state, sometimes called the "population vector," that moves through this landscape, representing the collective activity of all the neurons in the network (Figures 1E and 1I). In a linear dynamical system, the differential equations are linear functions of the current state, as represented by the dynamics matrix *A* in the first equation in Box 1.[54] Linear dynamical systems models are convenient because they are mathematically tractable, as discussed in Box 1 (for a more detailed treatment, see Yang et al.[61]). However, linear dynamical systems models are ultimately limited in their expressivity, and to model more complex features of neural circuits, we must turn to nonlinear models instead. Next, we describe how dynamical systems can aid our understanding of neural circuits and how these models can be estimated from measurements of neural activity.

UNSUPERVISED DATA-DRIVEN DISCOVERY OF POPULATION DYNAMICS IN NEURAL CIRCUITS

Population dynamics can emerge at multiple levels in a neural circuit: they can be inherited from shared inputs to the neurons constituting the circuit, they can emerge from intrinsic dynamics due to interconnectivity between these neurons, and they can even arise from biophysical properties of single neurons that give rise to persistent states. Complex dynamics may also emerge through meso-scale interactions between nuclei or brain regions.[63] In systems characterized by strong recurrence, emergent intrinsic dynamics resulting from the interconnectivity between regions and between neurons significantly influence neural dynamics. The framework of dynamical systems provides a conceptual foundation grounded in engineering and physical sciences for understanding how feedback influences ongoing dynamics in physical systems.[54] This approach has been instrumental in modeling how neural circuits, conceptualized as dynamical systems, perform a wide range of computations,[64] from generating movement[17] and timing actions[65] to tracking heading angle,[59,66] making decisions,[58,67] and mapping space in the brain.[64] Importantly, it can provide an understanding of the system's behavior and function at a level of abstraction that does not require a detailed understanding of the mechanistic implementation of feedback, and it is top down in describing computations, rather than bottom up in describing biological implementation.

Studying the properties of a dynamical system can uncover important emergent features of neural circuits that arise through the interactions between neurons (refer to Box 1). These features may include persistent activity and escalation or ramping that have been hypothesized to encode features of internal states.[4,20,68] These properties, if they exist, can be inferred by applying dynamical systems analysis to neural population activity. Importantly, the use of unsupervised analysis means that these features are not prior assumptions that constrain the model but rather will (or will not) emerge from the analysis depending on the system's behavior. In one instantiation of this approach, dynamical models of the type illustrated in the equation (Box 1; Figure 1A) are directly fit to the observed activity of a population of neurons. Inspection of the fitted model may reveal whether the system exhibits attractor dynamics, a property that can allow information to be maintained in persistent neural activity[26,62,69,70] (see the supplemental appendix for an intuitive description). While theoretical studies have long implicated the role of attractor dynamics in encoding variables such as movement,[71–74] memory,[75,76] spatial locations,[77] or eye position,[62] experimental studies have yielded increasing examples of biological attractor-like representations through the application of dynamical systems models to actual neural data.[24–26] For example, experimental evidence for "point" attractors (Figures 1G–1I) that encode working memory has been provided in the anterior lateral motor cortex (ALM).[67]

(D) Relationship between attractor basin of energy landscape and time constant. Systems with slow time constants approximate line attractor dynamics, with reported biological instantiations of line attractors being approximations of an ideal line attractor.[26,60] An ideal line attractor would show an infinitely long time constant ($\lambda = 0$) and would possess a perfectly flat attractor basin. See also Figure S1C.
(E) Scenario 1: a cartoon of this coupled neural system consisting of two populations with no functional connectivity between y_1 and y_2. This system also features self-inhibition within populations.
(F) Eigenvalues and time constants of dynamical system.
(G) The neural state space shows the possible configurations of activity in this system. The flow field, determined by the dynamics matrix *A*, shows how activity evolves in neural state space. In this example, the flow field reveals a single stable fixed point, which defines an approximate point attractor in neural state space.
(H) Same as (E), but showing the state evolution when the system is provided with a transient input that stimulates y_1 and y_2. Each input pushes the system out of the point attractor, following which it returns to the attractor (due to the strong self-inhibition).
(I) Same as (E), but showing the flow field as a 3D energy landscape. The slope of the landscape is proportional to the strength of flows (length of arrows in G) at any point in state space. Thus, the line attractor appears as a well, with the base of the trough indicating the stable point of the dynamical system.
(J) Scenario 2: a cartoon of a coupled neural system consisting of two populations that excite one another. The excitation is balanced by self-inhibition within populations.
(K) Resulting time constants with an eigenvalue that approaches 0.
(L) Neural state space of this system showing a line of stable points constituting a line attractor.
(M) Same as (L), but showing the time-evolving population neural trajectory with transient inputs (red arrows). The input pushes the neural state out of the line attractor, with the state returning to a different fixed point from its location when input was initiated. This property allows the neural system to "integrate" (i.e., to accumulate and store information about the history of the inputs) along the line attractor.
(N) Same as (M), but showing the flow field as a 3D topographic energy landscape. The line attractor appears as a trough, with the base of the trough indicating the stable points of the dynamical system. Toy examples modified from Yang et al.[61] Diagrams in (I) and (N) are modified from Seung.[62] For an intuitive description, see the supplemental appendix.

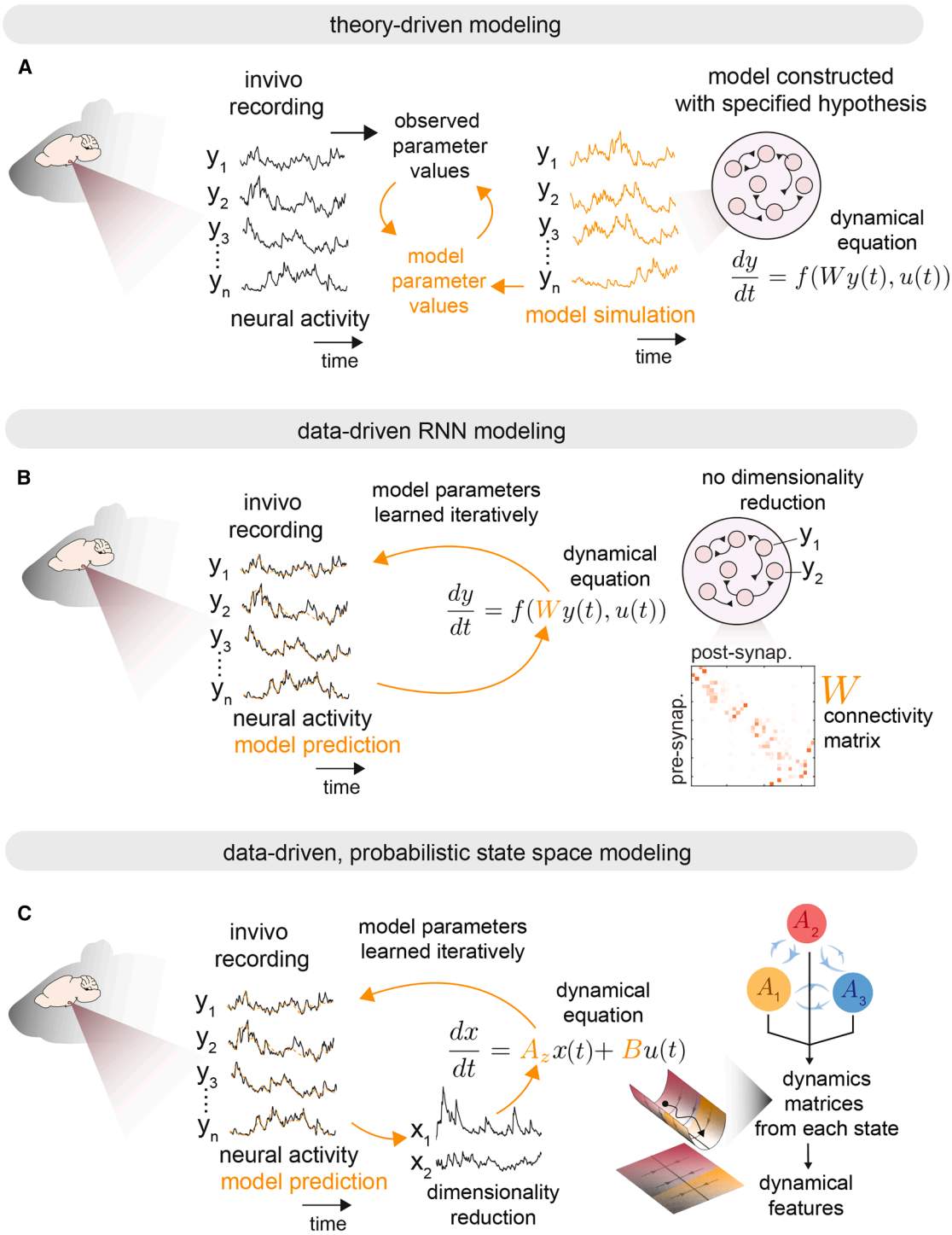


Figure 2. Theory vs. data-driven modeling to understand neural computation

(A) Illustration of the theory-driven modeling approach to understand neural computation. In this approach, multiple models are constructed with specific hypotheses, such as detailed connectivity patterns or intrinsic properties, and then each model is compared with neural data. The model whose parameters best fit the observed data is preferred.

(B) Illustration of the data-driven RNN modeling approach to understand neural computation. In this framework, nonlinear dynamical equations are fit to neural data in a manner that specifies the connectivity strength between individual neurons, and then the fit connectivity matrix can be studied to reveal neural computations.

(legend continued on next page)

Attractors can exist in other topologies, such as "line" or "continuous" attractors[25] (Figures 1L–1N). These continuous attractors can function as integrators, allowing a circuit to accumulate and represent a graded, integrated variable.[25,70] Each integrated value can be persistently represented within the circuit, enabling the encoding of continuous variables. The integrated positions function as "fixed points" that constrain activity within the circuit to exist within these points unless perturbed by external inputs to the circuit or by noise. Studies applying dynamical systems tools to neural activity have inferred representations of continuous attractors with ring topologies (a.k.a. "ring" attractors) important for encoding head direction in flies and rodents.[59,66,78] Here, the encoding of a continuous variable allows for different angles of head direction to be uniquely and persistently represented in activity.[59,66,78–80] Similar methods have also inferred continuous attractors encoding evidence in decision-making paradigms,[18] for reward history in learning paradigms,[81] and for encoding space.[79,80]

Because neural circuits are intrinsically nonlinear, they are traditionally modeled as nonlinear dynamical systems such as recurrent neural networks (RNNs). However, RNNs are difficult to interpret and make assumptions about implementation (i.e., circuit connectivity) that are not necessarily correct. Recent advances in machine learning have enabled an alternative approach to fitting neural data with dynamical systems models, which is particularly useful for single-trial recordings. In this approach, the algorithm "learns" the governing equations that produce the observed dynamics by iterative modeling of the data (Box 2). This data-driven approach contrasts with traditional modeling in neuroscience, where models are preconceived, intellectual constructs designed to approximate specific aspects of neural data (Box 2; Figure 2). This process of discovering dynamical properties employs new unsupervised machine-learning algorithms that fit various types of dynamical system models to neural data.[79] Since dynamical systems that represent neural circuits might be inherently nonlinear, some methods, such as latent factor analysis via dynamical systems (LFADS[81,82]), directly fit nonlinear models incorporating RNNs. Other approaches, such as recurrent switching linear dynamical systems (rSLDSs[28,83,84]), approximate nonlinear systems as a composition of linear dynamical systems, making them easier to interpret.

A key feature of both LFADS and rSLDS is that they use dimensionality reduction to identify dominant latent dimensions (also called "latent factors") with shared dynamical properties, which can explain a large fraction of variance in raw data. Dynamical models then explain the temporal evolution of these latent factors, which may arise from external inputs, noise, or intrinsic dynamics. Since these models are unsupervised, they do not assume any form of attractor dynamics or other dynamic mechanisms in the circuit. Furthermore, since they operate at the level of latent factors, they are agnostic to the exact implementation of the suggested dynamics of the underlying neural circuit. In other words, the dissection of these models can reveal the computation performed by the circuit but not the mechanism that implements the computation. Hence, these emerging methods permit unsupervised identification of putative dynamical features of neural activity and generate hypotheses for how those dynamics operate at a more abstract level.

DYNAMICAL ANALYSIS REVEALS AN ENCODING OF INTERNAL STATES AS CONTINUOUS ATTRACTORS IN STATE-CONTROLLING CIRCUIT NODES ACROSS BEHAVIORS

A set of recent studies has applied unsupervised dynamical systems analysis directly to neural activity from genetically marked neuron subsets within a key hypothalamic circuit node that controls innate social behaviors (male VMHvl^{Esr1} neurons[28] and female VMHvl$^{Esr1+,Npy2r-}$ neurons[27]; Figure 3). Functional manipulation of these and related genetically marked VMHvl populations (VMHvl^{PR}[61]; VMHvl^{Cckar}[112,113]) has revealed a causal role in regulating male[32,50,51,61] and female[49,114] aggressive and female sexual[112–114] behaviors, as well as in aggressive motivation[19] (reviewed in Mei et al.[115]). Although studies using fiber photometry in these populations have reported bulk calcium signals that are generally correlated with these behaviors,[49,112,114,116,117] microendoscopic single-cell calcium imaging studies have revealed relatively few neurons tuned or time-locked to specific social behaviors.[27,37,46,52]

Approaches incorporating statistical modeling of neural dynamics into the analysis of population activity in VMHvl have inferred latent dimensions with properties of ramping and stability that correlate with the computations of scalability and persistence that have been suggested to be cardinal features of many internal states.[4,20] One study inferred a latent slow dimension whose activity resembles that of a neural integrator that accumulates some input variable during aggression.[28] Activity along this dimension ramps up as an animal starts sniffing a male intruder and peaks as animals engage in attack behavior (Figure 3C, right), suggesting that it may represent an escalating internal state. Notably, activity remains persistent during the inter-bout intervals of attack, suggesting that the dimension resembles a state of aggressiveness rather than a representation of attack motor behavior. Importantly, by dissecting the fit dynamical system through examination of the fit parameters (e.g., eigendecomposition of the dynamics matrix *A*; see Figure 1), the slow dimension was suggestive of an approximate line attractor or leaky integrator, with the graded nature of the aggressive state corresponding to movement along the set of fixed points constituting the attractor (Figure 3D). Individual neurons display ramping activity as the population trajectory progresses along the basin of the line attractor, similar to that observed in other integrator networks.[119] This progression along the attractor correlates with escalating aggressive behavior, from sniffing to dominance mounting (a low-intensity aggressive behavior[46]) and ultimately attack.

A qualitatively similar slow ramping dimension was also inferred recently from activity in female VMHvl$^{Esr1+,\ Nyp2r-}$ neurons, which causally control sexual receptivity,[114] during mating

(C) Illustration of the data-driven, probabilistic state space modeling approach to understand neural computation. This framework uses dimensionality reduction to obtain latent states underlying neural data and fits a dynamical system model to the latent states. Analysis of the fit dynamics parameters—e.g., the linear dynamics matrices of an rSLDS—can reveal novel computations in the circuit under study.

Box 2. Discovering dynamical properties of neural circuits through data-driven modeling

Classical approaches to modeling neural circuits begin with a defined hypothesis about the circuit's properties, such as a specific connectivity pattern or other mechanistic implementation features.[85,86] This "theory-driven" approach attempts to recreate specific properties of neural activity observed in neural circuits (Figure 2A). For instance, numerous theoretical mechanistic models of the oculomotor integrator (which encodes eye position[62]) have been constructed to explain persistent activity seen in neurons from *in vivo* data.[69,70] In the study of internal states, mechanistic models have been used to explain persistent activity in VMHdm neurons involved in maintaining a defensive state.[85] In each system, different models were constructed that incorporated various ad hoc and various mechanistic features, such as recurrent connectivity, cell-intrinsic persistent activity, or neuromodulator release. These models can then be compared based on how faithfully they match actual neural measurements, with greater similarity suggesting that such models have proven sufficient to replicate observed dynamic phenomena. In this approach, the observed data are used as a point of reference rather than serving to directly constrain model construction.

Advances in machine learning over the past several years have enabled a different type of modeling framework that uses statistical methods to estimate the governing equations underlying data.[79] These "data-driven" approaches can fit dynamical systems directly to either behavioral or neural data. Task-trained networks comprise an important category of data-driven models where RNNs are trained to reproduce tasks performed by rodent or human agents. These models may be dissected to reveal hypothesis regarding neural computations, which may then be compared separately to neural data.[18,87] However, given the focus of this perspective on innate, naturalistic behavior, we refer the reader to other prominent reviews of task-trained networks for learned, artificial behaviors.[87–90]

Data-driven techniques can also be used to fit dynamical models to neural data. The fitting process adjusts model parameters to minimize the difference between a model's predicted neural trajectory and the actual trajectories in the data, or more precisely, to maximize the probability of the observed data under the model. Some data-driven approaches, such as partial in network training (PINning),[91] attempt to fit nonlinear RNNs of observed neural data in a 1-1 manner, providing predictions of underlying connectivity (Figure 2B). For instance, PINning has been used to create data-driven models of neural activity during decision-making tasks in the posterior parietal cortex (PPC).[91] By analyzing the fit connectivity matrix, the authors discover a new computational hypothesis for sequential activity in PPC that uses recurrent connectivity and external inputs. This capacity to discover new hypothesis of neural computations through analysis of the fit model parameters has evoked considerable interest in this data-driven approach.[79]

One class of data-driven approaches uses probabilistic state space models to provide a more abstract characterization of neural dynamics[92–107] (Figure 2C). These methods, including latent factor analysis via dynamical systems (LFADS)[82] and recurrent switching linear dynamical systems (rSLDSs)[83] among many recent methods,[92,97,100,101,103] use dimensionality reduction to find a set of latent dimensions/modes, which capture most of the variance in the original recordings, and simultaneously estimate dynamical equations that govern how these low-dimensional variables evolve in time.[83] In addition to discovering latent dimensions using dimensionality reduction, this approach can identify new candidate neural computations and functions performed by those latent modes through examination of the parameters of the learned dynamical models. These methods have been recently used to uncover integration signals in the hypothalamus that function as continuous attractors encoding aggressive and reproductive internal states,[27,28] as well as reward history the habenula.[81]

Methods using rSLDS models typically have fewer parameters than RNNs and are thus well suited for single-trial neural data such as those usually obtained during studies of innate behavior. Moreover, the piecewise linear nature of the rSLDS model allows for a comprehensive understanding of the fit model using methods from linear dynamical systems theory (see Box 1). However, unlike RNNs fit directly to neural data, probabilistic state space models like rSLDS do not have an immediate mechanistic interpretation. To understand how the latent states and dynamics of a model map onto cell types, synaptic connections, neuromodulators, and other biological mechanisms requires reverse engineering the fitted model[18,28,30,108,109] and, often, further experimental validation. Bridging the gap between state space models and more mechanistic circuit models is an active area of research.[60,110,111]

behavior[27] (Figure 3E). This study used similar statistical linear dynamical system models whose fit parameters suggested a leaky integrator or continuous line attractor, where movement along the basin of the attractor integrates male contact to create a continuous representation of sexual arousal or receptivity (Figure 3F). The integrated value along the attractor correlated strongly with the amount of receptive behavior displayed by females in a given mating trial.[27] Individual neurons displayed sequential activity and some ramping as the population trajectory progressed along the basin of the line attractor, correlating with increasing sexual behavior. Furthermore, this line attractor disappeared and then reappeared in a periodic, estrogen-dependent manner across the estrous cycle, with single neurons contributing to the attractor undergoing a transformation in their dynamics from persistent to transient (Figures 5G–5I). Importantly, in both studies, activity along the integration dimension was distinct from that of the population mean of all recorded neurons, underscoring the importance of dynamical methods to reveal neural signals with distinct, behaviorally relevant dynamics.

These studies suggest that continuous attractor dynamics might be a type of computation used by certain subcortical nuclei to encode key features of internal states underlying innate social behaviors, namely their persistence and scalability. However, further work remains to elucidate the nature of the external input variable that is integrated by the attractor. For example, while the receptivity-encoding continuous attractor was suggested to

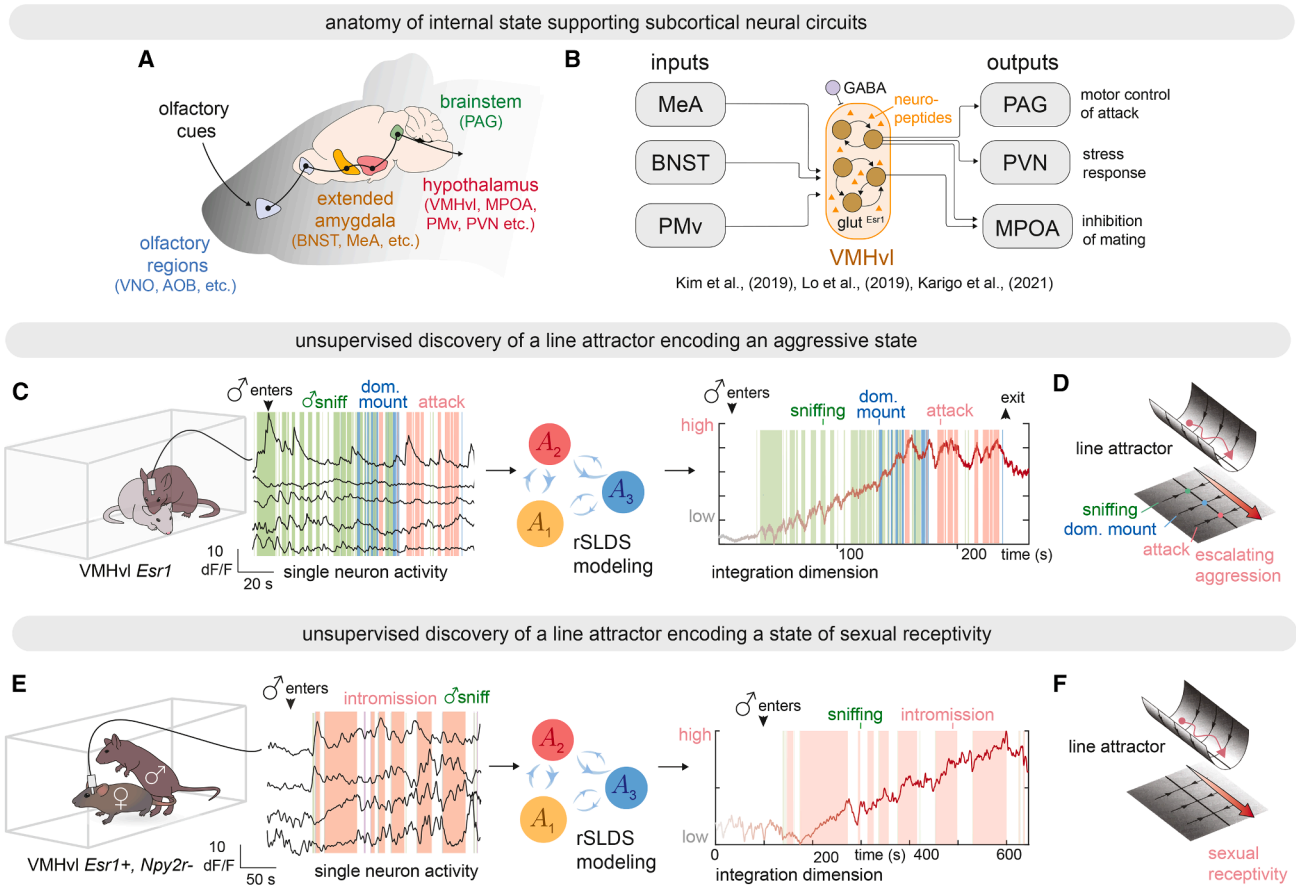


Figure 3. Neural circuitry and dynamics underlying internal state encoding

(A) Overview of subcortical circuits supporting innate behaviors and internal states such as aggression and mating.

(B) Overview of inputs and outputs to VMHvl^{Esr1} neurons.[118]

(C) Summary of results from Nair et al.[28] Unsupervised dynamical systems analysis is used to discover a dimension of activity termed the integration dimension in male VMHvl. This dimension possesses properties of ramping activity and persistence.

(D) This modeling approach discovers a line attractor in male VMHvl that correlates with an escalating aggressive state.

(E) Summary of results from Liu et al.[27] Unsupervised dynamical analysis uncovers a similar integration dimension during mating behavior in female mice with similar properties of ramping and persistence.

(F) Dynamical analysis discovers a line attractor in female VMHvl that correlates with escalating sexual receptivity.

integrate male contact,[27] inputs to the circuit that provide such a contact signal have not yet been characterized (although a recent study[120] identified such an input that controls male mounting). Independent studies in the habenula, another subcortical structure, have provided evidence of a line attractor that integrates reward history during performance of a reinforcement learning task.[81] Furthermore, the state output variable that is represented by VMHvl line attractor activity is not clear, and it might reflect arousal, motivational drive,[19] or both. Hence, elucidating the downstream behavioral functions of state-encoding continuous attractors represents an important area of future research.

It is important to note that while these data suggest a role for line attractors in the encoding of internal social behavioral states in VMHvl, other hypothalamic nuclei that causally control social behaviors do not appear to employ this mechanism. The MPOA controls several innate social behaviors, including parenting[45,121] and male mating.[46,47] Fitting of rSLDS models to data recorded from MPOA^{Esr1} neurons during male mating did not reveal a latent slow mode or line attractor dynamics; rather, mating behaviors were represented by specific, sequentially activated action-tuned populations such as mounting and intromission.[28] Nevertheless, a female-tuned subpopulation of VMHvl^{Esr1} neurons[37] in male mice was found to exhibit line attractor dynamics during mating.[28] These findings suggest that different hypothalamic nuclei may encode the same social behavior using different neural dynamics. The extent to which this generalizes to other hypothalamic nuclei and other behaviors is an interesting question for further studies.

REVEALING THE CAUSAL IMPLEMENTATION OF STATE-ENCODING CONTINUOUS ATTRACTORS

Understanding whether and how continuous attractor dynamics are implemented in subcortical nuclei requires first distinguishing whether the observed dynamics are computed locally in the nucleus under study or rather constitute a passive

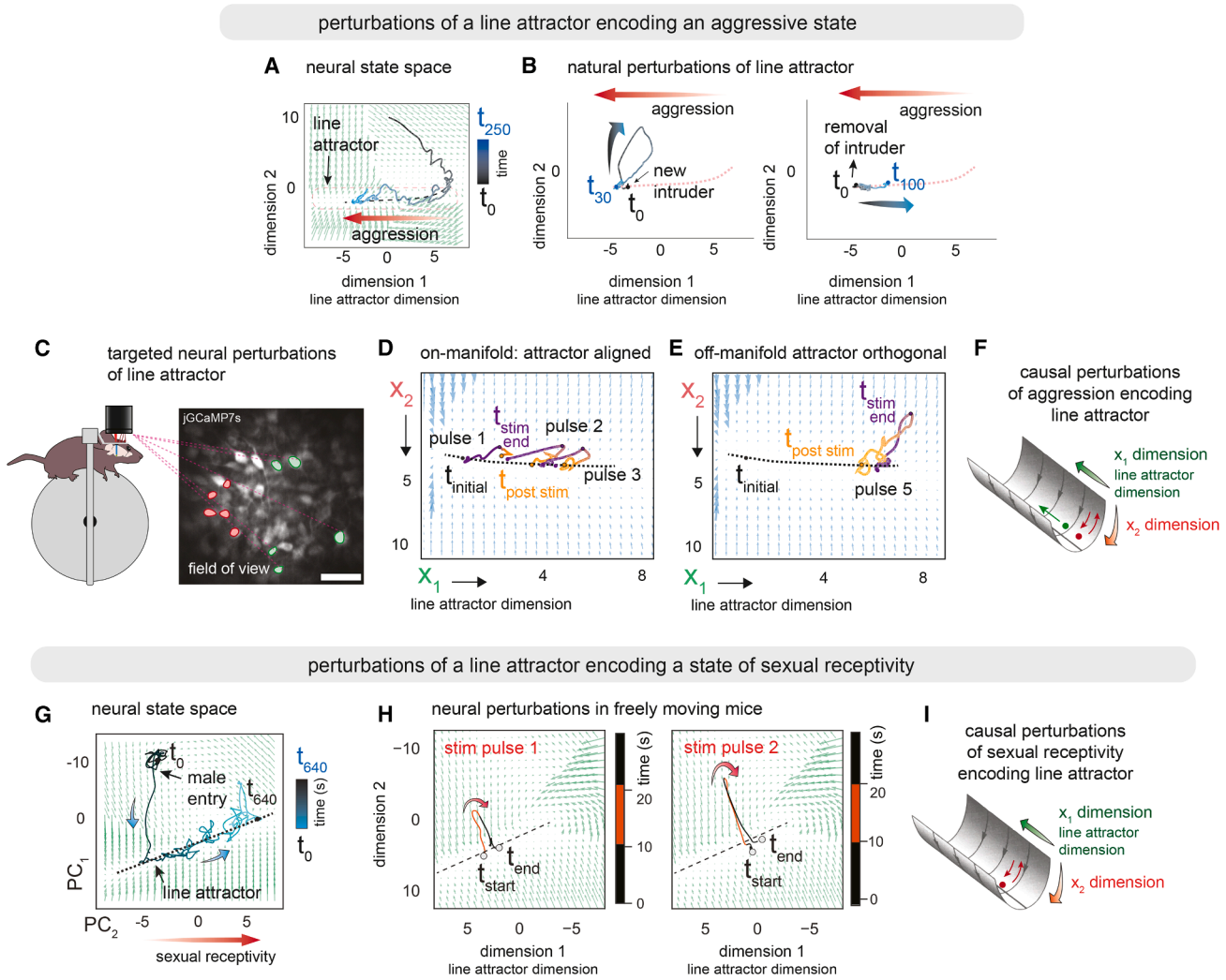


Figure 4. Causal perturbations of affective state-encoding line attractor dynamics

(A and B) Neural state space showing line attractor dynamics during aggression (A) and behavioral perturbations thereof (Nair et al.[28]) (B). Left: introduction of a male intruder results in a transient off-manifold response, with the neural trajectory returning to the line attractor as the resident interacts with the new intruder. Right: removal of all intruders from the cage results in a slow movement of the neural trajectory along the line attractor, as predicted by the time constant of the fit dynamical system model.

(C) Schematic of experiments performed in Vinograd et al.[30] Targeted perturbations are performed toward either attractor-aligned (x_1) or orthogonal neurons (x_2).

(D) Activation of attractor-aligned (x_1) neurons results in an on-manifold movement of the neural trajectory in neural state space. Each pulse of activation is integrated by the circuit, allowing the trajectory to move to the end of the line.

(E) Activation of attractor-orthogonal neurons results in an off-manifold response where the neural trajectory is perturbed in a transient manner orthogonal to the line. Activity returns to the line at the end of the stimulation period.

(F) Cartoon summarizing results in Vinograd et al.[30] showing effects of on- and off-manifold perturbation in neural state space.

(G) Neural state space showing line attractor dynamics in Liu et al.[27]

(H) Neural perturbations, in conjunction with imaging in freely moving animals, result in transient off-manifold movements of the neural trajectory in state space.

(I) Cartoon summarizing results in Liu et al.[27] showing effects of off-manifold perturbation in neural state space.

readout of a signal computed in an upstream attractor network, i.e., are "inherited" from an input. Furthermore, given that sources of persistent, time-varying sensory stimuli might be present in the environment,[122,123] the apparent continuous attractor observed in the hypothalamus could simply reflect ramping sensory input rather than true attractor dynamics. Arbitrating among these possibilities necessitates perturbation experiments to rigorously test predictions of a continuous attractor model.

In principle, such perturbations can be performed at two levels: manipulations of behavior or of circuit activity. Interruption of behaviorally relevant sensory input can be used to test whether persistent activity in attractor circuits is intrinsic to the observed area.[25] Examination of neural dynamics in VMHvl^{Esr1} neurons upon manual removal of the male intruder following attack by the resident mouse yielded persistent activity and slow decay within the integration dimension in the imaged resident, consistent with intrinsic attractor dynamics[28] (Figures 4A

and 4B). A more sophisticated "hands-off" behavioral perturbation was applied to test the intrinsic nature of the receptivity-encoding continuous attractor in female VMHvl$^{Esr1+,Npy2r-}$ neurons by using wireless optogenetics to induce abrupt male mounting pauses during an ongoing mating interaction. Analysis of activity in the integration dimension of the female indicated slow decay, with activity remaining persistently elevated and not collapsing at the offset of male mounting.[27] In some females, continued ramping activity was observed, suggestive of an intrinsic source of ramping in this system.[27]

A rigorous test of the intrinsic nature of attractor circuits, however, requires direct experimental perturbations of neural activity in conjunction with neural recordings. These perturbations may be "on-manifold" if they are targeted to neurons contributing to the attractor and "off-manifold" if they target other attractor-orthogonal neurons in the dimensionally reduced neural state space.[124] While on-manifold perturbations are crucial to reveal conclusive evidence of integration in continuous attractor networks, off-manifold perturbations are critical to illustrate the attractive nature of fixed points along the attractor manifold. Until recently, the only examples of *in vivo* neural perturbation of attractor dynamics were performed in a head direction-encoding continuous ring attractor in *Drosophila*[59,125] and in a murine point attractor encoding working memory in cortical area ALM.[67,126] In general, such experiments are challenging to perform in the context of innate social behaviors such as aggression because single-cell-targeted two-photon optical manipulation and imaging[127,128] require head fixation, and head-fixed mice cannot fight.

A recent study utilized the surprising observation that VMHvl$^{PR/Esr1}$ neurons appear to vicariously mimic ("mirror") aggression-related activity during passive observation of attack episodes performed by "demonstrator" mice.[117] rSLDS modeling of VMHvl^{Esr1} neural activity imaged in freely behaving mice yielded line attractor-like activity during both observation of and engagement in attack, with a ~70% overlap in attractor-weighted neurons under the two conditions.[30] This allowed instantiation of aggression-related continuous attractor dynamics in the VMHvl of head-fixed mice passively observing aggression[30] (Figure 4C). On-manifold two-photon holographic optogenetic activation combined with calcium imaging[127] demonstrated the capacity of Esr1$^+$ neurons to integrate direct input to various fixed points, while off-manifold activation demonstrated the attractive nature of those inferred fixed points (Figures 4D–4F). Neural perturbations in freely moving animals have been used to test the intrinsic nature of the female receptivity-encoding continuous attractor.[27] Off-manifold perturbations using generic optogenetic inactivation demonstrated the attractive nature of fixed points inferred from the data-driven models used in the study (Figures 4G–4I).

These studies demonstrate the intrinsic capacity of hypothalamic circuits to exhibit continuous line attractor-like dynamics that encode putative representations of internal states. Mechanistically, such intrinsic dynamics may emerge through either local (intra-nuclear) or long-range (inter-nuclear) recurrent connectivity[123,129] and may also involve cell-intrinsic mechanisms and inhibitory connections.[130,131] Given the densely interconnected nature of hypothalamic circuits,[118,121,132] further investigation is needed to understand how attractor dynamics are implemented within and across state-contributing nuclei.

LINKING INTERNAL STATE-ENCODING CONTINUOUS ATTRACTORS TO CIRCUIT MECHANISMS

The ability of dynamical systems models to reveal representations of state-encoding continuous attractors has relied on statistical approaches that are agnostic to implementation-level details of these emergent dynamics. The demonstration that these dynamics are intrinsic to the observed hypothalamic cell populations was crucial for showing that these signals are not simply graded inputs persistently present in the environment. However, significant open questions remain regarding the circuit-level implementations of these putative line attractor dynamics. More broadly, the mechanisms of how continuous attractor dynamics emerge from the neural circuits that embody them are yet to be fully understood, raising important questions about how population-manifold approaches can be linked to a circuit-level understanding of neural coding.[24] While advances toward this goal have been made for the head direction-computing ring attractor[59,66,125,133,134] in *Drosophila*, where highly specific genetic access to neural cell types[135,136] and an electron microscopy (EM) connectome[137,138] are available, such resources have been more limited in vertebrate systems, although progress is being made.[139]

Despite these limitations, studies on hypothalamic line attractors have begun to address questions of neural implementation. Evidence for functional connectivity among attractor-contributing neurons within the aggression-encoding continuous attractor in VMHvl^{Esr1} neurons has been provided by single-cell optogenetic stimulation and imaging experiments.[30] Interestingly, this functional connectivity appears to be specific to the attractor-contributing ensemble, as attractor-orthogonal neurons did not exhibit similar connectivity (Figures 5A–5C). Computational models suggest that this connectivity is not exclusively implemented through traditional glutamatergic connectivity, as theorized for cortico-hippocampal and oculomotor attractor networks,[70] but likely includes slower-timescale neurotransmission.[30] This finding implies a potential role for neuromodulatory signaling, potentially mediated by neuropeptides or biogenic amines, released either locally or through meso-scale inter-region connectivity.

Targeted CRISPR-based knockdown of neuropeptide receptor signaling in VMHvl^{Esr1} neurons[29] has recently provided evidence supporting the role of peptide signaling in hypothalamic line attractor dynamics (Figure 5D). A recent study reported that the genetic disruption of oxytocin (OXT) and vasopressin (AVP) receptor signaling in VMHvl^{Esr1} neurons resulted in the loss of the aggression-encoding continuous attractor, providing causal evidence of a requirement for peptide signaling to implement a hypothalamic line attractor.[29] Dynamical systems analysis performed on calcium imaging data from *OxtR/Avpr1a* gene-disrupted mice suggested that the relevant latent dimension no longer possessed the property of integration but rather showed activity time-locked to each bout of male interaction (Figure 5E, lower). As a result, the genetically perturbed system possessed a point attractor rather than a line attractor, where

CellPress
OPEN ACCESS

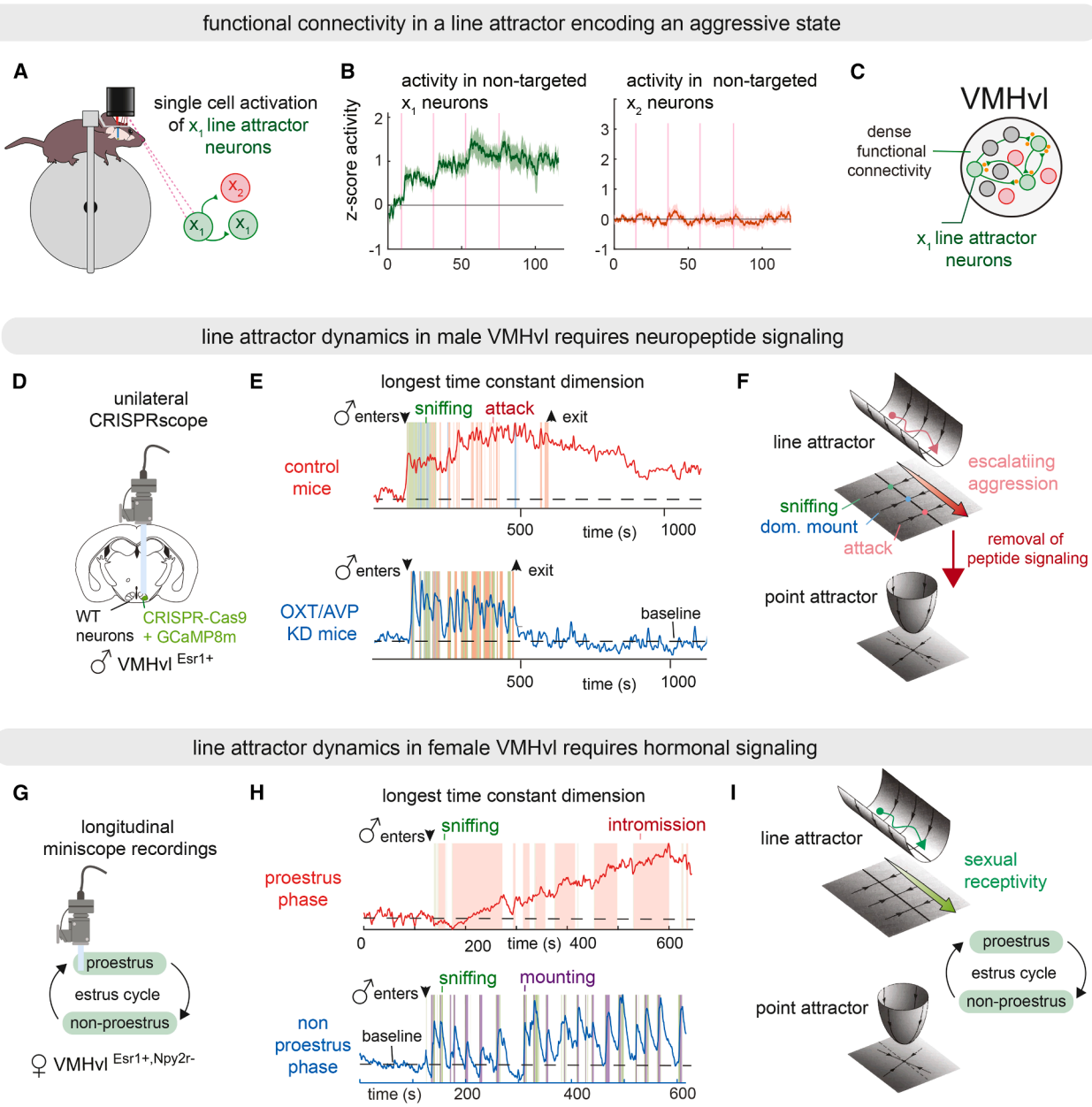


Figure 5. Mechanisms underlying affective state-encoding line attractor dynamics

(A) Schematic of experiments in Vinograd et al.[30]: single-line attractor-contributing neurons (x_1) are activated, and untargeted x_1 and attractor-orthogonal (x_2) neurons are examined.

(B) Left: activity of untargeted x_1 neurons showing functional connectivity within the line attractor-contributing ensemble. Right: activity of untargeted x_2 neurons showing lack of functional connectivity between x_1 and x_2 neurons.

(C) Summary of results in Vinograd et al.,[30] showing selective functional connectivity within the line attractor ensemble.

(D) Schematic of experiments in Mountoufaris et al.[29]: miniscope recordings are performed in animals where oxytocin (*OxtR*) and vasopressin (*Avpr1*) receptors in VMHvl are knocked down in Esr1 neurons using CRISPR unilaterally. This design allows researchers to study the effect of the knockdown (KD) on neural activity, independent of its effect on behavior.

(E) Longest time constant dimension from dynamical system modeling in control (top) and *OxtR*/*Avpr1* KD (bottom) mice. This dimension in control mice shows a long time constant with persistent activity that decays slowly in the post-intruder period. In the KD mice, however, this dimension possesses a smaller time constant and thus is unable to integrate, resulting in activity time-locked to interaction bouts.

(F) Summary of Mountoufaris et al.[29] showing the loss of line attractor dynamics upon removal of neuropeptide signaling.

(G) Schematic of experiments in Liu et al.[27]: miniscope recordings are performed longitudinally across the estrus cycle in female mice.

(H) Longest time constant dimension in the same mice during the proestrus phase (top) and non-proestrus phase (bottom). This dimension during the proestrus shows a long time constant with slow ramping and persistent activity. During the non-proestrus phase, this dimension possesses a smaller time constant and thus is unable to integrate, resulting in activity time-locked to interaction bouts.

(I) Summary of Liu et al.[27] showing the transformation of line attractor dynamics through the estrus cycle.

the point of stability represents the baseline activity of the nucleus (Figure 5F). The same manipulation also reduced aggression, consistent with (but not proving) a causal role for line attractor dynamics in this behavior.

These results establish a requirement for neuropeptide signaling in establishing or maintaining the hypothalamic line attractor. However, they also raise possible alternative hypotheses to explain the inferred line attractor. For instance, slow hypothalamic dynamics might also be approximated as a single point attractor with a shallow energy landscape along a single dimension. Decay along this dimension might correlate with the decay of neuropeptide dynamics. Thus, additional studies are required to understand the relevance of peptide-based modulation to the putative line attractor dynamics identified in this system. Furthermore, OXT and AVP are not synthesized by VMHvl^{Esr1} neurons[140] but rather by other hypothalamic nuclei that provide input to VMH.[141] Thus, further investigation is necessary to reveal the differential contribution of inter-region and intra-region connectivity, and/or of cell-intrinsic responses, to peptide-mediated continuous attractor dynamics in VMHvl.

The importance of slow neuromodulatory or hormonal signaling to hypothalamic continuous attractors also appears to generalize to other internal states. Recordings of neurons across the estrus cycle indicate that the female receptivity-encoding continuous attractor in VMHvl is lost during non-proestrus states when levels of cycling estrogen and progesterone are low, suggesting a strong modulation by steroidal hormones on a timescale of days[27] (Figures 5G and 5H). Dynamical analysis from mice in the non-proestrus stage of the estrus cycle indicated that the resulting dimensions no longer integrate but possess activity time-locked to periods of interaction with the male (Figure 5H, lower). This suggests that sex-steroid hormones may also play a role in instantiating hypothalamic line attractor dynamics (Figure 5I). In support of this conclusion, ovariectomy eliminated attractor dynamics, while priming of ovariectomized females with estrogen plus progesterone restored it.[27]

These new studies challenge dominant assumptions in two different domains of neuroscience. (1) They suggest that traditional theoretical frameworks for attractor dynamics[70] may need to be expanded to accommodate new classes of mechanisms, such as peptide-mediated attractor dynamics, as shown for hypothalamic continuous attractors. (2) They indicate that the hypothalamus performs more computation than previously recognized, challenging a prevailing view of the region as a "hydraulic" network of behavioral action-specific relay stations utilizing an underlying cell identity code.[6,48] This repositions the hypothalamus as a crucial platform to understand how attractor dynamics can emerge through multi-level mechanisms, including hormonal and neuropeptide signaling as well as recurrent connectivity, to elucidate the encoding of internal states.

TESTING THE IMPLICATIONS OF A DYNAMICAL SYSTEM MODEL: CAUSALITY, IMPLEMENTATION, READOUT, AND PLASTICITY

The finding that continuous attractors may encode internal affective states in the hypothalamus illustrates a new approach to identify and study such states through their unsupervised identification using neural dynamics, which should be generally applicable. At the same time, many important new questions and issues have been raised by these findings. Answering these questions is critical but technically challenging. For example, to advance beyond correlational studies, it is essential to test whether movement along hypothalamic continuous attractors causally influences behavior, using genetically based perturbation techniques. However, conducting such experiments requires genetic access to neurons that contribute to continuous attractor dynamics.[59,125,134] Since VMHvl^{Esr1} attractor neurons form a functional subpopulation,[28] activity-based tagging methods may be best suited for accessing them.[142] Whether these neuronal dynamics correspond neatly to genetic cell types marked by specific genes[140] remains an open and important question, whose answer may also further enable targeted perturbation studies.

The mechanistic implementation of the VMHvl^{Esr1} line attractors at the circuit level remains to be elucidated. Theories of line attractor implementation postulate a role for recurrent glutamatergic connectivity in generating the persistent activity and integration properties necessary for such dynamics.[62,70] While holographic optogenetic perturbations and imaging have provided evidence of functional connectivity among line attractor neurons in VMHvl,[30] whether such connectivity is glutamatergic or locally recurrent is not yet clear. RNN modeling[30,85] has indicated a requirement for slow neuromodulatory signaling in attractor implementation, an inference supported by genetic perturbations.[29] Consistent with this, paired slice recordings in VMH have suggested that local glutamatergic connectivity is sparse and neuropeptide signaling dominant[143]; however, attractor-weighted neurons are rare (<10%), and connections between them may have been missed by random sampling. EM connectomic reconstruction of VMH should resolve the issue of local recurrent connectivity and shed light on attractor implementation, as it has in other systems.[133,139]

Population activity-based studies of neural dynamics often assume that these dynamics are "read out" by downstream targets of the brain region under study: in other words, if we human observers can "see" and interpret attractor dynamics in a certain region of the brain using mathematics, it is likely that the brain itself does so as well. However, this assumption remains to be experimentally validated. Hypothalamic continuous attractors may be "read" by other recurrently connected hypothalamic regions and by pre-motor centers such as the periaqueductal gray (PAG).[116] Understanding these transformations will require simultaneous imaging of multiple hypothalamic regions, along with imaging of activity in putative downstream brain regions. New multi-site imaging techniques[144] will be crucial to achieve this objective.

A consistent feature of data-driven studies of hypothalamic continuous attractor dynamics is that key attractor parameters vary across individuals, even when the animals are genetically identical. For instance, the time constant of the aggression-encoding continuous attractor has been shown to strongly correlate with the fraction of time a mouse spends performing attack behavior, suggesting that the stability of the inferred continuous attractor may be tuneable[28] (Figures 6A and 6C). Further studies

CellPress
OPEN ACCESS

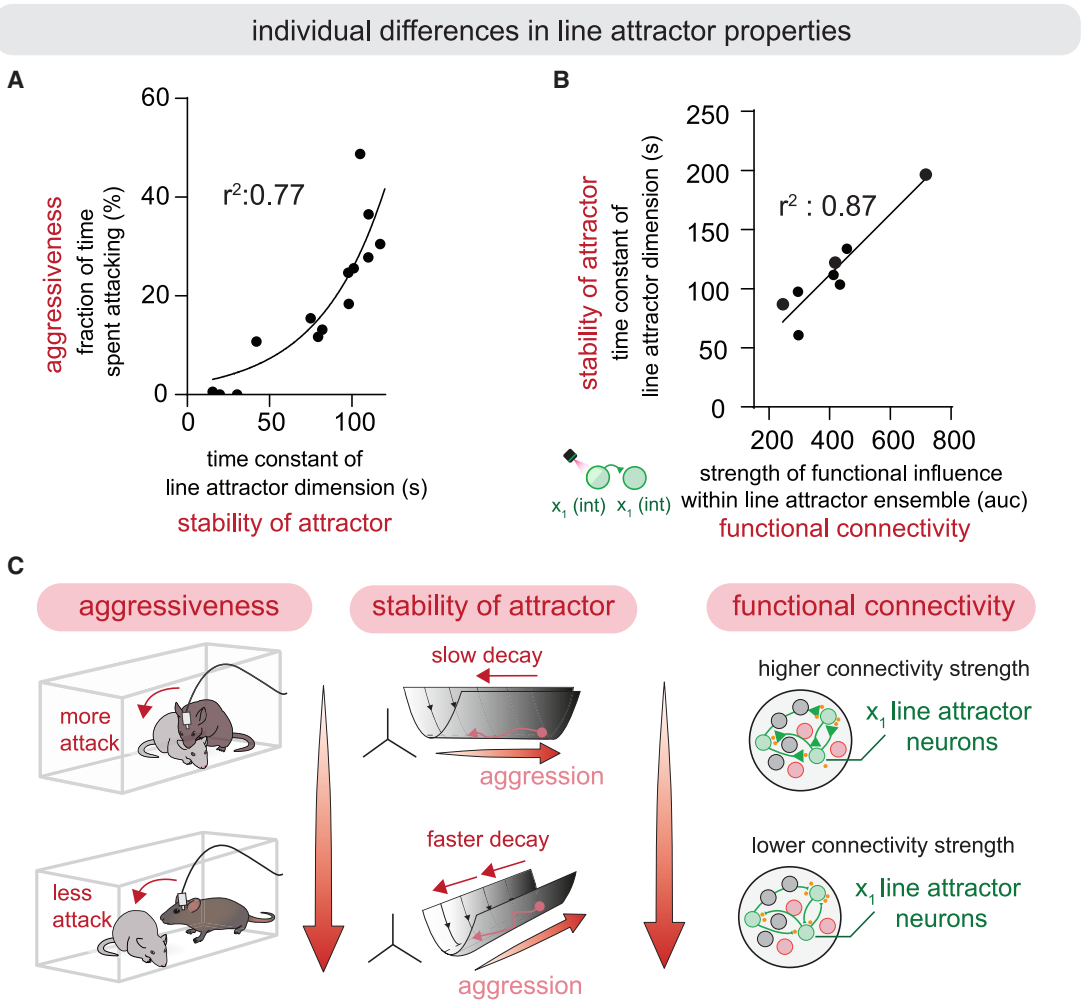


Figure 6. Individual differences in line attractor dynamics underlying aggression

(A) Results from Nair[28] showing a correlation between the stability of the line attractor (time constant of the line attractor dimension) and aggressiveness (fraction of time spent attacking) in individual mice.

(B) Results from Vinograd, Nair, et al.[30] showing a correlation between the stability of the line attractor and strength of functional connectivity within the line attractor-contributing ensemble (level of activity evoked in untargeted line attractor neurons upon activation of a single neuron) across individual mice.

(C) Summary of individual differences and line attractor properties in Nair et al.[28] and Vinograd, Nair, et al.[30] The line attractor cartoons are an idealization to facilitate the reader's intuition, and caveats are discussed in the section entitled "limitations of the dynamical systems perspective."

have demonstrated that the time constants of a similar attractor in head-fixed mice are positively correlated with the degree of functional connectivity within the attractor-weighted ensemble[30] (Figures 6B and 6C). Whether this inter-individual variation reflects a malleability of attractor dynamics within an individual, and the influences responsible for such plasticity, remains to be determined.

POTENTIAL TRANSLATIONAL RELEVANCE OF THE DYNAMICAL SYSTEM PERSPECTIVE

The finding that individual behavioral differences are correlated with differences in attractor properties raises the question of whether such properties may be altered in certain brain disorders and, if so, whether drug-based treatments might be used to reverse such changes.

Several neuropsychiatric disorders, such as major depressive disorder (MDD), feature persistent changes in behavior,[145] and it has been suggested that impaired attractor dynamics might explain aspects of disease etiology or pathophysiology.[146–148] Irrespective of their causal role in mental illness, altered attractor dynamics suggested by unsupervised analysis of neural data might afford more reliable biomarkers for psychiatric disorders with variable behavioral expression. Disease-relevant alterations in neural dynamics might also manifest as changes to neural manifold structure beyond attractor dynamics, such as in manifold geometric properties.[23,149] Indeed, recent studies that begin to apply dynamical systems analysis to human fMRI neural datasets have identified changes in neural manifold properties that correlate with learning deficiencies in adolescent populations.[150]

Thus, dynamical analysis of neural data may help to link studies of the neural basis of mental health disorders in humans

and animal models, where the behavioral phenotypes may not be homologous. They raise the possibility of using such dynamical models as digital twins, which could be used to identify disorder-specific changes in neural dynamics from human datasets (such as from fMRI) and ultimately make predictions of the contributions of mechanisms, such as excitability, functional connectivity, or neuropeptide expression, to disease etiology, which may be tested in mouse models.[151]

These exciting possibilities underscore the potentially transformative impact of understanding continuous attractor dynamics underlying internal affective states, not only for advancing basic neuroscience but also for developing targeted interventions in clinical settings.[146,147,152]

LIMITATIONS OF THE DYNAMICAL SYSTEM PERSPECTIVE

While the dynamical system perspective has the potential to reveal hypotheses for new computations underlying affective internal states in a data-driven manner, it is unlikely to be sufficient on its own to understand the full complexity of emotional experiences that accompany affective states. The inability to definitely measure subjective experience by verbal report in mice, together with a lack of behavioral assays to measure affect-related task variables,[9] limits the ability to directly relate latent manifold dynamics (or any other neural correlates of emotional behaviors identified in animal models[153,154]) to emotional feelings and mental states in humans.[155] Nevertheless, advances in recording human brain activity at the level of single units[156,157] may make it possible to search for persistent activity and latent variables associated with subjective emotional feelings in humans. Beyond persistence and scalability other emotion primitives, such as valence and generalization,[4,20] might not easily map onto latent dynamics suggested by dynamical systems modeling and would require new supervised forms of neural data analysis.

Finally, from a theoretical perspective, an important consideration is that biologically instantiated attractors are approximations of purely mathematically defined entities. Continuous attractors that have been uncovered in the literature, including the internal state-encoding line attractors described in this perspective and others such as the *Drosophila* ring attractor encoding head direction[66,134] or the toroidal attractor encoding allocentric position in rodents,[80] are approximations that are constrained by biological detail such as the size of the network, robustness to noise and perturbation, and other implementation details.[158,159] While theoretical studies have shown that these approximations function as "pure" attractors within a biologically relevant timescale,[158,160] we wish to emphasize that biological attractor-like systems may indeed be a distinct class of models that need to be understood regardless of their relation to their mathematical counterparts.

CONCLUSIONS

Advances in statistics, machine learning, and new data acquisition methods have facilitated data-driven discoveries of neural dynamics across a wide range of paradigms. When applied to neural data acquired during innate behaviors, these tools can uncover an encoding of internal states implemented as continuous attractors, with key features such as persistence and scalability reflected in the emergent properties of attractor dynamics. While future work is necessary to address critical questions of behavioral causality, implementation, and readout, the paradigm holds promise for linking the fields of manifold-level and circuit-level neuroscience,[24] a linkage that is becoming indispensable to understanding neural computations underlying internal states. We hope and anticipate that the ideas outlined in this perspective will inspire new research at the intersection of machine learning and observational and perturbational neuroscience in the years to come.

ACKNOWLEDGMENTS

We thank A. Kennedy, M.J. Schnitzer, and S. Ganguli for their seminal contributions to this project; T. Karigo and B. Yang for providing neural data; N. Coria for software engineering; members of the Anderson laboratory for helpful discussions; C. Chiu for laboratory management; and L. Chavarria for administrative assistance. D.J.A. is an investigator of the Howard Hughes Medical Institute. This work was supported in part by US NIH grants R01NS123916, R01MH123612, R01MH112593, and MH070053 to D.J.A. and by the Simons Collaboration on the Global Brain. A.N. is supported by the Nanyang Assistant Professorship Startup Grant by Nanyang Technological University, Singapore.

DECLARATION OF INTERESTS

D.J.A. is a member of the *Neuron* editorial board.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.neuron.2025.11.003.

REFERENCES

1. Berridge, K.C. (2004). Motivation concepts in behavioral neuroscience. Physiol. Behav. *81*, 179–209. https://doi.org/10.1016/j.physbeh.2004.02.004.

2. Flavell, S.W., Gogolla, N., Lovett-Barron, M., and Zelikowsky, M. (2022). The emergence and influence of internal states. Neuron *110*, 2545–2570. https://doi.org/10.1016/j.neuron.2022.04.030.

3. Anderson, D.J. (2016). Circuit modules linking internal states and social behaviour in flies and mice. Nat. Rev. Neurosci. *17*, 692–704. https://doi.org/10.1038/nrn.2016.125.

4. Anderson, D.J., and Adolphs, R. (2014). A framework for studying emotions across species. Cell *157*, 187–200. https://doi.org/10.1016/j.cell.2014.03.003.

5. Lorenz, K. (1966). On Aggression (Harcourt, Brace & World).

6. Ishii, K.K., Osakada, T., Mori, H., Miyasaka, N., Yoshihara, Y., Miyamichi, K., and Touhara, K. (2017). A Labeled-Line Neural Circuit for Pheromone-Mediated Sexual Behaviors in Mice. Neuron *95*, 123–137.e8. https://doi.org/10.1016/j.neuron.2017.05.038.

7. Chamero, P., Marton, T.F., Logan, D.W., Flanagan, K., Cruz, J.R., Saghatelian, A., Cravatt, B.F., and Stowers, L. (2007). Identification of protein pheromones that promote aggressive behaviour. Nature *450*, 899–902. https://doi.org/10.1038/nature05997.

8. Zelikowsky, M., Hui, M., Karigo, T., Choe, A., Yang, B., Blanco, M.R., Beadle, K., Gradinaru, V., Deverman, B.E., and Anderson, D.J. (2018). The Neuropeptide Tac2 Controls a Distributed Brain State Induced by Chronic Social Isolation Stress. Cell *173*, 1265–1279.e19. https://doi.org/10.1016/j.cell.2018.03.037.

9. LeDoux, J. (2012). Rethinking the emotional brain. Neuron *73*, 653–676. https://doi.org/10.1016/j.neuron.2012.02.004.

10. Adolphs, R. (2013). The biology of fear. Curr. Biol. *23*, R79–R93. https://doi.org/10.1016/j.cub.2012.11.055.

11. Cheung, K.Y.M., Nair, A., Li, L.Y., Shapiro, M.G., and Anderson, D.J. (2025). Population coding of predator imminence in the hypothalamus. Neuron *113*, 1259–1275.e4. https://doi.org/10.1016/j.neuron.2025.02.003.

12. Grewe, B.F., Gründemann, J., Kitch, L.J., Lecoq, J.A., Parker, J.G., Marshall, J.D., Larkin, M.C., Jercog, P.E., Grenier, F., Li, J.Z., et al. (2017). Neural ensemble dynamics underlying a long-term associative memory. Nature *543*, 670–675. https://doi.org/10.1038/nature21682.

13. Allen, W.E., Chen, M.Z., Pichamoorthy, N., Tien, R.H., Pachitariu, M., Luo, L., and Deisseroth, K. (2019). Thirst regulates motivated behavior through modulation of brainwide neural population dynamics. Science *364*, 253. https://doi.org/10.1126/science.aav3932.

14. Allen, W.E., DeNardo, L.A., Chen, M.Z., Liu, C.D., Loh, K.M., Fenno, L.E., Ramakrishnan, C., Deisseroth, K., and Luo, L. (2017). Thirst-associated preoptic neurons encode an aversive motivational drive. Science *357*, 1149–1155. https://doi.org/10.1126/science.aan6747.

15. Richman, E.B., Ticea, N., Allen, W.E., Deisseroth, K., and Luo, L. (2023). Neural landscape diffusion resolves conflicts between needs across time. Nature *623*, 571–579. https://doi.org/10.1038/s41586-023-06715-z.

16. Mathis, M.W., Perez Rotondo, A., Chang, E.F., Tolias, A.S., and Mathis, A. (2024). Decoding the brain: From neural representations to mechanistic models. Cell *187*, 5814–5832. https://doi.org/10.1016/j.cell.2024.08.051.

17. Churchland, M.M., Cunningham, J.P., Kaufman, M.T., Foster, J.D., Nuyujukian, P., Ryu, S.I., and Shenoy, K.V. (2012). Neural population dynamics during reaching. Nature *487*, 51–56. https://doi.org/10.1038/nature11129.

18. Mante, V., Sussillo, D., Shenoy, K.V., and Newsome, W.T. (2013). Context-dependent computation by recurrent dynamics in prefrontal cortex. Nature *503*, 78–84. https://doi.org/10.1038/nature12742.

19. Falkner, A.L., Grosenick, L., Davidson, T.J., Deisseroth, K., and Lin, D. (2016). Hypothalamic control of male aggression-seeking behavior. Nat. Neurosci. *19*, 596–604. https://doi.org/10.1038/nn.4264.

20. Adolphs, R., and Anderson, D.J. (2018). The Neuroscience of Emotion: A New Synthesis (Princeton University Press). https://doi.org/10.23943/9781400889914.

21. Russell, J.A. (1980). A circumplex model of affect. J. Pers. Soc. Psychol. *39*, 1161–1178. https://doi.org/10.1037/h0077714.

22. Glaser, J., Whiteway, M., Cunningham, J.P., Paninski, L., and Linderman, S. (2020). Recurrent switching dynamical systems models for multiple interacting neural populations. Adv. Neural Inf. Process. Syst. *33*, 14867–14878.

23. Perich, M.G., Narain, D., and Gallego, J.A. (2025). A neural manifold view of the brain. Nat. Neurosci. *28*, 1582–1597. https://doi.org/10.1038/s41593-025-02031-z.

24. Langdon, C., Genkin, M., and Engel, T.A. (2023). A unifying perspective on neural manifolds and circuits for cognition. Nat. Rev. Neurosci. *24*, 363–377. https://doi.org/10.1038/s41583-023-00693-x.

25. Khona, M., and Fiete, I.R. (2022). Attractor and integrator networks in the brain. Nat. Rev. Neurosci. *23*, 744–766. https://doi.org/10.1038/s41583-022-00642-0.

26. Miller, P. (2016). Dynamical systems, attractors, and neural circuits. F1000Res *5*. F1000 Faculty Rev-992. https://doi.org/10.12688/f1000research.7698.1.

27. Liu, M., Nair, A., Coria, N., Linderman, S.W., and Anderson, D.J. (2024). Encoding of female mating dynamics by a hypothalamic line attractor. Nature *634*, 901–909. https://doi.org/10.1038/s41586-024-07916-w.

28. Nair, A., Karigo, T., Yang, B., Ganguli, S., Schnitzer, M.J., Linderman, S.W., Anderson, D.J., and Kennedy, A. (2023). An approximate line attractor in the hypothalamus encodes an aggressive state. Cell *186*, 178–193.e15. https://doi.org/10.1016/j.cell.2022.11.027.

29. Mountoufaris, G., Nair, A., Yang, B., Kim, D.W., Vinograd, A., Kim, S., Linderman, S.W., and Anderson, D.J. (2024). A line attractor encoding a persistent internal state requires neuropeptide signaling. Cell *187*, 5998–6015.e18. https://doi.org/10.1016/j.cell.2024.08.015.

30. Vinograd, A., Nair, A., Kim, J.H., Linderman, S.W., and Anderson, D.J. (2024). Causal evidence of a line attractor encoding an affective state. Nature *634*, 910–918. https://doi.org/10.1038/s41586-024-07915-x.

31. Kruk, M.R. (2014). Hypothalamic attack: a wonderful artifact or a useful perspective on escalation and pathology in aggression? A viewpoint. Curr. Top. Behav. Neurosci. *17*, 143–188. https://doi.org/10.1007/7854_2014_313.

32. Lin, D., Boyle, M.P., Dollar, P., Lee, H., Lein, E.S., Perona, P., and Anderson, D.J. (2011). Functional identification of an aggression locus in the mouse hypothalamus. Nature *470*, 221–226. https://doi.org/10.1038/nature09736.

33. Jun, J.J., Steinmetz, N.A., Siegle, J.H., Denman, D.J., Bauza, M., Barbarits, B., Lee, A.K., Anastassiou, C.A., Andrei, A., Aydın, Ç., et al. (2017). Fully integrated silicon probes for high-density recording of neural activity. Nature *551*, 232–236. https://doi.org/10.1038/nature24636.

34. Zhang, Y., Rózsa, M., Liang, Y., Bushey, D., Wei, Z., Zheng, J., Reep, D., Broussard, G.J., Tsang, A., Tsegaye, G., et al. (2023). Fast and sensitive GCaMP calcium indicators for imaging neural populations. Nature *615*, 884–891. https://doi.org/10.1038/s41586-023-05828-9.

35. Ghosh, K.K., Burns, L.D., Cocker, E.D., Nimmerjahn, A., Ziv, Y., Gamal, A.E., and Schnitzer, M.J. (2011). Miniaturized integration of a fluorescence microscope. Nat. Methods *8*, 871–878. https://doi.org/10.1038/nmeth.1694.

36. Ziv, Y., Burns, L.D., Cocker, E.D., Hamel, E.O., Ghosh, K.K., Kitch, L.J., El Gamal, A., and Schnitzer, M.J. (2013). Long-term dynamics of CA1 hippocampal place codes. Nat. Neurosci. *16*, 264–266. https://doi.org/10.1038/nn.3329.

37. Remedios, R., Kennedy, A., Zelikowsky, M., Grewe, B.F., Schnitzer, M.J., and Anderson, D.J. (2017). Social behaviour shapes hypothalamic neural ensemble representations of conspecific sex. Nature *550*, 388–392. https://doi.org/10.1038/nature23885.

38. Resendez, S.L., Jennings, J.H., Ung, R.L., Namboodiri, V.M.K., Zhou, Z.C., Otis, J.M., Nomura, H., McHenry, J.A., Kosyk, O., and Stuber, G.D. (2016). Visualization of cortical, subcortical and deep brain neural circuit dynamics during naturalistic mammalian behavior with head-mounted microscopes and chronically implanted lenses. Nat. Protoc. *11*, 566–597. https://doi.org/10.1038/nprot.2016.021.

39. Li, Y., Mathis, A., Grewe, B.F., Osterhout, J.A., Ahanonu, B., Schnitzer, M.J., Murthy, V.N., and Dulac, C. (2017). Neuronal Representation of Social Information in the Medial Amygdala of Awake Behaving Mice. Cell *171*, 1176–1190.e17. https://doi.org/10.1016/j.cell.2017.10.015.

40. Kruk, M.R., Westphal, K.G., Van Erp, A.M., van Asperen, J., Cave, B.J., Slater, E., de Koning, J., and Haller, J. (1998). The hypothalamus: cross-roads of endocrine and behavioural regulation in grooming and aggression. Neurosci. Biobehav. Rev. *23*, 163–177. https://doi.org/10.1016/S0149-7634(98)00018-9.

41. Siegel, A., Roeling, T.A., Gregg, T.R., and Kruk, M.R. (1999). Neuropharmacology of brain-stimulation-evoked aggression. Neurosci. Biobehav. Rev. *23*, 359–389. https://doi.org/10.1016/S0149-7634(98)00040-2.

42. Kunwar, P.S., Zelikowsky, M., Remedios, R., Cai, H., Yilmaz, M., Meister, M., and Anderson, D.J. (2015). Ventromedial hypothalamic neurons control a defensive emotion state. eLife *4*, e06633. https://doi.org/10.7554/eLife.06633.

43. Wang, L., Chen, I.Z., and Lin, D. (2015). Collateral Pathways from the Ventromedial Hypothalamus Mediate Defensive Behaviors. Neuron *85*, 1344–1358. https://doi.org/10.1016/j.neuron.2014.12.025.

44. Hong, W., Kim, D.W., and Anderson, D.J. (2014). Antagonistic control of social versus repetitive self-grooming behaviors by separable amygdala

neuronal subsets. Cell *158*, 1348–1361. https://doi.org/10.1016/j.cell.2014.07.049.

45. Wu, Z., Autry, A.E., Bergan, J.F., Watabe-Uchida, M., and Dulac, C.G. (2014). Galanin neurons in the medial preoptic area govern parental behaviour. Nature *509*, 325–330. https://doi.org/10.1038/nature13307.

46. Karigo, T., Kennedy, A., Yang, B., Liu, M., Tai, D., Wahle, I.A., and Anderson, D.J. (2021). Distinct hypothalamic control of same- and opposite-sex mounting behaviour in mice. Nature *589*, 258–263. https://doi.org/10.1038/s41586-020-2995-0.

47. Wei, Y.C., Wang, S.R., Jiao, Z.L., Zhang, W., Lin, J.K., Li, X.Y., Li, S.S., Zhang, X., and Xu, X.H. (2018). Medial preoptic area in mice is capable of mediating sexually dimorphic behaviors regardless of gender. Nat. Commun. *9*, 279. https://doi.org/10.1038/s41467-017-02648-0.

48. Moffitt, J.R., Bambah-Mukku, D., Eichhorn, S.W., Vaughn, E., Shekhar, K., Perez, J.D., Rubinstein, N.D., Hao, J., Regev, A., Dulac, C., et al. (2018). Molecular, spatial, and functional single-cell profiling of the hypothalamic preoptic region. Science *362*, eaau5324. https://doi.org/10.1126/science.aau5324.

49. Hashikawa, K., Hashikawa, Y., Tremblay, R., Zhang, J., Feng, J.E., Sabol, A., Piper, W.T., Lee, H., Rudy, B., and Lin, D. (2017). Esr1(+) cells in the ventromedial hypothalamus control female aggression. Nat. Neurosci. *20*, 1580–1590.

50. Yang, T., Yang, C.F., Chizari, M.D., Maheswaranathan, N., Burke, K.J., Borius, M., Inoue, S., Chiang, M.C., Bender, K.J., Ganguli, S., et al. (2017). Social Control of Hypothalamus-Mediated Male Aggression. Neuron *95*, 955–970.e4. https://doi.org/10.1016/j.neuron.2017.06.046.

51. Lee, H., Kim, D.W., Remedios, R., Anthony, T.E., Chang, A., Madisen, L., Zeng, H., and Anderson, D.J. (2014). Scalable control of mounting and attack by Esr1+ neurons in the ventromedial hypothalamus. Nature *509*, 627–632. https://doi.org/10.1038/nature13169.

52. Yang, B., Karigo, T., and Anderson, D.J. (2022). Transformations of neural representations in a social behaviour network. Nature *608*, 741–749. https://doi.org/10.1038/s41586-022-05057-6.

53. Miczek, K.A., de Almeida, R.M.M., Kravitz, E.A., Rissman, E.F., de Boer, S.F., and Raine, A. (2007). Neurobiology of escalated aggression and violence. J. Neurosci. *27*, 11803–11806. https://doi.org/10.1523/JNEUROSCI.3500-07.2007.

54. Strogatz, S.H. (2018). Nonlinear Dynamics and Chaos: with Applications to Physics, Biology, Chemistry, and Engineering (CRC Press).

55. Luenberger, D.G. (1979). Introduction to Dynamic Systems, Theory, Models, and Applications (J. Wiley Sons).

56. Christodoulou, G., and Vogels, T. (2022) The Eigenvalue Value (in Neuroscience).

57. Aoi, M.C., Mante, V., and Pillow, J.W. (2020). Prefrontal cortex exhibits multidimensional dynamic encoding during decision-making. Nat. Neurosci. *23*, 1410–1420. https://doi.org/10.1038/s41593-020-0696-5.

58. Wong, K.F., and Wang, X.J. (2006). A recurrent network mechanism of time integration in perceptual decisions. J. Neurosci. *26*, 1314–1328. https://doi.org/10.1523/JNEUROSCI.3733-05.2006.

59. Kim, S.S., Rouault, H., Druckmann, S., and Jayaraman, V. (2017). Ring attractor dynamics in the Drosophila central brain. Science *356*, 849–853. https://doi.org/10.1126/science.aal4835.

60. Mastrogiuseppe, F., and Ostojic, S. (2018). Linking connectivity, dynamics, and computations in low-rank recurrent neural networks. Neuron *99*, 609–623.e29. https://doi.org/10.1016/j.neuron.2018.07.003.

61. Yang, C.F., Chiang, M.C., Gray, D.C., Prabhakaran, M., Alvarado, M., Juntti, S.A., Unger, E.K., Wells, J.A., and Shah, N.M. (2013). Sexually dimorphic neurons in the ventromedial hypothalamus govern mating in both sexes and aggression in males. Cell *153*, 896–909. https://doi.org/10.1016/j.cell.2013.04.017.

62. Seung, H.S. (1996). How the brain keeps the eyes still. Proc. Natl. Acad. Sci. USA *93*, 13339–13344. https://doi.org/10.1073/pnas.93.23.13339.

63. Inagaki, H.K., Chen, S., Ridder, M.C., Sah, P., Li, N., Yang, Z., Hasanbegovic, H., Gao, Z., Gerfen, C.R., and Svoboda, K. (2022). A midbrain-thalamus-cortex circuit reorganizes cortical dynamics to initiate movement. Cell *185*, 1065–1081.e23. https://doi.org/10.1016/j.cell.2022.02.006.

64. Vyas, S., Golub, M.D., Sussillo, D., and Shenoy, K.V. (2020). Computation Through Neural Population Dynamics. Annu. Rev. Neurosci. *43*, 249–275. https://doi.org/10.1146/annurev-neuro-092619-094115.

65. Remington, E.D., Egger, S.W., Narain, D., Wang, J., and Jazayeri, M. (2018). A Dynamical Systems Perspective on Flexible Motor Timing. Trends Cogn. Sci. *22*, 938–952. https://doi.org/10.1016/j.tics.2018.07.010.

66. Hulse, B.K., and Jayaraman, V. (2020). Mechanisms Underlying the Neural Computation of Head Direction. Annu. Rev. Neurosci. *43*, 31–54. https://doi.org/10.1146/annurev-neuro-072116-031516.

67. Inagaki, H.K., Fontolan, L., Romani, S., and Svoboda, K. (2019). Discrete attractor dynamics underlies persistent activity in the frontal cortex. Nature *566*, 212–217. https://doi.org/10.1038/s41586-019-0919-7.

68. Kennedy, A., Asahina, K., Hoopfer, E., Inagaki, H., Jung, Y., Lee, H., Remedios, R., and Anderson, D.J. (2014). Internal States and Behavioral Decision-Making: Toward an Integration of Emotion and Cognition. Cold Spring Harb. Symp. Quant. Biol. *79*, 199–210. https://doi.org/10.1101/sqb.2014.79.024984.

69. Major, G., and Tank, D. (2004). Persistent neural activity: prevalence and mechanisms. Curr. Opin. Neurobiol. *14*, 675–684. https://doi.org/10.1016/j.conb.2004.10.017.

70. Wang, X.J., and Goldman, M.S. (2009). Neural Integrator Models (Elsevier Limited).

71. Bizzi, E., Mussa-Ivaldi, F.A., and Giszter, S. (1991). Computations underlying the execution of movement: a biological perspective. Science *253*, 287–291. https://doi.org/10.1126/science.1857964.

72. Mussa-Ivaldi, F.A., Giszter, S.F., and Bizzi, E. (1994). Linear combinations of primitives in vertebrate motor control. Proc. Natl. Acad. Sci. USA *91*, 7534–7538. https://doi.org/10.1073/pnas.91.16.7534.

73. Graziano, M.S.A., Taylor, C.S.R., and Moore, T. (2002). Complex movements evoked by microstimulation of precentral cortex. Neuron *34*, 841–851. https://doi.org/10.1016/S0896-6273(02)00698-0.

74. Kleinfeld, D., and Sompolinsky, H. (1988). Associative neural network model for the generation of temporal patterns. Theory and application to central pattern generators. Biophys. J. *54*, 1039–1051. https://doi.org/10.1016/S0006-3495(88)83041-8.

75. Pereira, U., and Brunel, N. (2018). Attractor Dynamics in Networks with Learning Rules Inferred from In Vivo Data. Neuron *99*, 227–238.e4. https://doi.org/10.1016/j.neuron.2018.05.038.

76. Hopfield, J.J. (1982). Neural networks and physical systems with emergent collective computational abilities. Proc. Natl. Acad. Sci. USA *79*, 2554–2558. https://doi.org/10.1073/pnas.79.8.2554.

77. Tsodyks, M. (1999). Attractor neural network models of spatial maps in hippocampus. Hippocampus *9*, 481–489. https://doi.org/10.1002/(SICI)1098-1063(1999)9:4<481::AID-HIPO14>3.0.CO;2-S.

78. Chaudhuri, R., Gerçek, B., Pandey, B., Peyrache, A., and Fiete, I. (2019). The intrinsic attractor manifold and population dynamics of a canonical cognitive circuit across waking and sleep. Nat. Neurosci. *22*, 1512–1520. https://doi.org/10.1038/s41593-019-0460-x.

79. Durstewitz, D., Koppe, G., and Thurm, M.I. (2023). Reconstructing computational system dynamics from neural data with recurrent neural networks. Nat. Rev. Neurosci. *24*, 693–710. https://doi.org/10.1038/s41583-023-00740-7.

80. Gardner, R.J., Hermansen, E., Pachitariu, M., Burak, Y., Baas, N.A., Dunn, B.A., Moser, M.B., and Moser, E.I. (2022). Toroidal topology of population activity in grid cells. Nature *602*, 123–128. https://doi.org/10.1038/s41586-021-04268-7.

81. Sylwestrak, E.L., Jo, Y., Vesuna, S., Wang, X., Holcomb, B., Tien, R.H., Kim, D.K., Fenno, L., Ramakrishnan, C., Allen, W.E., et al. (2022).

Cell-type-specific population dynamics of diverse reward computations. Cell *185*, 3568–3587.e27. https://doi.org/10.1016/j.cell.2022.08.019.

82. Pandarinath, C., O'Shea, D.J., Collins, J., Jozefowicz, R., Stavisky, S.D., Kao, J.C., Trautmann, E.M., Kaufman, M.T., Ryu, S.I., Hochberg, L.R., et al. (2018). Inferring single-trial neural population dynamics using sequential auto-encoders. Nat. Methods *15*, 805–815. https://doi.org/10.1038/s41592-018-0109-9.

83. Linderman, S., Johnson, M., Miller, A., Adams, R., Blei, D., and Paninski, L. (2017) Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, pp. 914–922.

84. Zoltowski, D., Pillow, J., and Linderman, S. (2020). A general recurrent state space framework for modeling neural dynamics during decision-making. In International Conference on Machine Learning, *119*, pp. 11680–11691.

85. Kennedy, A., Kunwar, P.S., Li, L.Y., Stagkourakis, S., Wagenaar, D.A., and Anderson, D.J. (2020). Stimulus-specific hypothalamic encoding of a persistent defensive state. Nature *586*, 730–734. https://doi.org/10.1038/s41586-020-2728-4.

86. Linderman, S.W., and Gershman, S.J. (2017). Using computational theory to constrain statistical models of neural data. Curr. Opin. Neurobiol. *46*, 14–24. https://doi.org/10.1016/j.conb.2017.06.004.

87. Barak, O. (2017). Recurrent neural networks as versatile tools of neuroscience research. Curr. Opin. Neurobiol. *46*, 1–6. https://doi.org/10.1016/j.conb.2017.06.003.

88. Cohen, Y., Engel, T.A., Langdon, C., Lindsay, G.W., Ott, T., Peters, M.A.K., Shine, J.M., Breton-Provencher, V., and Ramaswamy, S. (2022). Recent advances at the interface of neuroscience and artificial neural networks. J. Neurosci. *42*, 8514–8523. https://doi.org/10.1523/JNEUROSCI.1503-22.2022.

89. Lindsay, G.W. (2024). Grounding neuroscience in behavioral changes using artificial neural networks. Curr. Opin. Neurobiol. *84*, 102816. https://doi.org/10.1016/j.conb.2023.102816.

90. Yang, G.R., and Molano-Mazón, M. (2021). Towards the next generation of recurrent network models for cognitive neuroscience. Curr. Opin. Neurobiol. *70*, 182–192. https://doi.org/10.1016/j.conb.2021.10.015.

91. Rajan, K., Harvey, C.D., and Tank, D.W. (2016). Recurrent Network Models of Sequence Generation and Memory. Neuron *90*, 128–142. https://doi.org/10.1016/j.neuron.2016.02.009.

92. Nassar, J., Linderman, S.W., Bugallo, M., and Park, I.M. (2018). Tree-Structured Recurrent Switching Linear Dynamical Systems for Multi-scale Modeling. Preprint at arXiv. https://doi.org/10.48550/arXiv.1811.12386.

93. Kim, T.D., Luo, T.Z., Pillow, J.W., and Brody, C.D. (2021). Inferring Latent Dynamics Underlying Neural Population Activity via Neural Differential Equations. In International Conference on Machine Learning, *139*, pp. 5551–5561.

94. Sani, O.G., Abbaspourazad, H., Wong, Y.T., Pesaran, B., and Shanechi, M.M. (2021). Modeling behaviorally relevant neural dynamics enabled by preferential subspace identification. Nat. Neurosci. *24*, 140–149. https://doi.org/10.1038/s41593-020-00733-0.

95. Buesing, L., Macke, J.H., and Sahani, M. (2012). Learning stable, regularised latent models of neural population dynamics. Network *23*, 24–47. https://doi.org/10.3109/0954898X.2012.677095.

96. Champion, K., Lusch, B., Kutz, J.N., and Brunton, S.L. (2019). Data-driven discovery of coordinates and governing equations. Proc. Natl. Acad. Sci. USA *116*, 22445–22451. https://doi.org/10.1073/pnas.1906995116.

97. Langdon, C., and Engel, T.A. (2022). Latent circuit inference from heterogeneous neural responses during cognitive tasks. Preprint at bioRxiv. https://doi.org/10.1101/2022.01.23.477431.

98. Hu, A., Zoltowski, D., Nair, A., Anderson, D., Duncker, L., and Linderman, S. (2024). Modeling Latent Neural Dynamics with Gaussian Process Switching Linear Dynamical Systems. Preprint at arXiv. https://doi.org/10.48550/arXiv.2408.03330.

99. Mudrik, N., Chen, Y., Yezerets, E., Rozell, C.J., and Charles, A.S. (2024). Decomposed linear dynamical systems (dlds) for learning the latent components of neural dynamics. J. Mach. Learn. Res. *25*, 1–44.

100. Kim, T.D., Luo, T.Z., Can, T., Krishnamurthy, K., Pillow, J.W., and Brody, C.D. (2023). Flow-field inference from neural data using deep recurrent networks. Preprint at bioRxiv. https://doi.org/10.1101/2023.11.14.567136.

101. Marine, S., Ta-Chu, K., Kristopher, T.J., and Guillaume, H. (2022). iLQR-VAE: control-based learning of input-driven dynamics with applications to neural data. International Conference on Learning Representations.

102. Genkin, M., Hughes, O., and Engel, T.A. (2021). Learning non-stationary Langevin dynamics from stochastic observations of latent trajectories. Nat. Commun. *12*, 5986. https://doi.org/10.1038/s41467-021-26202-1.

103. Dowling, M., Zhao, Y., and Park, I.M. (2024). Large-scale variational Gaussian state-space models. Preprint at arXiv. https://doi.org/10.48550/arXiv.2403.01371.

104. Archer, E., Park, I.M., Buesing, L., Cunningham, J., and Paninski, L. (2015). Black box variational inference for state space models. Preprint at arXiv. https://doi.org/10.48550/arXiv.1511.07367.

105. Macke, J.H., Buesing, L., Cunningham, J.P., Yu, B.M., Shenoy, K.V., and Sahani, M. (2011). Empirical models of spiking in neural populations. In Advances in Neural Information Processing Systems 24.

106. Paninski, L., Ahmadian, Y., Ferreira, D.G., Koyama, S., Rahnama Rad, K., Vidne, M., Vogelstein, J., and Wu, W. (2010). A new look at state-space models for neural data. J. Comput. Neurosci. *29*, 107–126. https://doi.org/10.1007/s10827-009-0179-x.

107. Smith, A.C., and Brown, E.N. (2003). Estimating a state-space model from point process observations. Neural Comput. *15*, 965–991. https://doi.org/10.1162/089976603765202622.

108. Driscoll, L.N., Shenoy, K., and Sussillo, D. (2024). Flexible multitask computation in recurrent networks utilizes shared dynamical motifs. Nat. Neurosci. *27*, 1349–1363. https://doi.org/10.1038/s41593-024-01668-6.

109. Sussillo, D., and Barak, O. (2013). Opening the black box: low-dimensional dynamics in high-dimensional recurrent neural networks. Neural Comput. *25*, 626–649. https://doi.org/10.1162/NECO_a_00409.

110. Valente, A., Pillow, J.W., and Ostojic, S. (2022). Extracting computational mechanisms from neural data using low-rank RNNs. Adv. Neural Inf. Process. Syst. *35*, 24072–24086.

111. Costacurta, J.C., Bhandarkar, S., Zoltowski, D.M., and Linderman, S.W. (2024). Structured flexibility in recurrent neural networks via neuromodulation. Preprint at bioRxiv. https://doi.org/10.1101/2024.07.26.605315.

112. Yin, L., Hashikawa, K., Hashikawa, Y., Osakada, T., Lischinsky, J.E., Diaz, V., and Lin, D. (2022). VMHvlIcckar cells dynamically control female sexual behaviors over the reproductive cycle. Neuron *110*, 3000–3017.e8. https://doi.org/10.1016/j.neuron.2022.06.026.

113. Knoedler, J.R., Inoue, S., Bayless, D.W., Yang, T., Tantry, A., Davis, C.H., Leung, N.Y., Parthasarathy, S., Wang, G., Alvarado, M., et al. (2022). A functional cellular framework for sex and estrous cycle-dependent gene expression and behavior. Cell *185*, 654–671.e22. https://doi.org/10.1016/j.cell.2021.12.031.

114. Liu, M., Kim, D.W., Zeng, H., and Anderson, D.J. (2022). Make war not love: The neural substrate underlying a state-dependent switch in female social behavior. Neuron *110*, 841–856.e6. https://doi.org/10.1016/j.neuron.2021.12.002.

115. Mei, L., Osakada, T., and Lin, D. (2023). Hypothalamic control of innate social behaviors. Science *382*, 399–404. https://doi.org/10.1126/science.adh8489.

116. Falkner, A.L., Wei, D., Song, A., Watsek, L.W., Chen, I., Chen, P., Feng, J.E., and Lin, D. (2020). Hierarchical representations of aggression in a hypothalamic-midbrain circuit. Neuron *106*, 637–648.e6. https://doi.org/10.1016/j.neuron.2020.02.014.

117. Yang, T., Bayless, D.W., Wei, Y., Landayan, D., Marcelo, I.M., Wang, Y., DeNardo, L.A., Luo, L., Druckmann, S., and Shah, N.M. (2023). Hypothalamic neurons that mirror aggression. Cell *186*, 1195–1211.e19. https://doi.org/10.1016/j.cell.2023.01.022.

118. Lo, L., Yao, S., Kim, D.W., Cetin, A., Harris, J., Zeng, H., Anderson, D.J., and Weissbourd, B. (2019). Connectional architecture of a mouse hypothalamic circuit node controlling social behavior. Proc. Natl. Acad. Sci. USA *116*, 7503–7512. https://doi.org/10.1073/pnas.1817503116.

119. Aksay, E., Olasagasti, I., Mensh, B.D., Baker, R., Goldman, M.S., and Tank, D.W. (2007). Functional dissection of circuitry in a neural integrator. Nat. Neurosci. *10*, 494–504. https://doi.org/10.1038/nn1877.

120. Salay, L.D., Tsao, D.Y., and Anderson, D.J. (2025). A circuit that integrates drive state and social contact to gate mating. Nature *646*, 394–403. https://doi.org/10.1038/s41586-025-09327-x.

121. Kohl, J., Babayan, B.M., Rubinstein, N.D., Autry, A.E., Marin-Rodriguez, B., Kapoor, V., Miyamishi, K., Zweifel, L.S., Luo, L., Uchida, N., et al. (2018). Functional circuit architecture underlying parental behaviour. Nature *556*, 326–331. https://doi.org/10.1038/s41586-018-0027-0.

122. Cichy, A., Dewan, A., Zhang, J., Kaye, S., Teng, T., Blanchard, K., Feinstein, P., and Bozza, T. (2021). Map-independent representation of an aggression-promoting social cue in the main olfactory pathway. Preprint at bioRxiv. https://doi.org/10.1101/2021.12.30.474554.

123. Itakura, T., Murata, K., Miyamichi, K., Ishii, K.K., Yoshihara, Y., and Touhara, K. (2022). A single vomeronasal receptor promotes intermale aggression through dedicated hypothalamic neurons. Neuron *110*, 2455–2469.e8. https://doi.org/10.1016/j.neuron.2022.05.002.

124. Jazayeri, M., and Afraz, A. (2017). Navigating the Neural Space in Search of the Neural Code. Neuron *93*, 1003–1014. https://doi.org/10.1016/j.neuron.2017.02.019.

125. Green, J., Adachi, A., Shah, K.K., Hirokawa, J.D., Magani, P.S., and Maimon, G. (2017). A neural circuit architecture for angular integration in Drosophila. Nature *546*, 101–106. https://doi.org/10.1038/nature22343.

126. Daie, K., Svoboda, K., and Druckmann, S. (2021). Targeted photostimulation uncovers circuit motifs supporting short-term memory. Nat. Neurosci. *24*, 259–265. https://doi.org/10.1038/s41593-020-00776-3.

127. Packer, A.M., Russell, L.E., Dalgleish, H.W.P., and Häusser, M. (2015). Simultaneous all-optical manipulation and recording of neural circuit activity with cellular resolution in vivo. Nat. Methods *12*, 140–146. https://doi.org/10.1038/nmeth.3217.

128. Zhang, Z., Russell, L.E., Packer, A.M., Gauld, O.M., and Häusser, M. (2018). Closed-loop all-optical interrogation of neural circuits in vivo. Nat. Methods *15*, 1037–1040. https://doi.org/10.1038/s41592-018-0183-z.

129. Stagkourakis, S., Spigolon, G., Williams, P., Protzmann, J., Fisone, G., and Broberger, C. (2018). A neural network for intermale aggression to establish social hierarchy. Nat. Neurosci. *21*, 834–842. https://doi.org/10.1038/s41593-018-0153-x.

130. Boucheny, C., Brunel, N., and Arleo, A. (2005). A continuous attractor network model without recurrent excitation: maintenance and integration in the head direction cell system. J. Comput. Neurosci. *18*, 205–227. https://doi.org/10.1007/s10827-005-6559-y.

131. Couey, J.J., Witoelar, A., Zhang, S.J., Zheng, K., Ye, J., Dunn, B., Czajkowski, R., Moser, M.B., Moser, E.I., Roudi, Y., et al. (2013). Recurrent inhibitory circuitry as a mechanism for grid formation. Nat. Neurosci. *16*, 318–324. https://doi.org/10.1038/nn.3310.

132. Hahn, J.D., Sporns, O., Watts, A.G., and Swanson, L.W. (2019). Macroscale intrinsic network architecture of the hypothalamus. Proc. Natl. Acad. Sci. USA *116*, 8018–8027. https://doi.org/10.1073/pnas.1819448116.

133. Turner-Evans, D.B., Jensen, K.T., Ali, S., Paterson, T., Sheridan, A., Ray, R.P., Wolff, T., Lauritzen, J.S., Rubin, G.M., Bock, D.D., et al. (2020). The Neuroanatomical Ultrastructure and Function of a Biological Ring Attractor. Neuron *108*, 145–163.e10. https://doi.org/10.1016/j.neuron.2020.08.006.

134. Green, J., and Maimon, G. (2018). Building a heading signal from anatomically defined neuron types in the Drosophila central complex. Curr. Opin. Neurobiol. *52*, 156–164. https://doi.org/10.1016/j.conb.2018.06.010.

135. Jenett, A., Rubin, G.M., Ngo, T.T.B., Shepherd, D., Murphy, C., Dionne, H., Pfeiffer, B.D., Cavallaro, A., Hall, D., Jeter, J., et al. (2012). A GAL4-driver line resource for Drosophila neurobiology. Cell Rep. *2*, 991–1001. https://doi.org/10.1016/j.celrep.2012.09.011.

136. Pfeiffer, B.D., Ngo, T.T.B., Hibbard, K.L., Murphy, C., Jenett, A., Truman, J.W., and Rubin, G.M. (2010). Refinement of Tools for Targeted Gene Expression in Drosophila. Genetics *186*, 735–755. https://doi.org/10.1534/genetics.110.119917.

137. Scheffer, L.K., Xu, C.S., Januszewski, M., Lu, Z., Takemura, S.Y., Hayworth, K.J., Huang, G.B., Shinomiya, K., Maitlin-Shepard, J., Berg, S., et al. (2020). A connectome and analysis of the adult Drosophila central brain. eLife *9*, e57443. https://doi.org/10.7554/eLife.57443.

138. Schlegel, P., Yin, Y., Bates, A.S., Dorkenwald, S., Eichler, K., Brooks, P., Han, D.S., Gkantia, M., Dos Santos, M., Munnelly, E.J., et al. (2024). Whole-brain annotation and multi-connectome cell typing of Drosophila. Nature *634*, 139–152. https://doi.org/10.1038/s41586-024-07686-5.

139. Vishwanathan, A., Sood, A., Wu, J., Ramirez, A.D., Yang, R., Kemnitz, N., Ih, D., Turner, N., Lee, K., Tartavull, I., et al. (2024). Predicting modular functions and neural coding of behavior from a synaptic wiring diagram. Nat. Neurosci. *27*, 2443–2454. https://doi.org/10.1038/s41593-024-01784-3.

140. Kim, D.W., Yao, Z., Graybuck, L.T., Kim, T.K., Nguyen, T.N., Smith, K.A., Fong, O., Yi, L., Koulena, N., Pierson, N., et al. (2019). Multimodal Analysis of Cell Types in a Hypothalamic Node Controlling Social Behavior. Cell *179*, 713–728.e17. https://doi.org/10.1016/j.cell.2019.09.020.

141. Osakada, T., Yan, R., Jiang, Y., Wei, D., Tabuchi, R., Dai, B., Wang, X., Zhao, G., Wang, C.X., Liu, J.J., et al. (2024). A dedicated hypothalamic oxytocin circuit controls aversive social learning. Nature *626*, 347–356. https://doi.org/10.1038/s41586-023-06958-w.

142. DeNardo, L., and Luo, L. (2017). Genetic strategies to access activated neurons. Curr. Opin. Neurobiol. *45*, 121–129. https://doi.org/10.1016/j.conb.2017.05.014.

143. Shao, Y.Q., Fan, L., Wu, W.Y., Zhu, Y.J., and Xu, H.T. (2022). A developmental switch between electrical and neuropeptide communication in the ventromedial hypothalamus. Curr. Biol. *32*, 3137–3145.e3. https://doi.org/10.1016/j.cub.2022.05.029.

144. Garcia-Garcia, M.G., Kapoor, A., Akinwale, O., Takemaru, L., Kim, T.H., Paton, C., Litwin-Kumar, A., Schnitzer, M.J., Luo, L., and Wagner, M.J. (2024). A cerebellar granule cell-climbing fiber computation to learn to track long time intervals. Neuron *112*, 2749–2764.e7. https://doi.org/10.1016/j.neuron.2024.05.019.

145. Czéh, B., Fuchs, E., Wiborg, O., and Simon, M. (2016). Animal models of major depression and their clinical implications. Prog. Neuropsychopharmacol. Biol. Psychiatry *64*, 293–310. https://doi.org/10.1016/j.pnpbp.2015.04.004.

146. LeDuke, D.O., Borio, M., Miranda, R., and Tye, K.M. (2023). Anxiety and depression: A top-down, bottom-up model of circuit function. Ann. N. Y. Acad. Sci. *1525*, 70–87. https://doi.org/10.1111/nyas.14997.

147. Rolls, E.T. (2016). A non-reward attractor theory of depression. Neurosci. Biobehav. Rev. *68*, 47–58. https://doi.org/10.1016/j.neubiorev.2016.05.007.

148. Rolls, E.T., Loh, M., and Deco, G. (2008). An attractor hypothesis of obsessive-compulsive disorder. Eur. J. Neurosci. *28*, 782–793. https://doi.org/10.1111/j.1460-9568.2008.06379.x.

149. Mitchell-Heggs, R., Prado, S., Gava, G.P., Go, M.A., and Schultz, S.R. (2023). Neural manifold analysis of brain circuit dynamics in health and disease. J. Comput. Neurosci. *51*, 1–21. https://doi.org/10.1007/s10827-022-00839-3.

150. Strock, A., Mistry, P.K., and Menon, V. (2024). Digital twins for understanding mechanisms of learning disabilities: Personalized deep neural networks reveal impact of neuronal hyperexcitability. Preprint at bioRxiv. https://doi.org/10.1101/2024.04.29.591409.

151. Strock, A., Mistry, P.K., and Menon, V. (2025). Personalized deep neural networks reveal mechanisms of math learning disabilities in children. Sci. Adv. *11*, eadq9990. https://doi.org/10.1126/sciadv.adq9990.

152. Holtzheimer, P.E., and Mayberg, H.S. (2011). Stuck in a rut: rethinking depression and its treatment. Trends Neurosci. *34*, 1–9. https://doi.org/10.1016/j.tins.2010.10.004.

153. Hultman, R., Ulrich, K., Sachs, B.D., Blount, C., Carlson, D.E., Ndubuizu, N., Bagot, R.C., Parise, E.M., Vu, M.T., Gallagher, N.M., et al. (2018). Brain-wide Electrical Spatiotemporal Dynamics Encode Depression Vulnerability. Cell *173*, 166–180.e14. https://doi.org/10.1016/j.cell.2018.02.012.

154. Walder-Christensen, K., Abdelaal, K., Klein, H., Thomas, G.E., Gallagher, N.M., Talbot, A., Adamson, E., Rawls, A., Hughes, D., Mague, S.D., et al. (2024). Electome network factors: Capturing emotional brain networks related to health and disease. Cell Rep. Methods *4*, 100691. https://doi.org/10.1016/j.crmeth.2023.100691.

155. Taschereau-Dumouchel, V., Michel, M., Lau, H., Hofmann, S.G., and LeDoux, J.E. (2022). Putting the "mental" back in "mental disorders": a perspective from research on fear and anxiety. Mol. Psychiatry *27*, 1322–1330. https://doi.org/10.1038/s41380-021-01395-5.

156. Courellis, H.S., Valiante, T.A., Mamelak, A.N., Adolphs, R., and Rutishauser, U. (2024). Neural dynamics underlying minute-timescale persistent behavior in the human brain. Preprint at bioRxiv. https://doi.org/10.1101/2024.07.16.603717.

157. Daume, J., Kamiński, J., Salimpour, Y., Gómez Palacio Schjetnan, A., Anderson, W.S., Valiante, T.A., Mamelak, A.N., and Rutishauser, U. (2024). Persistent activity during working memory maintenance predicts long-term memory formation in the human hippocampus. Neuron *112*, 3957–3968.e3. https://doi.org/10.1016/j.neuron.2024.09.013.

158. Ságodi, Á., Martín-Sánchez, G., Sokol, P., and Park, M. (2024). Back to the continuous attractor. Adv. Neural Inf. Process. Syst. *37*, 66856–66906.

159. Noorman, M., Hulse, B.K., Jayaraman, V., Romani, S., and Hermundstad, A.M. (2024). Maintaining and updating accurate internal representations of continuous variables with a handful of neurons. Nat. Neurosci. *27*, 2207–2217. https://doi.org/10.1038/s41593-024-01766-5.

160. Ganguli, S., Bisley, J.W., Roitman, J.D., Shadlen, M.N., Goldberg, M.E., and Miller, K.D. (2008). One-dimensional dynamics of attention and decision making in LIP. Neuron *58*, 15–25. https://doi.org/10.1016/j.neuron.2008.01.038.