nature reviews neuroscience

https://doi.org/10.1038/s41583-022-00642-0

Review article

Check for updates

Attractor and integrator networks in the brain

Mikail Khona[1,2,3,4] & Ila R. Fiete[1,2,3]

Abstract

In this Review, we describe the singular success of attractor neural network models in describing how the brain maintains persistent activity states for working memory, corrects errors and integrates noisy cues. We consider the mechanisms by which simple and forgetful units can organize to collectively generate dynamics on the long timescales required for such computations. We discuss the myriad potential uses of attractor dynamics for computation in the brain, and showcase notable examples of brain systems in which inherently low-dimensional continuous-attractor dynamics have been concretely and rigorously identified. Thus, it is now possible to conclusively state that the brain constructs and uses such systems for computation. Finally, we highlight recent theoretical advances in understanding how the fundamental trade-offs between robustness and capacity and between structure and flexibility can be overcome by reusing and recombining the same set of modular attractors for multiple functions, so they together produce representations that are structurally constrained and robust but exhibit high capacity and are flexible.

Sections

Introduction

What are attractors?

Construction and mechanisms

Attractors for neural computation

Evidence of attractors in the brain

Departures from attractor dynamics

Flexibility despite rigidity

Looking ahead

[1]Department of Brain and Cognitive Sciences, MIT, Cambridge, MA, USA. [2]K. Lisa Yang ICoN Center, MIT, Cambridge, MA, USA. [3]McGovern Institute, MIT, Cambridge, MA, USA. [4]Department of Physics, MIT, Cambridge, MA, USA. e-mail: fiete@mit.edu

Introduction

One of biology's grand challenges is to explain how order and complex function spring from inanimate physical systems composed of much simpler parts. The brain creates order in its representations of the world and performs complex functions through the collective interactions of simpler elements. In this Review, we describe and evaluate the hypothesis that attractor dynamics in widespread regions of the CNS have a key role in constructing some of these representations, generating long timescales to support integration and memory functions and endowing all these functions with robustness. We review the specific predictions of attractor-based models and the now extensive body of work testing these predictions. Thus, we illustrate that the theory and validation of computation with attractor dynamics in the brain is one of the biggest success stories in systems neuroscience.

Some of the first formal circuit-level models of brain function focused on the problem of associative memory and how neural circuits might generate spatially distributed, stable patterns of activity that could function as such a memory[1–4]. Hopfield networks, with multiple stable states constructed by inscribing input patterns into connection weights, were proposed more than four decades ago[3,5,6]. Network models possessing a continuous set of stable states that could be used to represent continuous variables were also first proposed in the same period[7]. Subsequently, many canonical brain circuits for motor control, sensory amplification and memory, motion integration, evidence integration, decision-making and spatial navigation have been modelled using the same general principle — that a set of states can be stabilized through collective positive feedback[8–17].

Because these are circuit-level models, but were typically inspired by experimental characterization of neurons recorded singly or a few at a time, the patterns of connectivity and the cell–activity correlations in the models automatically became novel and relatively specific predictions about the population dynamics and architecture of such circuits. As we discuss below, the combination of these prediction-rich (yet conceptually simple) models, modern experimental breakthroughs in the acquisition of cellular-resolution population activity data and novel and rigorous analyses of such data on the basis of the model predictions has provided much evidence that the brain constructs and exploits attractor networks for performing several essential computations.

We begin by defining attractors, and then describe proposed mechanisms for the construction of attractor network models in neuroscience. We provide an overview of why attractor networks can be important for computation in the brain and highlight criteria for determining whether a system has non-trivial attractor dynamics. We also discuss examples of brain circuits with non-trivial attractor dynamics. We end with a summary of new directions in our understanding of how these simple circuits could contribute to flexible computation through reuse in multiple contexts.

What are attractors?

To define an attractor, we first define a dynamical system and its states. A dynamical system is a set of variables together with all the rules that determine their changes in value with the passage of time. The value of these variables at any given instant is called the state of the system at that moment. The state is a point (vector) in the state space of the dynamical system. An attractor is the minimal set of states in a state space, to which all nearby states eventually flow with time[18]. One simple example of an attractor is a stable fixed point: all neighbouring states flow to it. Transferring these crisp mathematical definitions to the context of the brain involves challenges and simplifications that revolve around identifying a sufficiently self-contained system and the variables necessary to determine its dynamics.

Defining the state of a neural system

Inherent in the definition of a dynamical system is the assumption that there are no external dynamical inputs to the system (or, equivalently, that the system definition includes all such external variables).

The first simplification in characterizing the dynamics of a neural circuit is to assume that, at least on the timescale of interest, the system evolves in an autonomous way. Given that subcircuits in the brain are interconnected with others, and that the brain itself interacts with the world, it is impossible to isolate these circuits completely into autonomous systems. However, we may define a notion of 'effectively autonomous' dynamics, whereby inputs do not vary over time and are untuned, in the sense that they do not provide differential drive to subsets of the putative set of attractor states.

The second simplification is in defining the states of the system. The changes in state of a circuit in the brain over time may depend on the detailed pattern of all the spikes in all neurons, the levels of associated ions, neurotransmitters and modulators, and even the states of the ion channels. The weights and connections between neurons may be considered as parameters (rather than variables) on short timescales, but are themselves variables if considering a longer timescale. One widely used simplification in describing a neural circuit on the timescale of seconds is to use just the spiking outputs of the neurons in the circuit as the states, often further simplified as time-varying spike rates. If such a description is sufficient to predict the state changes of the system at the relevant timescales, it can be viewed as a reasonable dynamical system model of the circuit. Although spike or spike-rate descriptions ignore subcellular and molecular variables to make the grossly simplifying assumption that the relevant circuit dynamics are governed by spikes, the state space of a vertebrate microcircuit described in this way is nevertheless very high-dimensional, comprising the number of neurons in the circuit, which can be in the order of 10^2–10^7 cells. As we discuss below, such simplified models can nevertheless yield rich and accurate predictions about neural circuits.

Attractors exist in various flavours: an attractor may consist of a single state, a set of discrete states, a set of states that effectively behave as a continuous set or many such near-continuous sets (Fig. 1). If a set of attractor states traces out a shape in state space that is approximately continuous and locally Euclidean, it is known as an attractor manifold. Nonlinear continuous-attractor manifolds can be curved and topologically complex (for example, resembling rings, tori and so on; Fig. 1c,d, rightmost column)[19,20]. States on an attractor may be stationary, or might flow along the attractor to trace out trajectories that are periodic (known as limit cycles; Fig. 1f, rightmost column) or chaotic (that is, with dynamics that are inherently unpredictable owing to high sensitivity to small changes in the state[21]).

Various combinations of such attractors, of different dimensions, geometries and topologies, may coexist in different regions of the state space of a single dynamical system. Typically, the set of attractors in a dynamical system comprises a small subset of the state space, and attractor manifolds are usually much lower-dimensional than the state space. In cases in which a system has multiple attractor states, the initial condition determines the attractor state to which the system flows.

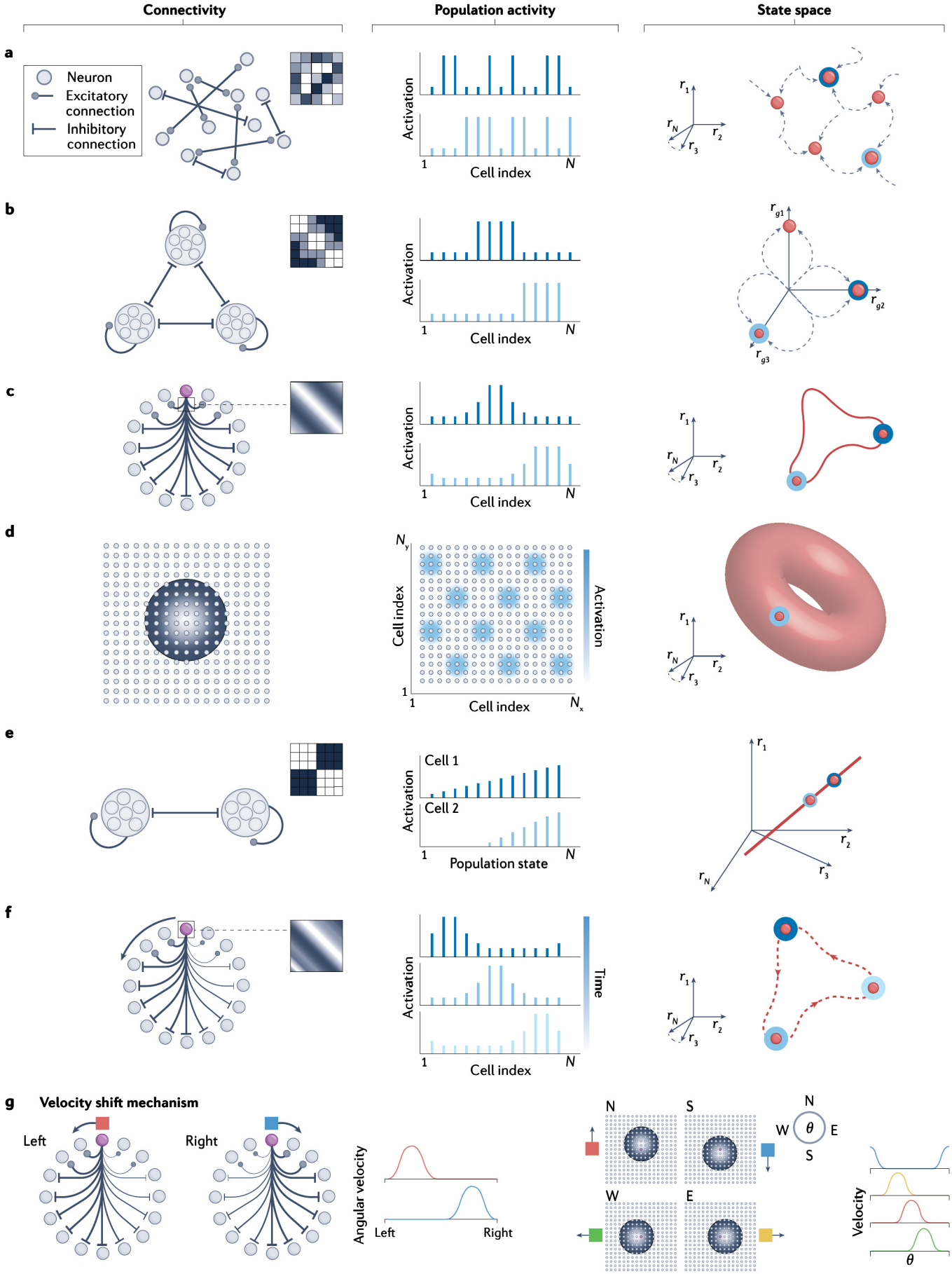

Connectivity
Population activity
State space
a
Neuron
Excitatory connection
Inhibitory connection
Activation
Cell index
1
N
b
c
d
Cell index
Activation
e
Cell 1
Cell 2
Population state
f
Time
g
Velocity shift mechanism
Left
Right
Angular velocity
N
S
W
E
θ
Velocity

Fig. 1 | Mechanisms of attractor formation. Left columns: open grey circles represent neurons, and connections between them are excitatory (black lines ending in bars) or inhibitory (black lines ending in circles). For layout of neurons and connections, connectivity matrices are shown as the inset, with black to white colours indicating strongly inhibitory to excitatory interactions, respectively. Middle columns: examples of stable population activity patterns. Right columns: state-space views of population states and dynamics. Red circles with shades of blue rings indicate the activity states shown in middle column; grey lines denote transient dynamic trajectories and red denotes attracting states. **a**, A network with dense symmetric connections determined by associative Hebbian learning on a set of input patterns (middle) stores them as stable attractor states. This defines a Hopfield network. **b**, Disjoint groups of neurons that interact through within-group excitation and across-group inhibition lead to group winner-takes-all (WTA) dynamics. Stable states are any patterns with one winning group. The state-space plot collapses all activities of neurons in group g_i along the axis r_{gi}. **c**, Neurons arranged in a ring with global inhibition and either local excitation or a lack of local inhibition, combined with uniform excitatory input to all neurons, produce localized activity bumps (middle) as the stable states. Bumps may be centred anywhere on the neural ring, defining a near-continuum of attractor states that form a ring in state space (right). **d**, Neurons arranged on a two-dimensional neural sheet, interacting through local inhibition and either centre excitation or a lack of inhibition near the centre with uniform excitatory input to all neurons, result in a pattern of multiple periodically spaced activity bumps (middle). Any two-dimensional phase shift of the periodic pattern up to the lattice periodicity results in distinct but equivalent stable states, and then the states repeat; thus, the result is a torus of stable states. **e**, Two neuron groups with in-group excitation and across-group inhibition, precisely tuned interaction strengths and quasi-linear neural input–output responses can counteract activity decay in the network and produce persistent activity over a continuum of activity levels in the two populations, defining ramp-like neural tuning and a line of attractor states. **f**, Neurons arranged on a ring with asymmetric connections drive a flow of neural activity in a particular direction. The network forms localized activity bumps that sequentially move around the ring in that direction (middle). The state space contains a limit-cycle attractor (right). **g**, The copy-and-offset mechanism for constructing integrators, illustrated for the ring (left) and grid (right) attractor circuits. Each network copy receives velocity inputs tuned to the corresponding shift direction.

Attractors in the presence of noise

Any real physical system unavoidably behaves non-deterministically from the perspective of a model of the system. This is because one cannot observe and describe all variables, and all uncharacterized variables together with true stochastic sources of variation (such as synaptic signalling noise from stochastic vesicle release[22]; fluctuations in ion concentrations during processes such as spike initiation[23] and calcium signalling; or fluctuations in small copy numbers of proteins[24]) serve as effective sources of noise in the model. Noise can disrupt states so they do not strictly localize to the attractor described in a noise-free version of the model, and can drive the system to escape from an attractor over time. However, the general idea of attractor states remains, in that, if the system is initialized near such a state, it tends to flow towards it and subsequently remains localized around it, for extended periods.

Because attractor states are where systems tend to localize (when not externally driven), they should be observable in the autonomous dynamics of real systems. This basic property is the basis for the most fundamental and robust tests of attractor dynamics in neural systems, as we discuss below. In a nutshell, the central signatures of attractors in real systems (discussed in more detail in later sections of this Review) can be summarized as: the localization of the states of a system to a lower-dimensional subset; the flow of the states towards the subset after perturbation; and the long-time and (effectively) autonomous stability of states in that subset.

Construction and mechanisms

The general principle underlying the formation of non-trivial attractor states in neural circuits is strong recurrent positive feedback. Positive feedback fights activity decay to stabilize certain states, and has been posited[25–28] to be the basis for the stabilization of memory traces and persistent activity in the brain. Which states become stabilized into attractors depends on how the network sculpts the positive feedback, which, according to the synaptic hypothesis, is determined by synaptic weights[29–31].

In general, characterizing the relationship between structure and function in a large collection of interacting elements is extremely difficult[32]. For example, a large collection of simple polar three-atom molecules of hydrogen and oxygen give rise to the emergent phenomena we associate with water — such as liquidness, wetness and freezing into a solid — that cannot be predicted through intuition or by drawing box and arrow diagrams. Nevertheless, the transitions and properties of emergent states can be described relatively simply, with very few key parameters and variables.

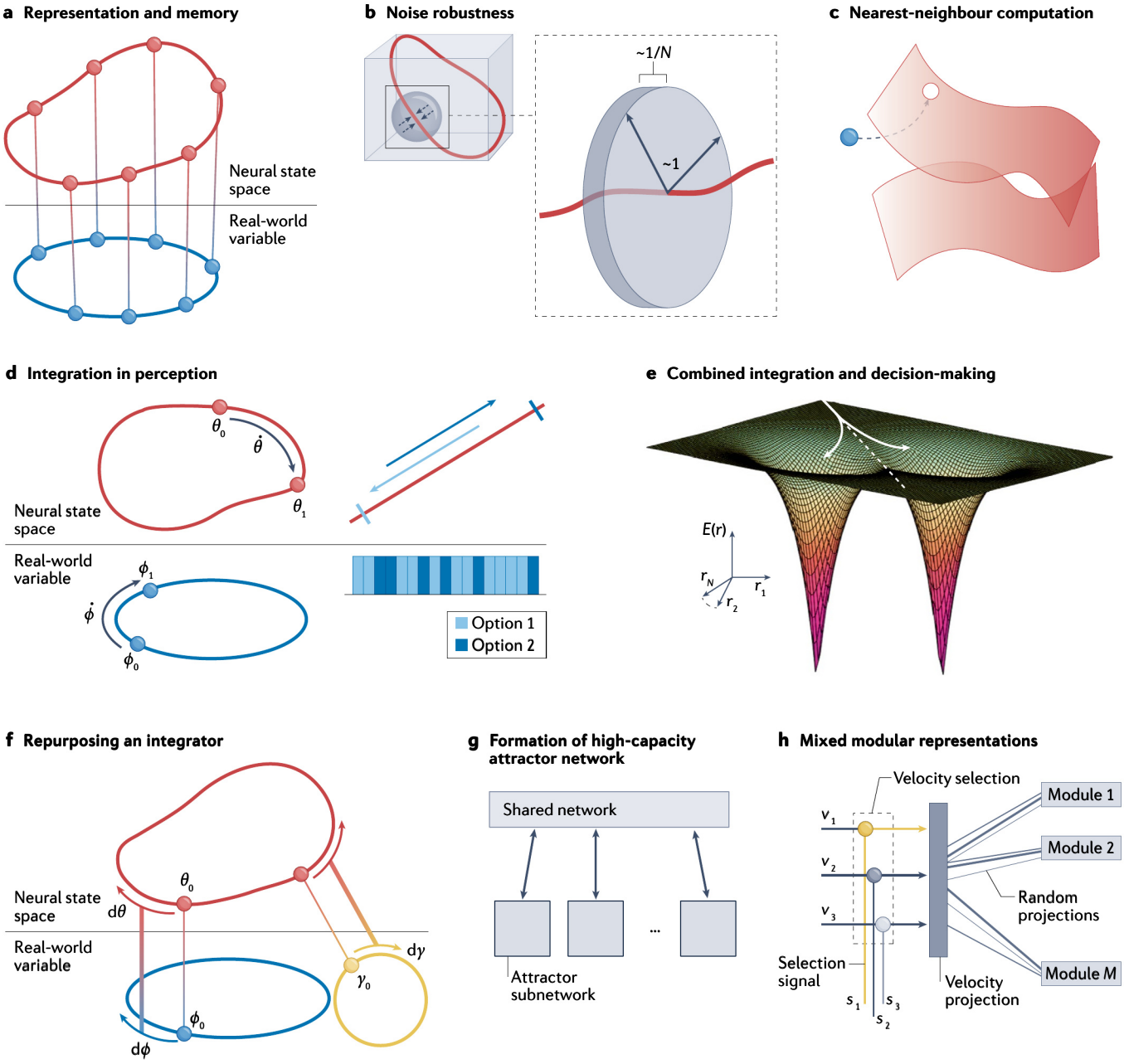
One way to characterize the relationship between synaptic weights and attractor dynamics is to ask what attractor states a given set of weights produces (the 'forward' problem). With a given set of weights, one can simulate a circuit and explore the resulting dynamics to find attractors of the system. A more powerful method, the Lyapunov function approach, holds for symmetric weight matrices ($W_{ij} = W_{ji}$) and rate-based neural dynamics. For this class of models, a generalized energy function (the Lyapunov function), which is a function of the weights and neural activation function[2,5,6], analytically specifies the network's dynamics. Stable and unstable attractor states are the energy minima and maxima of the derived landscape, respectively, and the network's state flows downhill towards the attractors (Fig. 2e) in the way a ball rolls down a gravitational potential.

Another way to characterize the relationship between attractors and network structure is to consider the 'inverse' problem: given a set of attractors, what network structure could generate it? Neuroscientists want to solve the inverse problem to make predictions about underlying mechanisms and, because neural activations are more readily observed than synaptic weights, the inverse problem is more frequently encountered than the forward problem. By contrast, evolution, the brain and artificially intelligent systems must solve the inverse problem to be able to perform computations that require a given type of attractor dynamics (discussed below). Theoretical neuroscience has discovered some solutions to the inverse problem for different types of attractors, as we describe below.

Discrete attractors

A well-known prescription for creating a set of discrete attractors at user-defined points is given by the Hopfield model[5] (Fig. 1a). Input patterns of neural activation are inscribed into the network weights through a Hebbian-like learning rule, such that co-active neurons are connected by excitatory interactions and inhibit all the rest. Thus, these patterns stabilize themselves and become attractor states. If a sufficiently small number of patterns are learned, they can be retrieved from partial or corrupted versions of the stored states, and thus the network can be said to store content-addressable memories. More generally, the attractors of simple rate-based networks with arbitrary symmetric weight matrices and without communication delays consist entirely of fixed points. Some non-symmetric networks can also support point attractors[33], but not generically, and they can require additional mechanisms such as homeostatic plasticity[34,35].

Attractor states in Hopfield-like networks typically have highly overlapping neural memberships, even when they are well separated in



the state space (Fig. 1a, middle column). Thus, there is not a clear notion of distinct 'cell assemblies'. In a special case of Hopfield networks, neurons are partitioned into largely disjointed groups with self-excitation within groups and inhibition between groups. In these winner-take-all (WTA) networks, the attractor states consist of largely non-overlapping active cell groups, which might then be called 'assemblies' (Fig. 1b).

Continuous attractors

How can one construct networks with a continuum of stationary attractor states? Weight matrices with a particular symmetry (across the diagonal) give rise to discrete attractors, as we have seen. If the weights instead exhibit a continuous symmetry — for example, if the weight profiles are invariant across neurons (they look the same at each neuron, thus the symmetry is translational) — then the set of formed attractors will be related by the same symmetry and could thus form a continuous set.

The general principle for the formation of stationary continuous attractors is pattern formation[36–42]. Simple and spatially local competitive interactions across the neural sheet lead to the emergence of spatially structured activity patterns that are stable states: neurons with excitatory coupling between them become co-active and suppress the rest of their neighbours through inhibition in what is known as a linear Turing instability[36].

Three conditions are generally sufficient (although not strictly necessary) to provide a solution to the inverse problem for forming stationary continuous attractors (Box 1). First, the system must include nonlinear neurons with saturating responses or inhibition-dominated

Fig. 2 | The utility of low-dimensional attractor networks. a, Persistent and stable states generated by attractor networks (red) can be used to represent and remember external variables (blue) by constructing an appropriate mapping between them (vertical lines). **b**, Attractor networks can correct errors by mapping noisy states to the nearest attractor state[262]. *N*-dimensional noise drawn from the unit sphere centred on a one-dimensional attractor has a projection strength of only $1/N$ along the attractor: in this counter-intuitive high-dimensional geometry, a ball is more similar to a pancake, with the attractor orthogonal to the large dimensions[19]. **c**, Flow to the nearest (continuous or discrete) attractor can perform a nearest-neighbour computation and, thus, perform classification. For example, the two attractors may represent 'cat' and 'dog' perceptual manifolds, and the blue dot a specific input data point. **d**, Left: continuous attractors can become integrators if velocities or movements in the external space are inputs to the network and induce proportional shifts in the internal attractor state. The current state on the attractor is then the integral of past velocity inputs relative to the starting state. Right: if the input to an integrating attractor consists of temporally varying evidence pulses (bottom, evidence about one option in dark blue and evidence about the opposing option in light blue), these will move the state on the attractor (top) so the system's current state reflects the integral of the total evidence. **e**, The energy (E) landscape of a combined integration and decision-making network: inputs push the state left or right, and as the system integrates, the network state also moves towards one of two discrete attractors (left and right; white arrows, two sample trajectories). Arrival in the basin of one of the discrete attractors is a decision point[65,66]. **f**, An integrator can be quickly re-purposed to represent multiple different and new external variables simply by yoking its velocity shift mechanism to different external velocities cues through feedforward learning. This mechanism also supports zero-shot learning and inference: given an initial state and an input velocity trajectory, it will generate a self-consistent representation for the current state even if the trajectory is different and new each time[218,230,232]. **g**, A set of (continuous or discrete) attractor subnetworks (red boxes at bottom) can interact bidirectionally with a shared network to form a high-capacity attractor network[62,236,237,263]. **h**, Mixed modular representations can enable representation of inputs of different dimensions, by reusing the same attractors of fixed dimension each. Velocities (v_i) from external spaces of potentially different dimension are selected by a set of selection signals (s_i). The selected velocity (green) is routed through random projections to a set of M modular integrator networks of dimension K each. This kind of mixed modular circuit can interchangeably represent various input spaces of dimension $D \leq MK$ while smoothly trading off resolution for dimension[230].

recurrent interactions and a uniform excitatory drive[7,10,15,17,43,44] to keep network activity bounded. Second, the system must involve sufficiently strong recurrent weights with competitive dynamics in the form of local excitation or disinhibition, with broader inhibition, to drive spontaneous pattern formation through the Turing instability[10,15,17,36–42,45–47]; these patterns become the attractor states. Last, the system requires some continuous symmetry in the weights (a continuous weight symmetry is one where as some variable is varied continuously, the weights remain invariant), such as translational or rotational invariance (Fig. 1c,d), to ensure a continuum of attractor states.

A special set of networks generate continuous-attractor dynamics without pattern formation: those with linear, planar or hyperplanar attractors that are generated by neurons with linear or near-linear response functions. In circuits of linear neurons, the feedback within the network is a linear function of activity (Wr, where W is the weight matrix and r are the neural activities), as is the activity decay (given by $-r$). Such networks can stabilize non-zero activity states simply by tuning positive feedback to cancel the decay. The matrix W can direct feedback in state space; if feedback is directed largely along one dimension, the network can support a line attractor (Fig. 1e). If it is directed equally along two or more dimensions, it can support a plane or hyperplane attractor. To create long-lived attractors requires that the network feedback magnitude is finely tuned to precisely cancel the decay[9,48], in contrast to pattern-forming continuous-attractor systems where the weight shapes (but not magnitudes) are tuned to maintain continuous symmetry across neurons.

Non-stationary continuous attractors

Large non-symmetric networks with nonlinear neurons and strong connectivity generically exhibit limit-cycle attractors or chaotic dynamics[49,50]. Just as point attractors emerge generically in large networks with strong symmetric weights and bounded state spaces, chaotic attractors emerge generically in large recurrent networks with strong asymmetric weights. Adequate asymmetries are easily achieved if excitatory and inhibitory synapses emerge from distinct sets of neurons[49], as biologically necessitated by Dale's law.

Despite the complexity of chaotic dynamics, chaotic attractors are also highly structured in that they typically exist in a relatively low number of dimensions compared with the number of neurons in the network[51]. Non-symmetric networks that are dominated by inhibition exhibit a single attractor at zero activity, although the flow towards the attractor in response to perturbations can involve large transients in neural activation that temporarily move the state further away from the attractor[52,53].

Attractors for neural computation

A system could theoretically be perfectly tuned such that every point in state space is a neutrally stable attractor, and thus the system has maximally high-dimensional attractor dynamics. However, because the robustness of attractor networks is related to the low-dimensionality of the attractor states (as discussed below), the system would lose most of its interesting computational properties: error correction or noise tolerance, nearest-neighbour computation, pattern completion and content-addressable memory. It could perform integration, but with no robustness to noise. As such, networks with low-dimensional attractor dynamics exhibit myriad properties that can be vital for computation in the brain These include robust representation, memory, sequence generation, integration, and robust classification and decision-making — ideas that have been extensively explored in the literature. In a later section, we describe how, although attractor dynamics may be rigid and invariant as needed for the roles listed above, recent theoretical and experimental findings are beginning to reveal how these rigid constructions may also be exploited to perform flexible computation through reuse and recombination across tasks.

Representation and memory

A representation of a set of inputs means the assignment of inputs to representational states (not necessarily on a one-to-one basis), with the ability to reproducibly retrieve those states ('labels') when cued. Attractor networks provide a stable internal set of states that can be used for reproducible representation of discrete or analogue variables, by mapping states in the world to the attractor states. One way to achieve this mapping is through a feedforward learning process that associates each external state with an internal attractor state (Fig. 2a).

An attractor network can exhibit two kinds of memory. The first is in the structure of the weights, which specify the set of all attractors.

If these weights are specified through an input-driven learning process, this is a form of long-term memory about the inputs. The second kind of memory is the ability to maintain persistent activity in a stationary attractor state: if a system with multiple stationary attractor states is initialized in one of them, it will tend to remain at or near the same state for some time. In other words, the activation levels of the neurons contributing to that state persist while the system remains in the state. This persistent activity response is thus a form of short-term memory of the input that initialized the circuit. If these persistent memory states can be activated without an explicit address, using just the content (or partial content) of the memory, they are content-addressable.

The short-term memory function of attractors depends on the prior formation of stable states through long-term plasticity. For instance, in Hopfield-like networks, states cannot persist if they were not first trained to be attractor states. Even models of short-term memory that are based on presynaptic facilitation, rather than persistent activity, rely implicitly on prior long-term associative plasticity to construct recurrently stabilized neural ensembles that can be reinstated by random inputs[54]. (Additionally, these models are not activity-silent in the delay period, in the sense that they would require ongoing activity to refresh the facilitation state over longer delays and to generate robustness against random background activity that would facilitate different synapses.) In other words, these presynaptic facilitation models cannot explain short-term memory for entirely novel inputs; however, combinations of attractors could enable more flexible short-term memory, as we discuss later.

De-noising representations and memories

If representational states are attractors, then the representations are robust in the sense that they perform de-noising: if the input cues or initial conditions reflect noisy or corrupted versions of an attractor state, the dynamics drive the state to a point on the representational attractor (Fig. 2b, inset). When attractors form a continuous manifold of dimension $K \ll N$, where N is the number of neurons in the circuit, all noise in $N–K$ dimensions is erased. A noise ball of unit radius in N dimensions (corresponding to random independent noise per neuron) has a projection of size only $\sim\sqrt{K/N} \ll 1$ along K dimensions. If K is low-dimensional, as is often the case, and N ranges from 10^2 to 10^7 as estimated before for common microcircuits, this constitutes a massive reduction in the sensitivity of the state to internal or input noise (Fig. 2b). Thus, most noise is rendered impotent by attractor dynamics.

De-noising owing to attractor dynamics is especially important for memory maintenance as, otherwise, noise-induced deviations would accumulate and grow over time. Discrete attractors continually erase all noise by mapping perturbed states back to the point attractor, resulting in zero drift. With continuous attractors as memory states,

Box 1

Attractor dynamics, anatomical topography and weight symmetries

Anatomical topography, in which functionally similar neurons are near one another, is neither a necessary nor a sufficient condition for the existence of an attractor, because any low-dimensional attractor network is mathematically unchanged if all weights are preserved but neuron locations are scrambled. However, if the network is merely a spatially scrambled version of the idealized model, then the symmetries of the weight matrix can be revealed after an appropriate reordering of the neurons. An advantage of anatomical topography from a biological perspective is that it can reduce the complexity of development, in that wiring decisions can be guided by spatial proximity rather than depending entirely on activity or other target cell-signalling mechanisms. For example, the locally competitive interactions of grid and head-direction circuit models could be largely constructed through local arborization. Anatomical topography also reduces overall wiring length in the mature circuit[269]. However, a circuit with three-dimensional dynamics or higher that are represented in an unfactorizable form cannot be embedded topographically in a two-dimensional cell layout, limiting the feasibility of topographic layouts for circuits that represent higher-dimensional unfactorizable manifolds.

In addition, the posited weight symmetries in simple models of attractors need not exist in a biological instance of the circuit with the same dynamics: unscrambling or reordering neurons may not be sufficient to reveal the symmetries. Consider, for example, a scenario in which low-dimensional attractor dynamics are generated by a recurrent network of N neurons, but are only needed downstream in a set of $M < N$ neurons. In this situation, the weight symmetries needed for continuous-attractor dynamics can be spread across both the recurrent and readout networks, such that the weights of the recurrent network alone will not reflect the relevant symmetries. Unveiling the symmetry in the circuit weights will require combining the readout weights with the recurrent ones[247].

These considerations give rise to a hypothesis for circuits with continuous attractors of dimension ≤2: evolutionarily conserved circuits that do not require extensive early experience[270,271] should be topographically organized. We might thus predict that the circuit that originates head-direction signals in mammals should be topographically organized. By contrast, if low-dimensional dynamics only emerge on the basis of activity-dependent plasticity with repetitive training, we may not expect the circuit to be topographically organized (or even localized to a single brain region).

Remarkably, despite these caveats, and in a beautiful example of the predictive power of simple theories in neuroscience, empirical evidence from the anatomy of the zebrafish oculomotor integrator and the fly head-direction circuit in the past few years shows that nature has used precisely the hypothesized constructions proposed in simple circuit models to build some integrator networks.

Glossary

Associative memory
The ability to remember and recall the relationship (association) between arbitrary items or concepts.

Autonomous
Characterized by time evolution through internal dynamics, without external driving forces.

Eccentricity
The degree of deflection of the gaze in the horizontal plane relative to a neutral centred position.

Error backpropagation
A procedure for updating the weights of all layers in artificial neural networks (ANNs) based on gradients of an objective function.

Euclidean
A space where it is possible to construct an orthogonal coordinate system and define a particular metric structure.

Hippocampal replay
Ordered sequences of place cell activity during rest or sleep, typically corresponding to sequences that occurred during normal behaviour or their time-reversed counterparts.

Homeostatic plasticity
Plasticity mechanisms that maintain the state of a system by counteracting induced changes.

Hopfield networks
Content-addressable associative memory networks, in which distributed activity states are stabilized as attractor states by synaptic weights using Hebbian learning.

Nearest-neighbour computation
Identifying the closest target out of a set of target states from any starting state, where closest is usually defined by a standard distance metric (for example, Euclidean or Hamming).

Nonlinear neurons
Neurons with input–output response relationships that are nonlinear; that is, the change in the output is not directly proportional to the change of the input.

Non-trivial attractor states
Any attractor states other than the null activity state.

Persistent activity
Maintenance of the firing rate of a neuron about a non-trivial value after removal of the stimulus that induced elevated firing, for durations that exceed the membrane time constant.

Positive feedback
Interactions between elements in which increasing the level of one element increases the level of the other. Positive feedback includes mutual excitation and disinhibition or inhibition of one's inhibitor.

Presynaptic facilitation
A form of short-term synaptic plasticity where the effect of presynaptic activity on the post-synaptic response is enhanced following recent presynaptic activity.

Simple cells
Neurons in the primary visual cortex (V1) of many vertebrate species that respond strongly to oriented edges and gratings of a particular spatial phase.

State space
The coordinate system in which each dimension corresponds to one of the variables of the dynamical system; often, the space is approximated by the spike counts of single neurons.

Synaptic hypothesis
The hypothesis that synaptic change is the substrate of learning and memory in the brain.

Symmetric weight matrices
Weight matrices W that satisfy $W^T = W$; that is, that are invariant to reflection of their entries about their diagonal.

Turing pattern formation
A dynamic process dependent on positive feedback in which a spatial pattern of a particular wavelength is amplified whereas others are suppressed.

Unsupervised
Characterization of the structure in data without any prior training data that contains information about the relationship between the data and external variables.

all noise orthogonal to the manifold is corrected; thus, there is a net reduction of the effects of noise by the factor $\sim\sqrt{K/N} \ll 1$ (refs.[45,55]). However, all states on the attractor manifold are neutrally stable, so the state can drift along the attractor. As such, components of noise along the K attractor dimensions are not internally corrected and cause an accumulating drift away from the initial state, with variance proportional to KT/N, where T is the elapsed time[15,45,55,56]. Thus, through the $1/N$ decrease in variance, even continuous memory states can be well stabilized in sufficiently large attractor networks.

Although content-addressable long-term memory and error reduction can be instantiated through feedforward computations involving only a few steps[57–59] in place of attractor dynamics, recurrent attractor dynamics are indispensable for the generation of persistent activity states (and thus for short-term memory through persistent activity[60,61]) and integration, as we discuss below.

Robust classification

When there are finitely many separated attractors (each a discrete attractor or a continuous manifold), states that are not initially on one of the attractors will flow to one of the attractors. An input to the network can then be classified according to the attractor to which the network state flows after initialization by the input. We can now identify inputs based on the attractors they flow to, a mechanism of classification. If the dynamics of the network further correctly assign corrupted versions of an input to the same attractor state as the uncorrupted input, this constitutes robust classification. In other words, the dynamical basins of attraction of the network must align with the Voronoi regions of the attractor states (that is, corrupted inputs that are closest in distance to one of the uncorrupted inputs should flow to that input's attractor through the dynamics and not another). This is approximately the case for attractor networks operating well below capacity, but typically deteriorates when attractor networks are pushed towards their capacity[62].

Integration

Single neurons integrate their inputs, but usually can only do this over the timescales associated with their membrane capacitances, typically 10–100 ms. Continuous-attractor dynamics can enable neural circuits to integrate over much longer timescales (in the order of about 1–100 s).

A pattern-forming continuous-attractor network requires an additional mechanism to gain the functionality of an integrator: a way to shift the internal state along the attractor in response to an input that encodes changes in the external variable (Fig. 2d, left). Conceptually, the simplest way to build a shift mechanism is by a copy-and-offset construction: construct multiple copies or subpopulations of the

attractor network, each with slightly offset (asymmetric) weights in the sense that active neurons centre their excitation or point of maximal disinhibition slightly offset from themselves on the neural sheet (for example, see that the network in Fig. 1g is a slightly asymmetric version of the network in Fig. 1c). The states in each such network will then form a limit-cycle attractor, with patterns of activity flowing in the direction of the asymmetry in each copy. If opposing copies are coupled together, the pattern is stabilized through a push–pull balance. A velocity input whose components project differentially to the copies will break the push–pull balance, driving the pattern along the flow direction of the more active copy (Fig. 1g). Thus, the total direction and magnitude of the shift of the pattern, corresponding to movement along the attractor manifold, represents the time integral of the velocity input to the network. This common principle unifies the mechanisms across diverse integrator models[12,13,15,63,64].

Decision-making

If, instead of a velocity signal, the input to an integrator network consisted of temporally varying positive and negative evidence in support of each of two options[65] (Fig. 2d, right) (or in the case of multiple options, evidence vectors instead of velocity vectors[66]), the network would integrate those inputs and thus perform evidence accumulation.

Decision-making can be viewed as a selection process applied to an integrator that is based on a readout that detects when the integrator state has accumulated enough evidence and moved past a decision threshold[56,67]. The selection process can be external to the integrator, in the form of a readout circuit that detects such threshold crossings and outputs the decision. Alternatively, the selection process can be built into the dynamics of the integrator itself, in the form of a more complex attractor landscape, in which the states move along a continuous attractor but, at some point, the continuous attractor gives way to a pair of discrete attractors, towards which the states flow (Fig. 2e). Neural WTA models implement such a hybrid analogue–discrete computation[16,65,66,68–70]. The parameters of WTA networks determine the balance between integration dynamics and competitive dynamics, and thus how well the network integrates later evidence: when the network is tuned to be a perfect integrator, its response to inputs is gradual, and small amounts of evidence cause (reversible) flow along the continuous-attractor manifold. In cases in which competition dominates, the response to evidence is a fast flow towards one of the discrete attractors; beyond a point, the flow is nearly irreversible, leading to rapid decision-making and the discounting of later evidence[71].

Neural WTA networks can leverage specific neural non-linearities to accurately and rapidly (in $\sim\log(N)$ time) make the best decision among N alternatives, even if the presented data are noisy (fluctuating over time around their means)[66,70] and even if the number of options varies over orders of magnitude[66].

Sequence generation

Attractor dynamics can be important for stabilizing another long-timescale behaviour: the generation of sequences. Robust sequences can be constructed as low-dimensional limit-cycle attractors, in which high-dimensional perturbations are corrected while along the attractor, there is a systematic, periodic or quasiperiodic flow of states[72–76]. The attractor property that affords ongoing de-noising is important for preventing spatial dispersion and temporal dissipation of the activity packet during sequence generation.

Similar to the case for stationary attractor manifolds, the small components of noise along the limit-cycle attractors are not correctable and lead to a gradual accumulation of drift, which for sequence generation is manifest as timing variability: the standard deviation in the time of reaching the Tth state in the sequence is predicted to grow as $\sqrt{T}$ for unbiased random drift along the attractor[45].

Evidence of attractors in the brain

Criteria for attractor dynamics

The fundamental predictions of attractor models centre on the state-space dynamics of the circuit, as initially explicitly discussed and tested in refs.[9,15,77,78]. First, a system's states should be found localized at or around a low-dimensional set of states that correspond to the attractors in the state space. Second, a system's state should flow quickly back to the low-dimensional state after perturbation. Third, the set of attractor states — quantified either by direct characterization of the full state space or by the relationships between cells — should be invariant, persisting over time and after removal of tuned input, across conditions, across behavioural states and even when there are induced variations in the mapping from internal states to external inputs[15,77,78]. Fourth, integrator networks should further exhibit the property of isometry, whereby lengths of coding space along a dimension are allocated to equal displacements along a dimension of the external variable. Additional predictions of attractor dynamics models, that are not as fundamental in the sense that they are not theoretically necessary or sufficient but are nevertheless of high importance because they are highly supportive of the mechanisms of attractor dynamics, are anatomical and structural correlates: the existence of low-dimensional physical structures and directly visible symmetries in connectivity between cells.

As we have seen, attractor networks dynamics need not be used by the brain in an autonomous setting: inputs that drive attractor networks can be an important part of their function, for instance in integration and evidence accumulation. Nevertheless, because attractor systems are characterized by their internally generated or autonomous dynamics, putative attractor networks are best tested in conditions that minimize external cues that are time-varying or tuned to provide localized inputs along the putative attractor — that is, in an effectively autonomous setting.

Innovations in recording methods that have made it possible to record multiple neurons simultaneously in animals performing naturalistic behaviours[79–82] have enabled crucial tests of these state-space predictions of attractor models described above. The newest methods provide activity data from thousands of neurons in a circuit[83–85], enabling characterization of the low-dimensional state-space dynamics of whole circuits[19,20,86–88].

When the attractor manifolds have three or fewer dimensions, one can directly visualize them by projecting or embedding the high-dimensional state spaces into dimension ≤3. This can be done using methods such as principle components analysis, multidimensional scaling, tensor factorization or other linear methods for projection; or Isomap, locally linear embedding, t-distributed stochastic neighbour embedding, variational autoencoders, latent factor analysis via dynamical systems and nonlinear tensor factorization, among others, for nonlinear embedding[89–92]. These methods can also be useful when manifolds have dimension ≥3 but are topologically simple[88,93]. For topologically non-trivial structures (such as rings and tori), especially those of dimension ≥3, topological data analysis methods become important[19,20,94–98].

Testing the first, second and third predictions of attractor models described above requires examination of the state-space structure of the population, rather than the more conventional characterization of relationships (tuning curves) between cell activity and input or output variables. The most direct way to examine state-space structure is to record enough cells simultaneously that it is possible to characterize the full state-space manifold[19,20,97]. However, the existence, stability and invariance of low-dimensional state-space structures (the first three predictions) can be inferred indirectly from smaller samples of simultaneously recorded cells, for example by characterizing invariant structure in pairwise cell–cell relationships, as has been successfully done in several studies[77,78,99–102].

The existence and stability of low-dimensional state-space structures are necessary but not sufficient for identification of recurrent attractor dynamics in a target network. First, if the behaviours, circuit fluctuations and inputs to the network are themselves low-dimensional, then any observed low-dimensionality of the circuit states may be ascribed to those inputs and reveals little about intrinsic constraints imposed by the circuit. Second, even if inputs and behaviours are high-dimensional, a low-dimensional feedforward projection into the target network would generate low-dimensional states, and high-dimensional perturbations to the circuit would not persist. The essential, defining prediction of attractor dynamics is that of invariance: because the states are internally generated and stabilized by strong recurrent connectivity, the population states and cell–cell relationships should be invariant when probed across time and across various input conditions, including when tuned input is removed and across waking and sleep. In simple terms, the stable low-dimensional states should be invariant across a broad range of conditions[15,78].

Next is the question of circuit localization: does a circuit exhibiting the key signatures of attractor dynamics give rise to these dynamics, or are they a readout of some other region? Localization need not be a primary goal of establishing attractor dynamics: an important problem is to simply characterize whether the brain solves certain problems through attractor dynamics, regardless of which local circuits create these dynamics. Nevertheless, the persistence of activity states in attractors can lend a helping hand to localization efforts. If a region gives rise to or is upstream (but not downstream) of the attractor dynamics, perturbations that alter its state along the set of attractors should persist after the perturbing drive is removed[103].

As we describe next, theoretically motivated analyses of population activity data have firmly established that low-dimensional attractor dynamics are ubiquitous in the brain, across levels in the brain's hierarchy and across species.

Discrete attractors

Up and down states. The simplest example of non-trivial discrete attractor dynamics (that is, beyond a single point attractor) is bistability. Bistable dynamics are a feature of cortical activity in the form of up and down states[4,104–108], in which the subthreshold membrane potential of neurons switches between a hyperpolarized state and a relatively depolarized one, with long persistence (in the order of hundreds of milliseconds to seconds) per state (Fig. 3a). The two states are relatively invariant over time, as seen in the relatively sharply peaked histograms (Fig. 3a), and despite presumed internal noise in the system the peaks are well separated, suggesting relatively rapid corrective dynamics towards the two states. There is little evidence of a strong contribution from cellular bistability in supporting these states, suggesting that it is a network-driven phenomenon involving self-excitation and global inhibition[4,104,105,107–111]. Transitions are believed to be driven through adaptation (from up to down) and by stochastic as well as external coordinating events (from down to up)[106]. Although these states and switches can occur in the cortex without input from the thalamus and striatum, they tend to be synchronous across the cortex and striatum[112,113]. Thus, the origin of up and down states may be highly distributed.

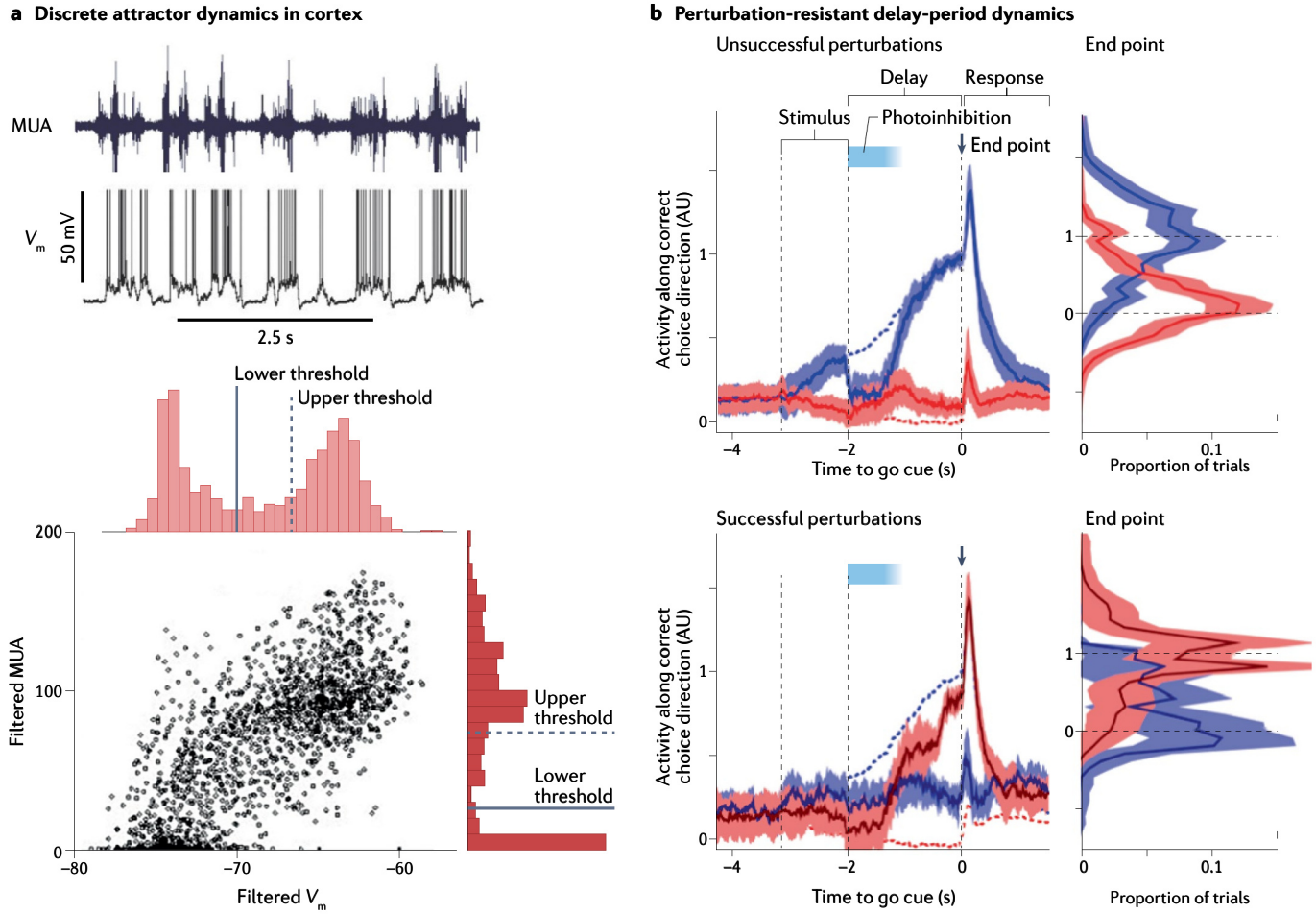
Perceptual bistability. Visual and auditory percepts including binocular rivalry, the Necker cube and some auditory illusions[114–120] offer clear examples of bistability in neural processing, suggesting the operation of a dynamical system with two attractors. In these illusions, the brain (at the level of perceptual reports) selects one possible interpretation of an ambiguous input, often switching between possibilities. Although the phenomenon has long been known and studied, no localized bistable attractor circuit has been identified as the basis of perceptual bistability. Indeed, some percepts may involve top-down activation and modulation of activity across many brain areas[118], suggesting once again a widely distributed circuit for bistability.

Bistability in a premotor area. Recent studies identify and localize discrete attractor dynamics in a mouse premotor area, the anterior lateral motor cortex (ALM)[121–124]. In a cued two-alternative delayed response task, ALM neurons exhibit persistent activity over a 1-s delay period. During the post-cue delay period, activity evolves towards one of two states that guide the response (Fig. 3b), fulfilling the first prediction of attractor dynamics. The delay-period terminal states are similar for cues from different sensory modalities[125], partially meeting the prediction of invariance. ALM perturbations during the delay are either erased (corrected) by the circuit (Fig. 3b, top) or drive a jump to the opposite state (Fig. 3b, bottom), which results in the animal making the wrong action, suggesting bistable switching dynamics similar to the mechanism shown in either Fig. 1b or Fig. 2e.

Given the long training time required for the task and the resulting tailoring of the ALM dynamics to the specific task structure — bistability for a two-choice task — it is likely that this system acquires its dynamics through slow plasticity and, thus, that the network's recurrent structure is malleable in adult animals. New results showing the existence of small (on the scale of about 100 µm) clusters of locally recurrent neurons in the ALM that can maintain persistent responses to microstimulation[126] may provide experimental evidence of the theoretically posited mixed modular networks (below) that are hypothesized to support robust and high-capacity memory states[62].

Discrete multistability. Hopfield networks and WTA networks[69,127–134] (which can be viewed as a special type of Hopfield network, with bistable switch networks as a special type of WTA network) are models of multistability beyond bistability.

At present, the evidence for discrete multistability as a circuit-level brain process is less direct and less exhaustive than that for continuous-attractor networks (described below). However, there are many likely candidate systems and brain regions with dynamics that are suggestive of and consistent with discrete multistability, at least of the special case of WTA attractor dynamics — including in the mammalian hippocampus and auditory cortex, and in the fly and mammalian olfactory system[132–137]. In particular, many of these circuits exhibit global inhibition that clearly narrows and refines activity in the circuit (Fig. 3c, left), and also show evidence of selective recurrent excitation that leads to multiple distinct and stably correlated input


a Discrete attractor dynamics in cortex
MUA
V_m
50 mV
2.5 s
Lower threshold
Upper threshold
200
100
0
Filtered MUA
−80
−70
−60
Filtered V_m
Upper threshold
Lower threshold
b Perturbation-resistant delay-period dynamics
Unsuccessful perturbations
Delay
Response
Stimulus
Photoinhibition
End point
Activity along correct choice direction (AU)
1
0
−4
−2
0
Time to go cue (s)
End point
1
0
0
0.1
Proportion of trials
Successful perturbations
Activity along correct choice direction (AU)
1
0
−4
−2
0
Time to go cue (s)
End point
1
0
0
0.1
Proportion of trials

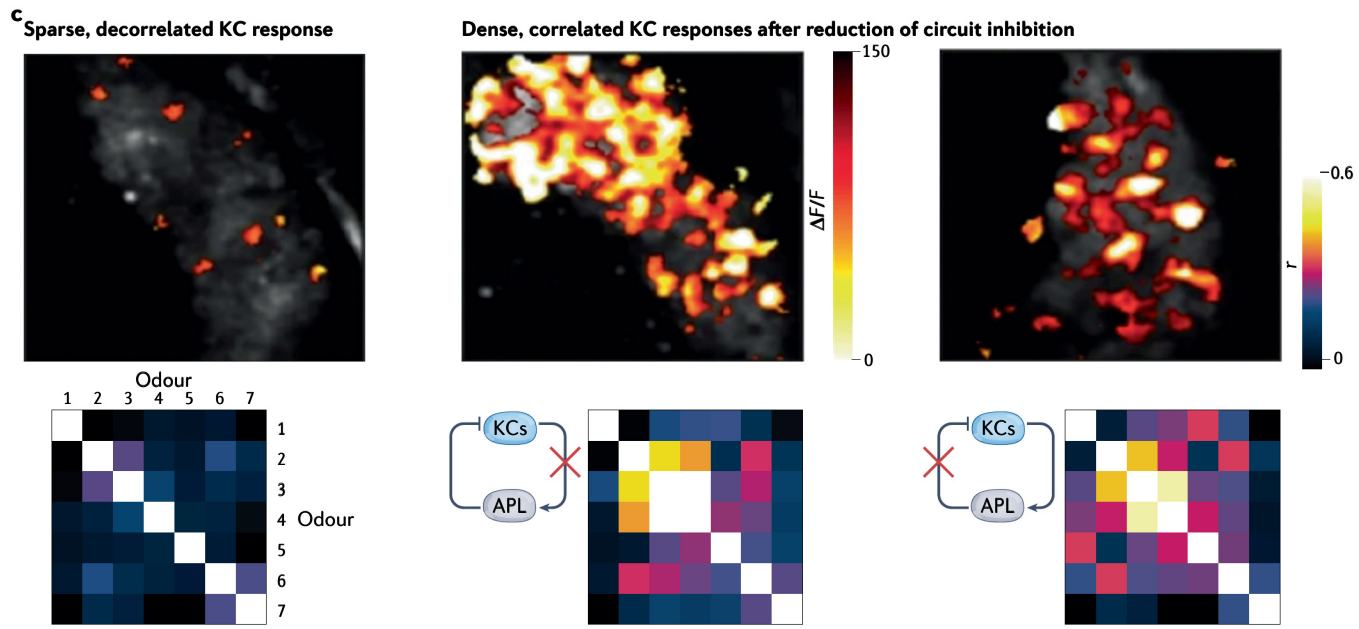

c Sparse, decorrelated KC response
Dense, correlated KC responses after reduction of circuit inhibition
150
ΔF/F
0
−0.6
r
0
Odour
1 2 3 4 5 6 7
1
2
3
4 Odour
5
6
7
KCs
APL
KCs
APL

Fig. 3 | Evidence of discrete attractor dynamics in the brain. a, Multi-unit activity (MUA) and single-unit activity (V_m) during cortical up states and down states show signatures of bistability (clusters and histograms at bottom). **b**, Delay-period dynamics in rodent premotor area (anterolateral motor cortex (ALM)) during a binary decision task (blue and red correspond to correct and incorrect direction choices, respectively). Before the animal makes a motor report of its decision (at the 'go' cue delivery), ALM activity seems to converge to one of two discrete end points (blue and red curves and histograms, top). Perturbations (optogenetic inhibition, denoted by pale blue) are either robustly erased (top; dashed lines show the unperturbed trajectory, and solid line shows a return to the unperturbed trajectory) or flip the dynamics so that the end points are reversed (bottom) and the animal reports the incorrect decision. **c**, Evidence of all-to-all inhibition and competitive winner-takes-all (WTA) recurrent dynamics in the fly olfactory system. Kenyon cells (KCs) activate anterior paired lateral (APL) inhibitory neurons, which in turn globally inhibit KCs. KC responses to odours, when input from the APL neurons is intact, are sparse: top-left image shows calcium fluorescence responses of KCs to odorant isoamyl acetate. KC responses are also decorrelated across odours (left). Blocking either KC drive to APL neurons or APL inhibition of KCs results in dense and correlated odour responses (middle, right). Part **a** adapted with permission from ref.[264], Society for Neuroscience. Part **b** adapted from ref.[121], Springer Nature. Part **c** adapted from ref.[136], Springer Nature.

responses in distinct subpopulations of cells (Fig. 3c, middle and right)[132–137]. In our view, it is likely that these circuits exhibit multiple discrete attractor states, but quantitative testing of the first three predictions of attractor dynamics and direct demonstration of these states as stable and invariant remain an important future direction for characterizing these circuits.

Continuous attractors

The oculomotor integrator. The oculomotor integrator, together with the head-direction circuit, was one of the first systems in neuroscience to be studied theoretically[8,9,138] and experimentally[139] as a continuous-attractor network – specifically as a line attractor (Fig. 1e). This network, which is presynaptic to the motor neurons that control horizontal eye position, is highly conserved across vertebrates, from fish[139,140] to primates[141,142]. It integrates pulse-like saccadic eye movement-command signals to generate step-like stable muscle tension command signals (Fig. 4a) that persist autonomously at graded activity levels after removal of the movement cue and even in the dark in the absence of visual feedback (Fig. 4b; third prediction), and thus enable stable gaze fixation at various degrees of eccentricity. Saccadic inputs knock the system slightly off the linear response states, but the neural responses rapidly decay back towards the persistent firing states (in line with the second prediction). Remarkably, the same system also integrates smooth head-velocity signals to permit gaze stabilization during head movement.

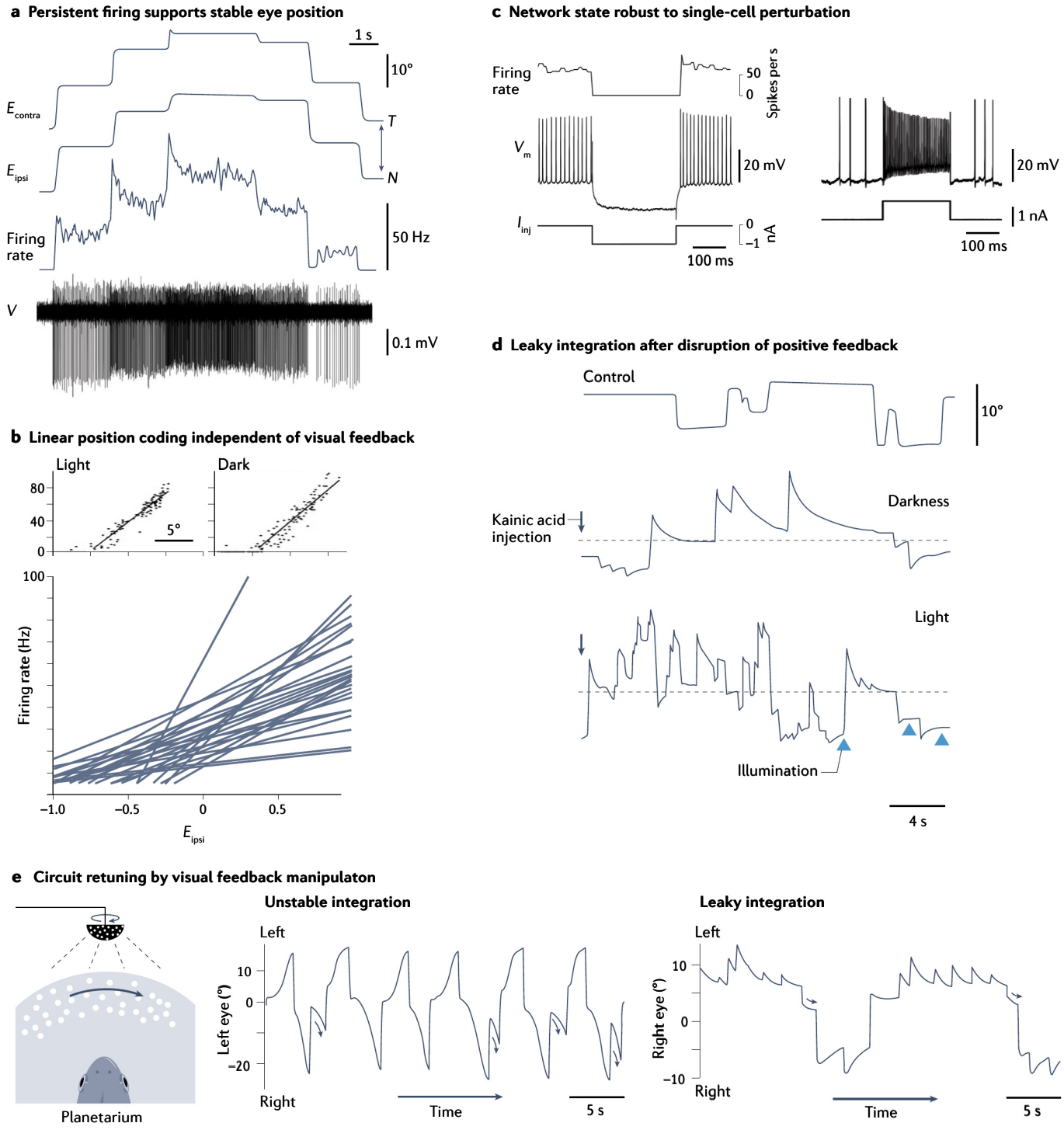
Integration functionality is a network-level rather than single-cell process: single neurons do not generate persistent responses to transient current injections (Fig. 4c, inset), whereas decreasing network feedback through the use of synaptic blockers reduces the time constant of integration and results in a leaky integrator[143] (Fig. 4c). It is possible to reduce or increase network feedback through training with a virtual surround that generates an artificial retinal-slip percept (Fig. 4d), implying that the system is capable of error-driven fine-tuning to maintain a high degree of persistence[144]. Finally, a recent electron microscopy reconstruction[145,146] finds recurrent synaptic interconnectivity between integrator neurons, with excitatory connections between ipsilateral neurons and primarily inhibitory contralateral projections, in excellent agreement with line-attractor models of the oculomotor circuit[9] (Fig. 1e).

Head-direction cells. Some of the earliest experiments to suggest the existence of low-dimensional continuous-attractor dynamics were done in the rodent head-direction circuit[77,99,147] (Fig. 5a,b). The head-direction circuit in mammals maintains an updated internal compass estimate of the heading direction, relative to some arbitrary external reference, as animals move around. It does so by integrating internal rotational velocity estimates during navigation and incorporating information from external cues[148–152]. The head-direction circuit is modelled as a ring-attractor network[10,12,13,17,64] (Fig. 1c,g, left). Before large population recordings became available, cell–cell correlations established that the network states remained invariant on a very low-dimensional manifold across environments[77,99,147] (Fig. 5a), in line with the first and third predictions. The complete set of states of the several thousand-neuron mammalian head-direction network was shown to consist solely of a one-dimensional ring[19,97] (Fig. 5b) (in line with the first prediction), revealing that the brain has completely factorized its navigational representations to dedicate a circuit only to head direction. Furthermore, intervals in the state-space ring manifold map isometrically to intervals of head direction (in line with the fourth prediction), as evidenced by a close match between the isometrically parameterized internal ring states and the measured head direction (Fig. 5b, inset and right).

After natural perturbations away from the ring attractor, the activity of the head-direction circuit flowed back to it[19] (Fig. 5d), meeting the second prediction, and the ring manifold was invariant across waking and rapid eye movement (REM) sleep[19,97] (Fig. 5e), meeting the third prediction. These findings explicitly validate the most fundamental predictions of ring attractor models and continuous attractor-based integrators, providing (together with the grid cell system; see below) the most direct and compelling evidence of continuous-attractor dynamics in the brain.

In a striking example of convergent evolution[151,153], *Drosophila* compute head-direction estimates using apparently very similar dynamics to mammals[148,152,154,155]. The fly neural compass circuit is topographically organized such that the neuropil forms a physical ring-shaped structure in the ellipsoid body, with a local moving activity peak that tracks head direction as the fly turns (Fig. 5f). Other notable advantages of the fly circuit in the effort to characterize its mechanisms are that the number of neurons is small and their morphology and connectivity have been fully traced[156] (Fig. 5g). This detailed view of the circuit permits quantitative, not just qualitative, comparisons with ring-attractor models.

The combined activity and connectivity data reveal that the fly head-direction system quite literally implements the copy-and-offset double-ring network architecture that has been proposed for velocity integration[13,157]. However, the dimensionality of the fly head-direction circuit and its full state-space dynamics remain to be characterized. Notably, although the circuit is organized physically as a ring network, recent evidence suggests that the insect head-direction circuit may be involved in performing two-dimensional path integration as well[158,159]. Thus, unlike the anterodorsal thalamic nucleus network in mammals, the insect head-direction circuit may not be confined to a one-dimensional



ring of attractor states that fully factorizes out the representation of head direction in its representation of spatial variables.

Finally, the head-direction system of both insects and mammals can be re-anchored and reset based on tuned external cues[148,152,160], and this can change the orientation tuning curves of cells and moment by moment firing rates of cells in a way that remains consistent with the third prediction for attractor dynamics.

Grid cells. A grid cell encodes spatial location through a periodic triangular-lattice discharge pattern that tiles explored two-dimensional spaces[161]. Grid cell phases update during movement in the light and in the dark[161] to reflect the animal's current position, as a two-dimensional phase. Continuous-attractor models of grid cells are based on collective Turing pattern formation[15,162,163], explain their velocity integration function and predict that grid cells should exist in large sets with

Fig. 4 | Linear attractor dynamics generated by network feedback in the oculomotor integrator. a, In the goldfish, the positions of the ipsilateral and contralateral eyes (E_{ipsi} and E_{contra}, respectively) can be maintained for a stable horizontal gaze during inter-saccadic fixation at different angular positions (top two traces). This is supported by stable steps in firing rate by oculomotor integrator neurons (bottom two traces show extracellularly recorded firing rate and voltage (V)), which integrate transient (in the order of about 100 ms) saccadic command bursts. **b**, Oculomotor neurons drive eye position with linearly ramping tuning curves (bottom). Their responses are the same in the light and the dark (top), and thus do not depend on visual input for gaze stabilization on the timescale of seconds. **c**, Transient current injection into individual oculomotor neurons results in only a transient (that is, not persistent) decrease (left) or increase (right) in firing rate, consistent with lack of a cellular origin for persistent intersaccadic firing. **d**, Injection of kainic acid into the oculomotor integrator produces leaky dynamics in horizontal eye position, consistent with network models. The leak is pronounced in the dark and is still present although reduced, presumably because of visual feedback, during illumination (triangles). **e**, Visual training (here, from the motion of dots of light in a planetarium-like set-up) that mimics leaky or unstable eye positions in goldfish can mistune the oculomotor integrator, making it unstable or leaky, respectively. Arrows highlight fixations following saccades towards the mid position. Part **a** is adapted from ref.[139], Springer Nature. Part **b** is adapted with permission from ref.[265], American Society of Physiology. Part **c** is adapted from ref.[139], Springer Nature. Part **d** is adapted with permission from ref.[266], Wiley. Part **e** is adapted with permission from ref.[144], National Academy of Sciences, USA.

identical spatial periodicity and orientation, but tile all possible two-dimensional phases. As with the first general prediction of continuous-attractor models, they specifically predict that the population states of such a set of cells should be confined to merely two dimensions along a torus-shaped manifold that remains unchanged across environments and behavioural states[15] (Fig. 1d, rightmost column).

Analyses of simultaneously recorded grid cells with similar periods revealed that their periods and orientations are identical down to estimation noise (thus defining a discrete population, subsequently called a 'module'[164]) and that they tile all possible two-dimensional phases[78,165], strongly suggesting a two-dimensional torus in line with the first prediction. Moreover, the relative firing phases and grid parameter ratios of co-modular cells are tightly conserved even as the spatial tuning of cells varies across time and environments[78] (Fig. 6a), with the dimensionality of the spatial environment[166] (Fig. 6b) and with large environmental rescaling-driven deformations of grid tuning[78], confirming the prediction of invariance. In addition, the detailed cell–cell relationships seen in waking exploration that define the low-dimensional response of a grid module are conserved across overnight sleep in grid cells but not in place cells[101,102] (Fig. 6c), establishing that the low-dimensional states are autonomously generated. In line with all of the fundamental predictions of continuous-attractor dynamics[15], these findings established that each grid module's response is very low-dimensional; is invariant across environments, time and behavioural states; and is internally stabilized and autonomously generated. Most recently, these findings were confirmed by large-scale recordings of grid cells that made it possible to directly characterize the grid cell population response by applying the topological analyses of state-space structure pioneered earlier[19,97] to grid cells (Fig. 6e), directly illustrating the low-dimensional, toroidal and invariant state-space structure of grid cell modules[20].

A corollary is that the grid cell response is not derived from upstream place cells, which remap across environments and during sleep (Fig. 6c): as shown in ref.[101], this finding renders models in which the place cell response is primary to grid cells[167–169] inconsistent with the data. Another corollary of the population states of grid cells remaining strictly preserved[20,78,170], even when their spatial tuning curves in two-dimensional and three-dimensional environments are altered so they do not form equilateral triangular grids[170–176], is that these variations must result from changes in how the invariant internal states are mapped to external states. Such changes may arise from, for example, alterations in velocity estimation[15,78] that stretch the grid or from external cues that shift the phase of the grid cell network[177–180], rather than because of alterations in the internal grid network dynamics.

Despite having periodic representations, and thus each only representing position as an ambiguous two-dimensional phase, collectively grid cells form a discrete set of modules with distinct but similar periodicities[164]. This allows grid cells to unambiguously represent position over a scale that grows exponentially in the number of grid modules[131,181].

In sum, the head-direction cell and grid cell systems show that the same pattern formation principle — based on local excitation or disinhibition, with broader inhibition — that is pivotal for morphogenesis in plants and animals[38] is also fundamental to the genesis of stationary continuous-attractor states for computation and representation in the brain.

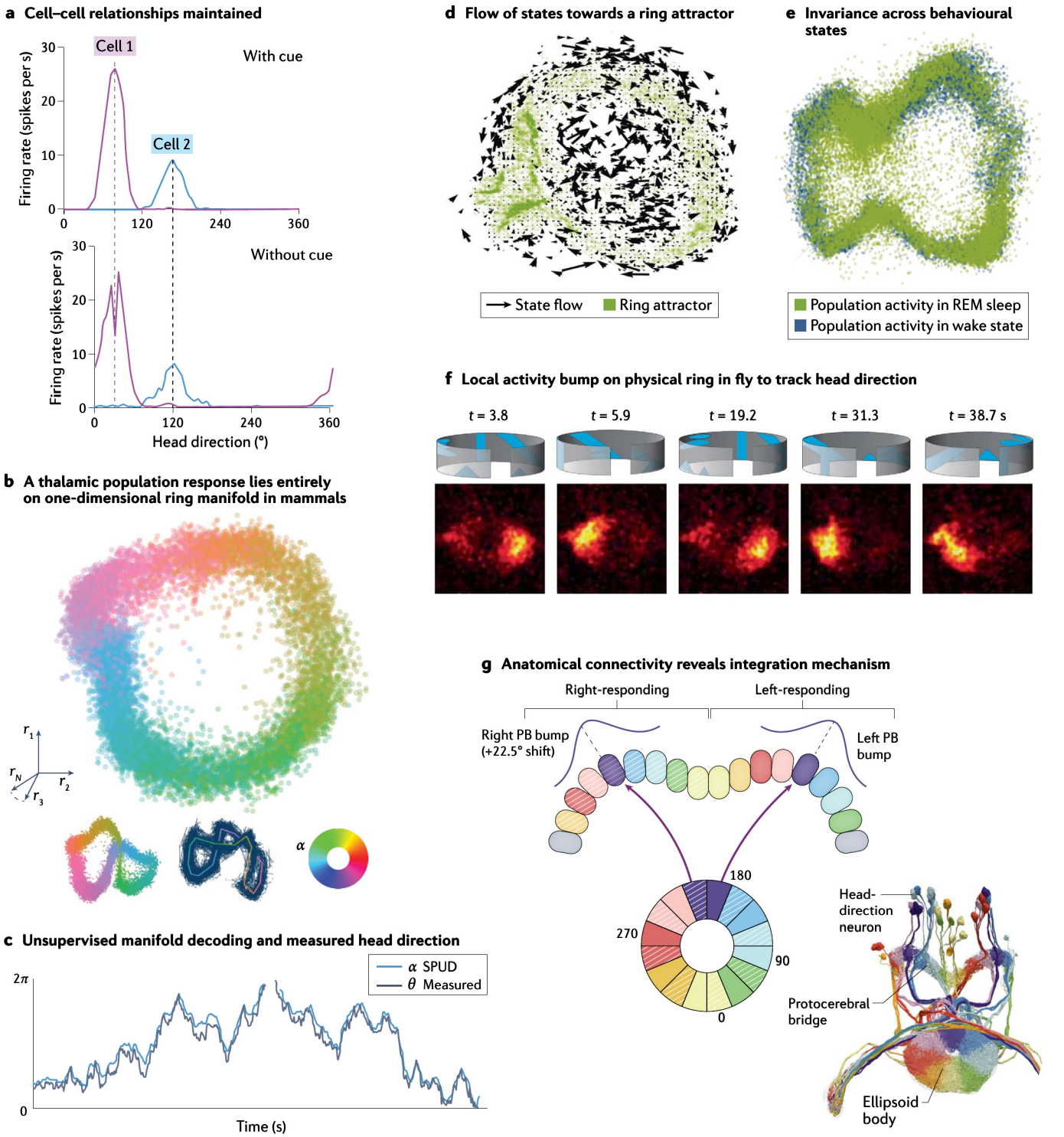
Graded working memory networks. In monkeys trained to saccade to a remembered cued location (selected from a set arranged in a circle), cells in the prefrontal cortex and posterior parietal cortex exhibit persistent activity across the delay period that is selective for the direction of the cue, consistent with the first and third predictions of attractor dynamics[182,183]. The delay period activity in the prefrontal cortex is a bump that moves apparently randomly along a one-dimensional manifold with the characteristics of a diffusion process[87]. Thus, the variance in bump location grows linearly with time during the delay, as predicted by continuous-attractor models[15,19,55], but the bump profile remains largely invariant (first and second predictions). Bump movement predicts subsequent behavioural errors[87], suggesting that these states are repositories or read-outs of the memory.

The need for extensive training and the resulting tailoring of the attractor states to this specific but not naturally encountered multi-cue task suggests that this attractor forms through learning in a flexible system. We might therefore also expect a loss of the neural correlation structure if the animal is subsequently trained on other tasks, unlike with the grid and head-direction cell networks.

Limit-cycle attractors

The CNS and peripheral nervous system contain numerous instances of periodic dynamics, from the spiking of single neurons[184,185] to circadian rhythms and sleep-cycle generation[186], to rhythmic activity in motor circuits. The amplitude of a linear oscillator is set by the initial condition (for example, the height at which a pendulum is released), whereas limit-cycle oscillators have an invariant intrinsic amplitude. Thus, oscillations that decay or whose long-term amplitude or frequency changes after transient perturbation are not limit cycles.

Many of the oscillations noted above maintain their amplitude over time and, given their robustness, are probably generated through



attractor dynamics. Experimentally well-characterized examples of sustained periodic dynamics are central pattern generators in spinal motor circuits that drive swimming, crawling, walking, breathing and digestion; these differ in specifics across species but have common principles of mechanism and operation, including high robustness[187,188]. Central pattern generator circuits typically integrate external feedback, but can operate in isolation without external drive[189]. However, driven (non-autonomous) systems could exhibit limit cycles that are attributable to their inputs rather than to intrinsic attractor dynamics[190].

Given the sizeable literature on these topics, we refer the reader to some excellent papers and reviews[186,191–195].

Fig. 5 | The head-direction circuit: a ring attractor in the brain. a, Activity of two cells in the rat head-direction circuit during free foraging in a two-dimensional circular arena with a globally orienting cue (top). When the cue is removed (bottom), the fields rotate, but the cells maintain their tuning shapes and relative tuning angles (pale curves show the cells' activity from the top plot, but globally rotated). **b**, The population-level states of the anterodorsal thalamus during free-foraging and other natural behaviour in a two-dimensional environment, shown through nonlinear embedding in two dimensions and independently validated by topological data analysis, are confined to a one-dimensional ring (as in Fig. 1c). Inset: another view of the same ring manifold in three dimensions (left). The manifold is colourized based on a computational approach called SPUD (spline parameterization for unsupervised decoding)[19]: the manifold is fit by a spline of matching dimension and topology (middle), and the spline is parameterized isometrically; equal changes in parameter value for equal distances along the manifold (right). Parameter changes are indicated by colour. **c**, There is a close match between unsupervised isometric parametrization of the manifold from part **b** and the externally measured head direction of the rodent. **d,e**, The same cells as in part **b** were recorded during rapid eye movement (REM) sleep (green): the states during REM sleep remain confined to a one-dimensional ring that precisely overlays the ring of waking states (blue, part **e**), and states off the ring exhibit large flows (black arrows) back towards the ring (part **d**). **f**, Calcium imaging of activity in the physically ring-shaped *Drosophila* ellipsoid body reveals a localized bump of excitation that follows the movement of a cue in the fly's visual field. **g**, A combination of electrophysiology and electron microscopy imaging of the central complex in flies[267] has provided detailed layout and connectivity data for comparison with predicted connectivity in ring attractor models. Part **a** adapted with permission from ref.[77], Society for Neuroscience. Parts **b**–**e** adapted from ref.[19], Springer Nature. Part **f** is adapted with permission from ref.[154], Science/AAAS. Part **g** is adapted with permission from ref.[267], CC BY 4.0 (https://creativecommons.org/licenses/by/4.0/).

Departures from attractor dynamics

Not all circuits hypothesized to exhibit low-dimensional attractor dynamics seem under further experimentation to do so, or currently lack sufficient evidence to establish such dynamics in the circuit. We discuss three such examples.

Orientation tuning in visual cortex

The circuit of simple cells in the primary visual cortex (V1) satisfies some key properties of attractor networks[10]: V1 and V2 cells exhibit orientation-tuned responses to real and illusory edges[196–198], and in V1 the activity of neurons with similar orientation tuning is correlated during spontaneous activity[199]. However, changing the state of an attractor requires strong inputs and is slow[200,201], inconsistent with the need for perceptual systems to respond sensitively and rapidly[202]. Moreover, the responses to illusory edges in V1 tend to occur at longer latency than responses to real edges, suggestive of top-down inputs rather than within-V1 dynamics. These observations lend weight to the possibility that responses might be dominated by feedforward drive[196,203], potentially with non-normal amplification processes[52,204]. Quantitative characterizations of response speed will be important to draw clear conclusions about V1 circuit dynamics.

Place cells

Place cells form stable representations of space[205] that can persist in the dark[206] and shortly after the animal has fallen asleep[207,208]. In any particular environment, the population response lies on a low-dimensional manifold in state space[88]. Accordingly, the place cell circuit has been modelled as a continuous-attractor network[209] with one or multiple overlapping maps[210], whereby each map is a different assignment of cells to spatial locations. However, the storage of multiple high-resolution maps in a homogeneous attractor network severely limits capacity[181,211–213]. Cell–cell correlations are not preserved across environments, as implied by the phenomenon of remapping[101,102,207,214,215]. Similar to V1 neurons, place cells might be better described as deriving their tuning by forming conjunctions between multiple feedforward inputs, including those from grid cells and cells that encode external cues such as borders, landmarks and reward sites[59,131,213,216–218]. At the same time, place cells exhibit sequential activation of previous trajectories during activity hippocampal replay[208,219–221]. This sequential activation is hypothesized to be generated by recurrent connections in hippocampal area CA3, suggesting that recurrent and feedforward dynamics may collaborate in the generation of place cell states; more recent models are beginning to capture this interplay[59,218,222]. Closing the book on the question of autonomous low-dimensional dynamics in what, in our view, is the far more complex response of place cells than grid cells requires more detailed experimentation, analysis and modelling.

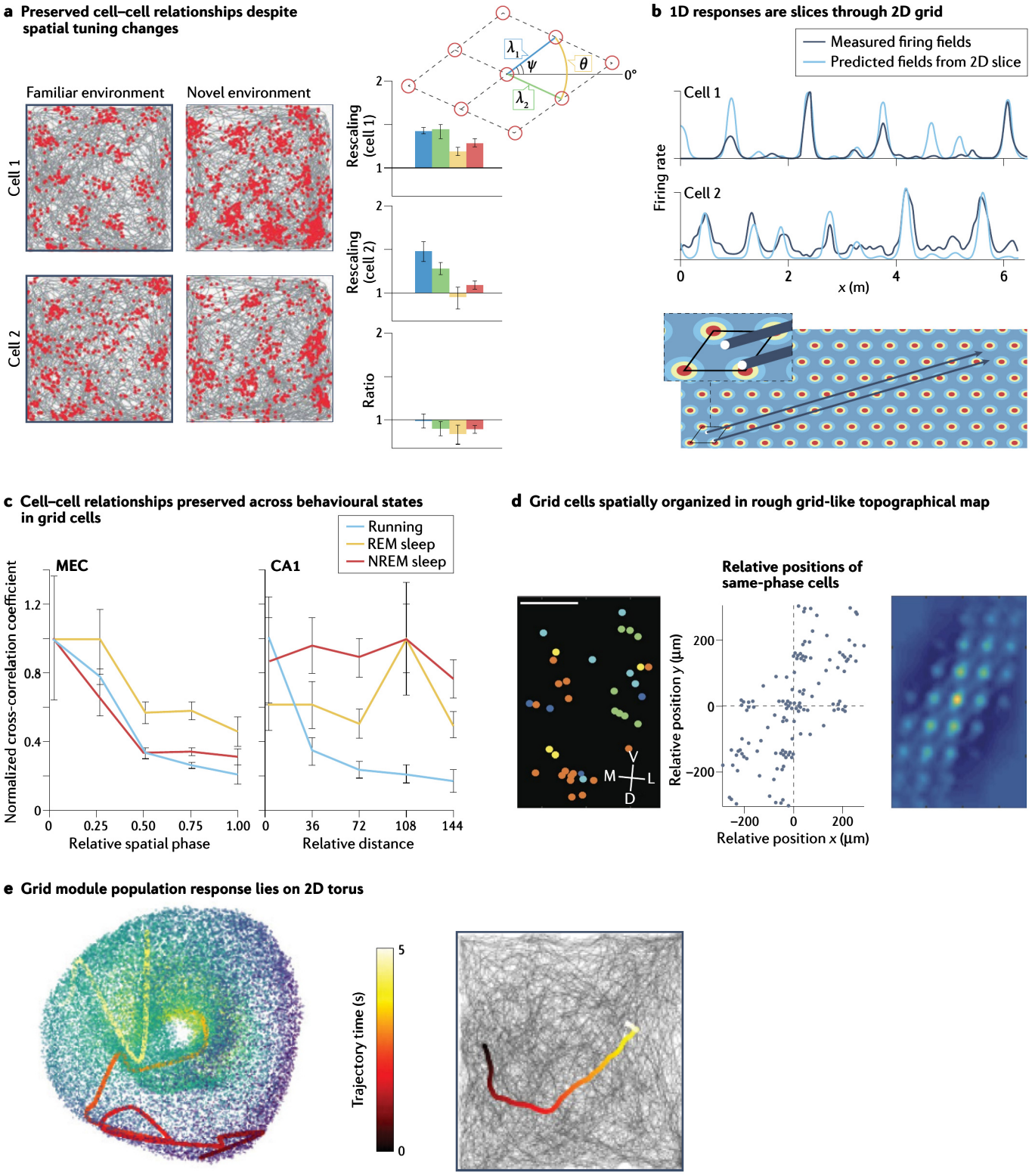
Motor cortical trajectories

Finally, recordings of motor cortical activity during stereotyped arm movements in primates reveal the existence of stable low-dimensional trajectories[86,223–226], similar to the trajectories in state space that were originally characterized in olfactory circuit responses to different odours[227]. Limit cycles and other low-dimensional attractors have been hypothesized to have a key role in cortical movement generation[228,229]. The behaviours typically performed during these neural recordings are themselves restricted to be stereotyped and low-dimensional, and thus it remains unclear whether activity would remain equally low-dimensional across richer behaviours (for example, over the set of all possible arm movements). Recent evidence from perturbation experiments[190] suggests that neural trajectories in the motor cortex during skilled movements are driven by input from the thalamus, and thus that the circuits for motor pattern generation in the CNS might be distributed across multiple brain regions. Characterizing the intrinsic dimensionality of motor cortical activity, and determining whether the command to make more-complex motions involves multiple upstream or distributed primitive attractors, remain important open questions for both clinical brain–machine interfaces and neuroscience.

Flexibility despite rigidity

The attractor networks we have described in this Review are typically rigid across time and conditions. However, recent experimental and theoretical work has suggested that low-dimensional and rigid attractor states could be reused and recombined to create versatile and efficient systems for representation and computation in new situations.

Building a representation (Fig. 2a) could proceed by painstakingly constructing a large set of associative feedforward correspondences, equivalent to a look-up table. By contrast, an attractor that is an integrator requires only two feedforward correspondences: an anchoring process that identifies one external state to one internal one, and then an association of external movement-based velocities with the internal shift mechanism in the integrator[230] (Fig. 2f). Thus, continuous attractors that are also integrators could enable, for example, the rapid construction[218,230,231] and even inference of states visited for



the first time through a new trajectory[218,230,232], and could be reused to represent multiple variables[230]. Indeed, the brain seems to (re)use grid cells and place cells when navigating in space and in non-spatial domains[233–235]; recent work shows how the dimensionality of the represented variable could be greater than the individual attractor networks[230].

Fig. 6 | Two-dimensional toroidal attractors in the grid cell system. a, The spatial tuning periods and orientations of grid cells reconfigure substantially in novel environments (left: firing patterns of an example pair of grid cells in a familiar and a novel environment), but cell–cell relationships remain the same, as seen from the tight covariance of changes across cells (right), implying an internally generated low-dimensional structure. Each colour corresponds to a variable that describes the lattice of the spatial tuning curve of the cell, as shown in the schematic. **b**, The non-periodic responses of two example co-modular cells (dark blue) on a one-dimensional linear track do not look like simple offsets of one another, raising the question of whether cell–cell relationships have reconfigured and the grid cell dynamics are not low-dimensional and invariant. However, the responses of the cells can be predicted (light blue) as parallel slices through the two-dimensional grid (bottom), and their two-dimensional relative phase offset is predicted by the separation of the one-dimensional response slices, showing that the cell relationships and two-dimensional circuit dynamics are preserved across diverse conditions. **c**, Pairwise correlations between grid cells in the medial entorhinal cortex (MEC) measured during navigation are preserved across overnight rapid eye movement (REM) sleep and non-REM (NREM) sleep, whereas those of place cells in hippocampal area CA1 are not. **d**, Grid cells are anatomically arranged according to their relative spatial firing phases. Left: cell positions in a field of view of the MEC coloured according to the phase of their spatial tuning curves. The relative cortical positions of same-phase cells make a triangular lattice pattern (middle), with a grid-like autocorrelation pattern (right). **e**, The population-level states of grid cells from one module (each dot represents the population state at one point in time) during free foraging in a two-dimensional environment are shown through nonlinear dimensionality reduction and confirmed by topological data analysis to lie on a two-dimensional torus (left) as predicted by models[15]. As the animal follows a spatial trajectory (right), the state moves along the torus manifold (left). Manifold colouring is a gradient along the first principal component of the data. Part **a** is adapted from ref.[78], Springer Nature. Part **b** is adapted with permission from ref.[166], Elsevier. Part **c** is adapted from ref.[101], Springer Nature. Part **d** is adapted with permission from ref.[268], Elsevier. Part **e** is adapted from ref.[20], Springer Nature.

A further line of work has posited that networks composed of modular subnetworks, each an attractor network, enable a given number of neurons to represent an exponentially larger number of representational or memory states[62,131,181,213,236–240] through combinations of states than fully connected, Hopfield-like networks can[241–245]. Although the combinatorial states expressed by the set of attractor networks are not themselves attractors, it is possible to couple together these subnetworks to generate an exponential number of attractor states such that they each have a reasonably sized basin and are thus robust[59,62,222,236,237,242] (Fig. 2). The states in these networks cannot have arbitrary form and content; they are defined by the rigid states of each module. Thus, a crucial question is how they could be leveraged for memory. Such high-capacity sets of attractor states have been shown to provide possible models for high-capacity and robust action selection[62], robust classification[62] and smoothly decaying associative memory[59]. Moreover, the principles described in this paragraph can be combined in a 'mixed modular coding scheme' to represent and store inputs of any dimensionality relative to the individual attractor networks, so long as it is lower than the summed attractor dimension across networks[230], without needing to reconfigure the recurrent network (Fig. 2h). Much of the potential for alternative uses, configurations or combinations of attractor networks remains unexplored and is ripe for further study.

Looking ahead

The theory of attractor dynamics in the brain has provided a powerful and unifying conceptual framework for understanding integration, representation, memory, error correction and efficient learning and inference in the brain. The experimental effort to study candidate attractor circuits and test their predictions has been a fertile field of research, and population-wide physiology techniques have led to breath-taking direct visualizations of attractor dynamics at work in the brain.

The theory is also proving to be a powerful tool in interpreting how artificial neural networks (ANNs) solve complex tasks. ANNs trained to robustly solve memory, integration and decision-making tasks in domains as diverse as spatial navigation, vision and language develop attractor dynamics[46,246–249], suggesting that attractor networks not only are able to solve such problems but also might be necessary when the computing elements are memoryless neurons. Furthermore, equipping ANNs with preconfigured attractor networks can help produce faster, more data-efficient and generalizable learning[59,230,231]. Because ANNs can be trained on complex tasks and then fully examined after learning, they will potentially more readily contribute to the next chapter in our understanding of how continuous-attractor networks can interact and combine with other mechanisms to enable the brain to solve rich problems associated with intelligence.

Notable mechanistic questions about attractor networks also remain open. One avenue may involve moving away from the high firing-rate asynchronous spiking regimens[250,251] to better understand whether low firing-rate synchronous spiking networks might support attractor dynamics — and thus permit a combination of fast timescale dynamics such as spike synchronization and oscillatory phase dynamics[250,252,253]. For continuous attractors, understanding how the brain deals with the problem of fine-tuning in linear networks or the imposition and maintenance of a continuous symmetry across neurons remains unknown and is ripe for resolution[34,254].

A few models of the development of continuous attractors show how they could emerge simply through unsupervised associative plasticity[17,179,210], whereas others are based on combining feedback of known or plausible error signals with neural activity in relatively simple learning rules[17,255,256]. The rest of such models train networks on a high-level goal through error backpropagation, combined with several other constraints on architecture or the form the solutions should take[46,231,247,249,257–259]. As recent work suggests, however, training ANNs to solve tasks is not a panacea for understanding the brain's solutions[260]. All models of attractor network development are incomplete for different reasons: the unsupervised models require uniform exploration of the input variable space and suppression of recurrent weights during their training, whereas backpropagation models do not offer an account of how loss functions, learning and additional constraints might be generated and implemented in biological systems.

There is much left to do in the field and an exciting vista ahead. On the experimental side, tools for high-resolution population-level neural recordings and perturbation across multiple brain areas[84,85,261] enable us to peer further and deeper than ever. On the theory side, future developments will help us conceptualize how such circuits could help underwrite intelligent computation through the formation, interaction and reuse of multiple low-dimensional attractors or attractor-like structures.

Published online: 3 November 2022

References

1. Amari, S.-I. Neural theory of association and concept-formation. *Biol. Cybern.* **26**, 175–185 (1977).
2. Hopfield, J. J. Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl Acad. Sci. USA* **79**, 2554–2558 (1982).
3. Little, W. A. The existence of persistent states in the brain. *Math. Biosci.* **19**, 101–120 (1974).
4. Wilson, H. R. & Cowan, J. D. A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue. *Kybernetik* **13**, 55–80 (1973).
5. Hopfield, J. J. Neurons with graded response have collective computational properties like those of two-state neurons. *Proc. Natl Acad. Sci. USA* **81**, 3088–3092 (1984).
6. Cohen, M. A. & Grossberg, S. Absolute stability of global pattern formation and parallel memory storage by competitive neural networks. *IEEE Trans. Syst. Man Cybern.* **SMC-13**, 815–826 (1983).
7. Amari, S. Dynamics of pattern formation in lateral-inhibition type neural fields. *Biol. Cybern.* **27**, 77–87 (1977).
8. Cannon, S. C., Robinson, D. A. & Shamma, S. A proposed neural network for the integrator of the oculomotor system. *Biol. Cybern.* **49**, 127–136 (1983).
9. Seung, H. S. How the brain keeps the eyes still. *Proc. Natl Acad. Sci. USA* **93**, 13339–13344 (1996).
   This work constructs and pedagogically describes a mathematical theory of line attractor dynamics for the oculomotor system.
10. Ben-Yishai, R., Bar-Or, R. L. & Sompolinsky, H. Theory of orientation tuning in visual cortex. *Proc. Natl Acad. Sci. USA* **92**, 3844–3848 (1995).
11. Ermentrout, B. Neural networks as spatio-temporal pattern-forming systems. *Rep. Prog. Phys.* **61**, 353 (1998).
12. Stringer, S., Trappenberg, T., Rolls, E. & Araujo, I. Self-organizing continuous attractor networks and path integration: one-dimensional models of head direction cells. *Network* **13**, 217–242 (2002).
13. Xie, X., Hahnloser, R. H. R. & Seung, H. S. Double-ring network model of the head-direction system. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **66**, 041902 (2002).
14. Fuhs, M. C. & Touretzky, D. S. A spin glass model of path integration in rat medial entorhinal cortex. *J. Neurosci.* **26**, 4266–4276 (2006).
15. Burak, Y. & Fiete, I. R. Accurate path integration in continuous attractor network models of grid cells. *PLoS Comput. Biol.* **5**, e1000291 (2009).
    For a single module of grid cells, this work construct a faithful continuous-attractor network model based on the principles of pattern formation.
16. Wang, X.-J. Probabilistic decision making by slow reverberation in cortical circuits. *Neuron* **36**, 955–968 (2002).
17. Zhang, K. Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: a theory. *J. Neurosci.* **15**, 2112–2126 (1996).
    This work constructs a continuous-attractor network model of the head-direction system, showing how intrinsic dynamics contribute to shaping population firing rates.
18. Milnor, J. W. Attractor. *Scholarpedia* **1**, 1815 (2006).
19. Chaudhuri, R., Gerçek, B., Pandey, B., Peyrache, A. & Fiete, I. The intrinsic attractor manifold and population dynamics of a canonical cognitive circuit across waking and sleep. *Nat. Neurosci.* **22**, 1512–1520 (2019).
    This work tests and verifies the predictions of continuous-attractor dynamics for the head-direction cell circuit in the anterodorsal thalamic nucleus in rodents by analysing data across behavioural states.
20. Gardner, R. J. et al. Toroidal topology of population activity in grid cells. *Nature* **602**, 123–128 (2022).
    This work using large-scale recordings of several hundred cells verifies predictions of continuous-attractor dynamics in single modules of grid cells.
21. Strogatz, S. H. *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering* (CRC, 2018).
22. Koch, C. *Biophysics of Computation: Information Processing in Single Neurons* (Oxford Univ. Press, 2004).
23. Shadlen, M. N. & Newsome, W. T. The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. *J. Neurosci.* **18**, 3870–3896 (1998).
24. Hanus, C. & Schuman, E. M. Proteostasis in complex dendrites. *Nat. Rev. Neurosci.* **14**, 638 (2013).
25. James, W. *The Principles of Psychology* (Henry Holt, 1890).
26. McDougall, W. On the seat of the psycho-physical processes. *Brain* **24**, 579–630 (1901).
27. Hebb, D. O. *The Organization of Behavior* (Wiley, 1949).
28. Brown, R. E., Bligh, T. W. B. & Garden, J. F. The Hebb synapse before Hebb: theories of synaptic function in learning and memory before, with a discussion of the long-lost synaptic theory of William McDougall. *Front. Behav. Neurosci.* **15**, 732195 (2021).
29. Abraham, W. C., Jones, O. D. & Glanzman, D. L. Is plasticity of synapses the mechanism of long-term memory storage? *NPJ Sci. Learn.* **4**, 1–10 (2019).
30. Takeuchi, T., Duszkiewicz, A. J. & Morris, R. G. The synaptic plasticity and memory hypothesis: encoding, storage and persistence. *Philos. Trans. R. Soc. B: Biol. Sci.* **369**, 20130288 (2014).
31. Martin, S., Grimwood, P. & Morris, R. Synaptic plasticity and memory: an evaluation of the hypothesis. *Annu. Rev. Neurosci.* **23**, 649–711 (2000).
32. Anderson, P. W. More is different. *Science* **177**, 393–396 (1972).
33. Zhang, H., Wang, Z. & Liu, D. A comprehensive review of stability analysis of continuous-time recurrent neural networks. *IEEE Trans. Neural Netw. Learn. Syst.* **25**, 1229–1262 (2014).
34. Renart, A., Song, P. & Wang, X.-J. Robust spatial working memory through homeostatic synaptic scaling in heterogeneous cortical networks. *Neuron* **38**, 473–485 (2003).
35. Itskov, V., Hansel, D. & Tsodyks, M. Short-term facilitation may stabilize parametric working memory trace. *Front. Comput. Neurosci.* **5**, 40 (2011).
36. Turing, A. M. The chemical basis of morphogenesis. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **237**, 37–72 (1952).
37. Cross, M. C. & Hohenberg, P. C. Pattern formation outside of equilibrium. *Rev. Mod. Phys.* **65**, 851 (1993).
38. Koch, A. J. & Meinhardt, H. Biological pattern formation: from basic mechanisms to complex structures. *Rev. Mod. Phys.* **66**, 1481–1507 (1994).
39. Schweisguth, F. & Corson, F. Self organization in pattern formation. *Dev. Cell* **49**, 659–677 (2019).
40. Shraiman, B. Mechanical feedback as a possible regulator of tissue growth. *Proc. Natl Acad. Sci. USA* **102**, 3318–3323 (2005).
41. Sekimura, T., Noji, S., Ueno, N. & Maini, P., *Morphogenesis and Pattern Formation in Biological Systems: Experiments and Models* (Springer, 2003).
42. Gierer, A. & Meinhardt, H. A theory of biological pattern formation. *Kybernetik* **12**, 30–39 (1972).
43. Boucheny, C., Brunel, N. & Arleo, A. A continuous attractor network model without recurrent excitation: maintenance and integration in the head direction cell system. *J. Comput. Neurosci.* **18**, 205–227 (2005).
44. Couey, J. J. et al. Recurrent inhibitory circuitry as a mechanism for grid formation. *Nat. Neurosci.* **16**, 318–324 (2013).
45. Burak, Y. & Fiete, I. R. Fundamental limits on persistent activity in networks of noisy neurons. *Proc. Natl Acad. Sci. USA* **109**, 17645–17650 (2012).
46. Sorscher, B., Mel, G., Ganguli, S. & Ocko, S. A unified theory for the origin of grid cells through the lens of pattern formation. In *Advances in Neural Information Processing Systems* 10003–10013 (NeurIPS, 2019).
47. Khona, M., Chandra, S. & Fiete, I. Spontaneous emergence of topologically robust grid cell modules: a multiscale instability theory. Preprint at *bioRxiv* https://doi.org/10.1101/2021.10.28.466284 (2021).
48. Seung, H. S. Amplification, attenuation, and integration. *Handb. Brain Theory Neural Netw.* **2**, 94–97 (2003).
49. Sompolinsky, H., Crisanti, A. & Sommers, H.-J. Chaos in random neural networks. *Phys. Rev. Lett.* **61**, 259 (1988).
50. van Vreeswijk, C. & Sompolinsky, H. Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science* **274** 5293, 1724–1726 (1996).
51. Engelken, R., Wolf, F. & Abbott, L. Lyapunov spectra of chaotic recurrent neural networks. Preprint at *arXiv* https://doi.org/10.48550/arXiv.2006.02427 (2020).
52. Murphy, B. K. & Miller, K. D. Balanced amplification: a new mechanism of selective amplification of neural activity patterns. *Neuron* **61**, 635–648 (2009).
53. Trefethen, L. N., Trefethen, A. E., Reddy, S. C. & Driscoll, T. A. Hydrodynamic stability without eigenvalues. *Science* **261**, 578–584 (1993).
54. Mongillo, G., Barak, O. & Tsodyks, M. Synaptic theory of working memory. *Science* **319**, 1543–1546 (2008).
55. Compte, A., Brunel, N., Goldman-Rakic, P. S. & Wang, X.-J. Synaptic mechanisms and network dynamics underlying spatial working memory in a cortical network model. *Cereb. Cortex* **10**, 910–923 (2000).
56. Bogacz, R., Brown, E., Moehlis, J., Holmes, P. & Cohen, J. D. The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced-choice tasks. *Psychol. Rev.* **113**, 700–765 (2006).
57. Baum, E. B., Moody, J. & Wilczek, F. Internal representations for associative memory. *Biol. Cybern.* **59**, 217–228 (1988).
58. Saul, L. K. & Jordan, M. I. Attractor dynamics in feedforward neural networks. *Neural Comput.* **12**, 1313–1335 (2000).
59. Sharma, S., Chandra, S. & Fiete, I. R. Content addressable memory without catastrophic forgetting by heteroassociation with a fixed scaffold. In *Int. Conf. Machine Learning, ICML 2022* (eds Chaudhuri, K. et al.) Vol. 162, 19658-19682 (PMLR, 2022).
60. Funahashi, S., Bruce, C. J. & Goldman-Rakic, P. S. Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *J. Neurophysiol.* **61**, 331–349 (1989).
61. Curtis, C. E. & D'Esposito, M. Persistent activity in the prefrontal cortex during working memory. *Trends Cogn. Sci.* **7**, 415–423 (2003).
62. Chaudhuri, R. & Fiete, I. Bipartite expander Hopfield networks as self-decoding high-capacity error correcting codes. In *Advances in Neural Information Processing Systems* 7686–7697 (NeurIPS, 2019).
63. Song, P. & Wang, X.-J. Angular path integration by moving "hill of activity": a spiking neuron model without recurrent excitation of the head-direction system. *J. Neurosci.* **25**, 1002–1014 (2005).
64. Redish, D., Elga, A. N. & Touretzky, D. S. A coupled attractor model of the rodent head direction system. *Netwk. Comput. Neural Syst.* **7**, 671–685 (1996).
65. Wang, X.-J. Decision making in recurrent neuronal circuits. *Neuron* **60**, 215–234 (2008).
66. Kriener, B., Chaudhuri, R. & Fiete, I. Robust parallel decision-making in neural circuits with nonlinear inhibition. *Proc. Natl Acad. Sci. USA* **117**, 25505–25516 (2020).

67. Usher, M. & McClelland, J. L. The time course of perceptual choice: the leaky, competing accumulator model. *Psychol. Rev.* **108**, 550 (2001).
68. Wong, K.-F. & Wang, X.-J. A recurrent network mechanism of time integration in perceptual decisions. *J. Neurosci.* **26**, 1314–1328 (2006).
69. Hahnloser, R. H., Sarpeshkar, R., Mahowald, M. A., Douglas, R. J. & Seung, H. S. Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit. *Nature* **405**, 947 (2000).
70. Bogacz, R. & Gurney, K. The basal ganglia and cortex implement optimal decision making between alternative actions. *Neural Comput.* **19**, 442–477 (2007).
71. Prat-Ortega, G., Wimmer, K., Roxin, A. & de la Rocha, J. Flexible categorization in perceptual decision making. *Nat. Commun.* **12**, 1–15 (2021).
72. Pfeiffer, B. E. & Foster, D. J. Autoassociative dynamics in the generation of sequences of hippocampal place cells. *Science* **349**, 180–183 (2015).
73. Laje, R. & Buonomano, D. V. Robust timing and motor patterns by taming chaos in recurrent neural networks. *Nat. Neurosci.* **16**, 925–933 (2013).
74. Kleinfeld, D. Sequential state generation by model neural networks. *Proc. Natl Acad. Sci. USA* **83**, 9469–9473 (1986).
75. Sompolinsky, H. & Kanter, I. Temporal association in asymmetric neural networks. *Phys. Rev. Lett.* **57**, 2861 (1986).
76. Fiete, I. R., Senn, W., Wang, C. Z. H. & Hahnloser, R. H. R. Spike-time-dependent plasticity and heterosynaptic competition organize networks to produce long scale-free sequences of neural activity. *Neuron* **65**, 563–576 (2010).
77. Taube, J. S., Muller, R. U. & Ranck, J. B. Jr. Head-direction cells recorded from the postsubiculum in freely moving rats. II. effects of environmental manipulations. *J. Neurosci.* **10**, 436–447 (1990).
78. Yoon, K., Buice, M., Barry, R. C., Hayman, B. N. & Fiete, I. Specific evidence of low-dimensional continuous attractor dynamics in grid cells. *Nat. Neurosci.* **16**, 1077–1084 (2013).
By analysing grid cell data across environments, this work shows that pairwise correlations are preserved within a grid cell module, in agreement with continuous-attractor models.
79. Jun, J. J. et al. Fully integrated silicon probes for high-density recording of neural activity. *Nature* **551**, 232–236 (2017).
80. Ahrens, M. B. et al. Brain-wide neuronal dynamics during motor adaptation in zebrafish. *Nature* **485**, 471–477 (2012).
81. McNaughton, B. L., O'Keefe, J. & Barnes, C. A. The stereotrode: a new technique for simultaneous isolation of several single units in the central nervous system from multiple unit records. *J. Neurosci. Methods* **8**, 391–397 (1983).
82. Wilt, B. A. et al. Advances in light microscopy for neuroscience. *Annu. Rev. Neurosci.* **32**, 435–506 (2009).
83. Obaid, A. M. et al. Massively parallel microwire arrays integrated with CMOS chips for neural recording. *Sci. Adv.* https://doi.org/10.1126/sciadv.aay2789 (2020).
84. Weisenburger, S. & Vaziri, A. A guide to emerging technologies for large-scale and whole-brain optical imaging of neuronal activity. *Annu. Rev. Neurosci.* **41**, 431–452 (2018).
85. Steinmetz, N. A. et al. Neuropixels 2.0: a miniaturized high-density probe for stable, long-term brain recordings. *Science* **372**, eabf4588 (2020).
86. Churchland, M. M. et al. Neural population dynamics during reaching. *Nature* **487**, 51–56 (2012).
87. Wimmer, K., Nykamp, D. Q., Constantinidis, C. & Compte, A. Bump attractor dynamics in prefrontal cortex explains behavioral precision in spatial working memory. *Nat. Neurosci.* **17**, 431–439 (2014).
88. Low, R. J., Lewallen, S., Aronov, D., Nevers, R. & Tank, D. W. Probing variability in a cognitive map using manifold inference from neural dynamics. Preprint at *bioRxiv* https://doi.org/10.1101/418939 (2018).
89. Pandarinath, C. et al. Inferring single-trial neural population dynamics using sequential auto-encoders. *Nat. Methods* **15**, 805–815 (2018).
90. Roweis, S. T. & Saul, L. K. Nonlinear dimensionality reduction by locally linear embedding. *Science* **290**, 2323–2326 (2000).
91. Van der Maaten, L. & Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).
92. Tenenbaum, J. B. A global geometric framework for nonlinear dimensionality reduction. *Science* **290**, 2319–2323 (2000).
93. Wu, A., Pashkovski, S., Datta, S. R. & Pillow, J. W. Learning a latent manifold of odor representations from neural responses in piriform cortex. In *Advances in Neural Information Processing Systems* 5378–5388 (NeurIPS, 2018).
94. Zomorodian, A. & Carlsson, G. Computing persistent homology. *Discret. Comput. Geom.* **33**, 249–274 (2005).
95. Ghrist, R. Barcodes: the persistent topology of data. *Bull. Am. Math. Soc.* **45**, 61–75 (2008).
96. Carlsson, G., Ishkhanov, T., de Silva, V. & Zornorodian, A. On the local behavior of spaces of natural images. *Int. J. Comput. Vis.* **76**, 1–12 (2008).
97. Rybakken, E., Baas, N. & Dunn, B. Decoding of neural data using cohomological feature extraction. *Neural Comput.* **31**, 68–93 (2019).
98. Singh, G. et al. Topological analysis of population activity in visual cortex. *J. Vis.* **8**, 11 (2008).
99. Taube, J. S., Muller, R. U. & Ranck, J. B. Jr. Head-direction cells recorded from the postsubiculum in freely moving rats. I. description and quantitative analysis. *J. Neurosci.* **10**, 420–435 (1990).
100. Yoganarasimha, D., Yu, X. & Knierim, J. J. Head direction cell representations maintain internal coherence during conflicting proximal and distal cue rotations: comparison with hippocampal place cells. *J. Neurosci.* **26**, 622–631 (2006).
101. Trettel, S., Trimper, J., Hwaun, E., Fiete, I. & Colgin, L. Grid cell co-activity patterns during sleep reflect spatial overlap of grid fields during active behaviors. *Nat. Neurosci.* **22**, 609–617 (2019).
102. Gardner, R. J., Lu, L., Wernle, T., Moser, M.-B. & Moser, E. I. Correlation structure of grid cells is preserved during sleep. *Nat. Neurosci.* **22**, 598–608 (2019).
103. Widloski, J., Marder, M. P. & Fiete, I. R. Inferring circuit mechanisms from sparse neural recording and global perturbation in grid cells. *eLife* **7**, e33503 (2018).
104. Cossart, R., Aronov, D. & Yuste, R. Attractor dynamics of network up states in the neocortex. *Nature* **423**, 283–288 (2003).
105. Jercog, D. et al. UP–DOWN cortical dynamics reflect state transitions in a bistable network. *eLife* **6**, e22425 (2017).
106. Sanchez-Vives, M. V., Massimini, M. & Mattia, M. Shaping the default activity pattern of the cortical network. *Neuron* **94**, 993–1001 (2017).
107. Scarpetta, S. & de Candia, A. Alternation of up and down states at a dynamical phase-transition of a neural network with spatiotemporal attractors. *Front. Syst. Neurosci.* **8**, 88 (2014).
108. Jercog, D. et al. Up-down cortical dynamics reflect state transitions in a bistable network. *eLife* **6**, e22425 (2017).
109. Latham, P. E., Richmond, B., Nelson, P. & Nirenberg, S. Intrinsic dynamics in neuronal networks. I. Theory. *J. Neurophysiol.* **83**, 808–827 (2000).
110. Compte, A., Sanchez-Vives, M. V., McCormick, D. A. & Wang, X.-J. Cellular and network mechanisms of slow oscillatory activity (<1Hz) and wave propagations in a cortical network model. *J. Neurophysiol.* **89**, 2707–2725 (2003).
111. Kasanetz, F., Riquelme, L. A., O'Donnell, P. & Murer, M. G. Turning off cortical ensembles stops striatal up states and elicits phase perturbations in cortical and striatal slow oscillations in rat in vivo. *J. Physiol.* **577**, 97–113 (2006).
112. Rigas, P. & Castro-Alamancos, M. A. Thalamocortical up states: differential effects of intrinsic and extrinsic cortical inputs on persistent activity. *J. Neurosci.* **27**, 4261–4272 (2007).
113. McCormick, D. A., McGinley, M. J. & Salkoff, D. B. Brain state dependent activity in the cortex and thalamus. *Curr. Opin. Neurobiol.* **31**, 133–140 (2015).
114. Deutsch, D. An auditory illusion. *Nature* **251**, 307–309 (1974).
115. Ward, E. J. & Scholl, B. J. Stochastic or systematic? Seemingly random perceptual switching in bistable events triggered by transient unconscious cues. *J. Exp. Psychol. Hum. Percept. Perform.* **41**, 929 (2015).
116. Blake, R. & Logothetis, N. K. Visual competition. *Nat. Rev. Neurosci.* **3**, 13–21 (2002).
117. McWalter, R. & McDermott, J. H. Illusory sound texture reveals multi-second statistical completion in auditory scene analysis. *Nat. Commun.* **10**, 1–18 (2019).
118. Wang, M., Arteaga, D. & He, B. J. Brain mechanisms for simple perception and bistable perception. *Proc. Natl Acad. Sci. USA* **110**, E3350–E3359 (2013).
119. Vattikuti, S. et al. Canonical cortical circuit model explains rivalry, intermittent rivalry, and rivalry memory. *PLoS Comput. Biol.* **12**, e1004903 (2016).
120. Moreno-Bote, R., Rinzel, J. & Rubin, N. Noise-induced alternations in an attractor network model of perceptual bistability. *J. Neurophysiol.* **98**, 1125–1139 (2007).
121. Inagaki, H. K., Fontolan, L., Romani, S. & Svoboda, K. Discrete attractor dynamics underlies persistent activity in the frontal cortex. *Nature* **566**, 212–217 (2019).
This work tests the predictions of discrete attractor dynamics in the rodent ALM using optogenetic perturbations.
122. Li, N., Daie, K., Svoboda, K. & Druckmann, S. Robust neuronal dynamics in premotor cortex during motor planning. *Nature* **532**, 459–464 (2016).
123. Piet, A. T., Erlich, J. C., Kopec, C. D. & Brody, C. D. Rat prefrontal cortex inactivations during decision making are explained by bistable attractor dynamics. *Neural Comput.* **29**, 2861–2886 (2017).
124. Erlich, J. C., Brunton, B. W., Duan, C. A., Hanks, T. D. & Brody, C. D. Distinct effects of prefrontal and parietal cortex inactivations on an accumulation of evidence task in the rat. *eLife* **4**, e05457 (2015).
125. Inagaki, H. K., Inagaki, M., Romani, S. & Svoboda, K. Low-dimensional and monotonic preparatory activity in mouse anterior lateral motor cortex. *J. Neurosci.* **38**, 4163–4185 (2018).
126. Daie, K., Svoboda, K. & Druckmann, S. Targeted photostimulation uncovers circuit motifs supporting short-term memory. *Nat. Neurosci.* **24**, 259–265 (2021).
127. Lazzaro, J., Ryckebusch, S., Mahowald, M. A. & Mead, C. A. Winner-take-all networks of O(N) complexity. In *Advances in Neural Information Processing Systems* 703–711 (NeurIPS, 1989).
128. Xie, X., Hahnloser, R. H. & Seung, H. S. Selectively grouping neurons in recurrent networks of lateral inhibition. *Neural Comput.* **14**, 2627–2646 (2002).
129. Majani, E., Erlanson, R. & Abu-Mostafa, Y. S. On the K-winners-take-all network. In *Advances in Neural Information Processing Systems* 634–642 (NeurIPS, 1989).
130. Bolding, K. A. & Franks, K. M. Recurrent cortical circuits implement concentration-invariant odor coding. *Science* **361**, eaat6904 (2018).
131. Sreenivasan, S. & Fiete, I. Grid cells generate an analog error-correcting code for singularly precise neural computation. *Nat. Neurosci.* **14**, 1330–1337 (2011).
132. de Almeida, L., Idiart, M. & Lisman, J. E. The input–output transformation of the hippocampal granule cells: from grid cells to place fields. *J. Neurosci.* **29**, 7504–7512 (2009).

133. Espinoza, C., Guzman, S. J., Zhang, X. & Jonas, P. Parvalbumin+interneurons obey unique connectivity rules and establish a powerful lateral-inhibition microcircuit in dentate gyrus. *Nat. Commun.* **9**, 4605 (2018).
134. Kurt, S. et al. Auditory cortical contrast enhancing by global winner-take-all inhibitory interactions. *PLoS ONE* **3**, e1735 (2008).
135. Josselyn, S. A. & Tonegawa, S. Memory engrams: recalling the past and imagining the future. *Science* **367 6473**, eaaw4325 (2020).
136. Lin, A. C., Bygrave, A. M., de Calignon, A., Lee, T. & Miesenböck, G. Sparse, decorrelated odor coding in the mushroom body enhances learned odor discrimination. *Nat. Neurosci.* **17**, 559–568 (2014).
137. Stevens, C. F. What the fly's nose tells the fly's brain. *Proc. Natl Acad. Sci. USA* **112**, 9460–9465 (2015).
138. Arnold, D. & Robinson, D. The oculomotor integrator: testing of a neural network model. *Exp. Brain Res.* **113**, 57–74 (1997).
139. Aksay, E., Gamkrelidze, G., Seung, H. S., Baker, R. & Tank, D. W. In vivo intracellular recording and perturbation of persistent activity in a neural integrator. *Nat. Neurosci.* **4**, 184–193 (2001).
This work tests the predictions of line attractor dynamics in the goldfish oculomotor integrator using in vivo intracellular current perturbations.
140. Pastor, A., Cruz, L. D. R. & Baker, R. Eye position and eye velocity integrators reside in separate brainstem nuclei. *Proc. Natl Acad. Sci. USA* **91**, 807–811 (1994).
141. Cannon, C. & Robinson, D. Loss of the neural integrator of the oculomotor system from brain stem lesions in monkey. *J. Neurophys.* **57**, 1383–1409 (1987).
142. Mettens, P., Godaux, E., Cheron, G. & Galiana, H. Effect of muscimol microinjections into the prepositus hypoglossi and the medial vestibular nuclei on cat eye movements. *J. Neurophysiol.* **72**, 785–802 (1994).
143. Kaneko, C. R. Eye movement deficits after ibotenic acid lesions of the nucleus prepositus hypoglossi in monkeys. I. Saccades and fixation. *J. Neurophysiol.* **78**, 1753–1768 (1997).
144. Major, G. et al. Plasticity and tuning by visual feedback of the stability of a neural integrator. *Proc. Natl Acad. Sci. USA* **101**, 7739–7744 (2004).
145. Vishwanathan, A. et al. Electron microscopic reconstruction of functionally identified cells in a neural integrator. *Curr. Biol.* **27**, 2137–2147 (2017).
146. Vishwanathan, A. et al. Predicting modular functions and neural coding of behavior from a synaptic wiring diagram. Preprint at *bioRxiv* https://doi.org/10.1101/2020.10.28.359620 (2021).
147. Taube, J. S. Head direction cells recorded in the anterior thalamic nuclei of freely moving rats. *J. Neurosci.* **15**, 70–86 (1995).
148. Kim, S. S., Hermundstad, A. M., Romani, S., Abbott, L. F. & Jayaraman, V. Generation of stable heading representations in diverse visual scenes. *Nature* **576**, 126–131 (2019).
149. Yoder, R. M. & Taube, J. S. The vestibular contribution to the head direction signal and navigation. *Front. Integr. Neurosci.* **8**, 32 (2014).
150. Yoder, R. M., Peck, J. R. & Taube, J. S. Visual landmark information gains control of the head direction signal at the lateral mammillary nuclei. *J. Neurosci.* **35**, 1354–1367 (2015).
151. Hulse, B. K. & Jayaraman, V. Mechanisms underlying the neural computation of head direction. *Annu. Rev. Neurosci.* **43**, 31–54 (2020).
152. Fisher, Y. E., Lu, J., D'Alessandro, I. & Wilson, R. I. Sensorimotor experience remaps visual input to a heading-direction network. *Nature* **576**, 121–125 (2019).
153. Angelaki, D. E. & Laurens, J. The head direction cell network: attractor dynamics, integration within the navigation system, and three-dimensional properties. *Curr. Opin. Neurobiol.* **60**, 136–144 (2020).
154. Kim, S. S., Rouault, H., Druckmann, S. & Jayaraman, V. Ring attractor dynamics in the *Drosophila* central brain. *Science* **356**, 849–853 (2017).
This work uses calcium imaging and optogenetics in the ellipsoid body to identify network motifs and dynamics corresponding to ring attractors.
155. Green, J. et al. A neural circuit architecture for angular integration in *Drosophila*. *Nature* **546**, 101–106 (2017).
156. Turner-Evans, D. B. et al. The neuroanatomical ultrastructure and function of a biological ring attractor. *Neuron* **108**, 145–163.e10 (2020).
157. Skaggs, W. E., Knierim, J. J., Kudrimoti, H. S. & McNaughton, B. L. A model of the neural basis of the rat's sense of direction. In *Advances in Neural Information Processing Systems*. 173–180 (NeurIPS, 1995).
158. Stone, T. et al. An anatomically constrained model for path integration in the bee brain. *Curr. Biol.* **27**, 3069–3085 (2017).
159. Lyu, C., Abbott, L. & Maimon, G. Building an allocentric travelling direction signal via vector computation. *Nature* **601**, 92–97 (2022).
160. Asumbisa, K., Peyrache, A. & Trenholm, S. Flexible cue anchoring strategies enable stable head direction coding in both sighted and blind animals. *Nat. Commun.* **13**, 5483 (2022).
161. Hafting, T., Fyhn, M., Molden, S., Moser, M.-B. & Moser, E. Microstructure of a spatial map in the entorhinal cortex. *Nature* **436**, 801–806 (2005).
162. Guanella, A., Kiper, D. & Verschure, P. A model of grid cells based on a twisted torus topology. *Int. J. Neural Syst.* **17**, 231–240 (2007).
163. Burak, Y. & Fiete, I. Do we understand the emergent dynamics of grid cell activity? *J. Neurosci.* **26**, 9352–9354 (2006).
164. Stensola, H. et al. The entorhinal grid map is discretized. *Nature* **492**, 72–78 (2012).
165. Fyhn, M., Hafting, T., Treves, A., Moser, M.-B. & Moser, E. I. Hippocampal remapping and grid realignment in entorhinal cortex. *Nature* **446**, 190–194 (2007).
166. Yoon, K., Lewallen, S., Kinkhabwala, A. A., Tank, D. W. & Fiete, I. R. Grid cell responses in 1D environments assessed as slices through a 2D lattice. *Neuron* **89**, 1086–1099 (2016).
This work shows that grid cell firing fields in linear tracks are well predicted by a one-dimensional slice through a two-dimensional hexagonal lattice, consistent with continuous-attractor dynamics.
167. Kropff, E. & Treves, A. The emergence of grid cells: intelligent design or just adaptation? *Hippocampus* **18**, 1256–1269 (2008).
168. Dordek, Y., Soudry, D., Meir, R. & Derdikman, D. Extracting grid cell characteristics from place cell inputs using non-negative principal component analysis. *eLife* **5**, e10094 (2016).
169. Stachenfeld, K. L., Botvinick, M. M. & Gershman, S. J. The hippocampus as a predictive map. *Nat. Neurosci.* **20**, 1643–1653 (2017).
170. Barry, C., Hayman, R., Burgess, N. & Jeffery, K. J. Experience-dependent rescaling of entorhinal grids. *Nat. Neurosci.* **10**, 682–684 (2007).
171. Boccara, C. N., Nardin, M., Stella, F., O'Neill, J. & Csicsvari, J. The entorhinal cognitive map is attracted to goals. *Science* **363**, 1443–1447 (2019).
172. Butler, W. N., Hardcastle, K. & Giocomo, L. M. Remembered reward locations restructure entorhinal spatial maps. *Science* **363**, 1447–1452 (2019).
173. Krupic, J., Bauza, M., Burton, S., Barry, C. & O'Keefe, J. Grid cell symmetry is shaped by environmental geometry. *Nature* **518**, 232–235 (2015).
174. Hayman, R. M. A., Casali, G., Wilson, J. J. & Jeffery, K. J. Grid cells on steeply sloping terrain: evidence for planar rather than tric encoding. *Front. Psychol.* **6**, 925 (2015).
175. Ginosar, G. et al. Locally ordered representation of 3D space in the entorhinal cortex. *Nature* **596**, 404–409 (2021).
176. Grieves, R. M. et al. Irregular distribution of grid cell firing fields in rats exploring a 3D volumetric space. *Nat. Neurosci.* **24**, 1567–1573 (2021).
177. Keinath, A. T., Epstein, R. A. & Balasubramanian, V. Environmental deformations dynamically shift the grid cell spatial metric. *eLife* **7**, e38169 (2018).
178. Welinder, P. E., Burak, Y. & Fiete, I. R. Grid cells: the position code, neural network models of activity, and the problem of learning. *Hippocampus* **18**, 1283–1300 (2008).
179. Widloski, J. & Fiete, I. R. A model of grid cell development through spatial exploration and spike time-dependent plasticity. *Neuron* **83**, 481–495 (2014).
180. Hardcastle, K., Ganguli, S. & Giocomo, L. M. Environmental boundaries as an error correction mechanism for grid cells. *Neuron* **86**, 827–839 (2015).
181. Fiete, I. R., Burak, Y. & Brookings, T. What grid cells convey about rat location. *J. Neurosci.* **28**, 6858–6871 (2008).
182. Gnadt, J. W. & Andersen, R. A. Memory related motor planning activity in posterior parietal cortex of macaque. *Exp. Brain Res.* **70**, 216–220 (1988).
183. Constantinidis, C., Franowicz, M. N. & Goldman-Rakic, P. S. Coding specificity in cortical microcircuits: a multiple-electrode analysis of primate prefrontal cortex. *J. Neurosci.* **21**, 3646–3655 (2001).
184. Izhikevich, E. M. *Dynamical Systems in Neuroscience* (MIT Press, 2007).
185. Ashwin, P., Coombes, S. & Nicks, R. Mathematical frameworks for oscillatory network dynamics in neuroscience. *Math. Neurosci.* **6**, 1–92 (2016).
186. Adamantidis, A. R., Herrera, C. G. & Gent, T. C. Oscillating circuitries in the sleeping brain. *Nat. Rev. Neurosci.* **20**, 746–762 (2019).
187. Bruno, A. M., Frost, W. N. & Humphries, M. D. A spiral attractor network drives rhythmic locomotion. *eLife* **6**, e27342 (2017).
188. Nichols, A. L., Eichler, T., Latham, R. & Zimmer, M. A global brain state underlies *C. elegans* sleep behavior. *Science* **356**, eaam6851 (2017).
189. Bucher, D., Haspel, G., Golowasch, J. & Nadim, F. Central pattern generators. *eLS* https://doi.org/10.1002/9780470015902.a0000032.pub2 (2015).
190. Sauerbrei, B. A. et al. Cortical pattern generation during dexterous movement is input-driven. *Nature* **577**, 386–391 (2020).
191. Marder, E. & Bucher, D. Central pattern generators and the control of rhythmic movements. *Curr. Biol.* **11**, R986–R996 (2001).
192. Marder, E. & Calabrese, R. L. Principles of rhythmic motor pattern generation. *Physiol. Rev.* **76**, 687–717 (1996).
193. Goulding, M. Circuits controlling vertebrate locomotion: moving in a new direction. *Nat. Rev. Neurosci.* **10**, 507–518 (2009).
194. Kiehn, O. Decoding the organization of spinal circuits that control locomotion. *Nat. Rev. Neurosci.* **17**, 224 (2016).
195. Yuste, R., MacLean, J. N., Smith, J. & Lansner, A. The cortex as a central pattern generator. *Nat. Rev. Neurosci.* **6**, 477–483 (2005).
196. Hubel, D. H. & Wiesel, T. N. Receptive fields of single neurones in the cat's striate cortex. *J. Physiol.* **148**, 574–591 (1959).
197. von der Heydt, R., Peterhans, E. & Baumgartner, G. Illusory contours and cortical neuron responses. *Science* **224**, 1260–1262 (1984).
198. Grosof, D. H., Shapley, R. M. & Hawken, M. J. Macaque V1 neurons can signal 'illusory' contours. *Nature* **365**, 550–552 (1993).
199. Grinvald, A., Lieke, E., Frostig, R. D., Gilbert, C. D. & Wiesel, T. N. Functional architecture of cortex revealed by optical imaging of intrinsic signals. *Nature* **324**, 361–364 (1986).
200. Zhong, W., Lu, Z., Schwab, D. J. & Murugan, A. Non-equilibrium statistical mechanics of continuous attractors. *Neural Comput.* **32**, 1033–1068 (2020).
201. Fung, C. C. A. et al. Discrete-attractor-like tracking in continuous attractor neural networks. *Phys. Rev. Lett.* **122**, 018102 (2019).
202. Thorpe, S., Fize, D. & Marlot, C. Speed of processing in the human visual system. *Nature* **381**, 520–522 (1996).

203. Ferster, D., Chung, S. & Wheat, H. Orientation selectivity of thalamic input to simple cells of cat visual cortex. *Nature* **380**, 249–252 (1996).
204. Hennequin, G., Ahmadian, Y., Rubin, D. B., Lengyel, M. & Miller, K. D. The dynamical regime of sensory cortex: stable dynamics around a single stimulus-tuned attractor account for patterns of noise variability. *Neuron* **98**, 846–860.e5 (2018).
205. O'Keefe, J. & Dostrovsky, J. The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. *Brain Res.* **34**, 171–175 (1971).
206. Quirk, G. J., Muller, R. U. & Kubie, J. L. The firing of hippocampal place cells in the dark depends on the rat's recent experience. *J. Neurosci.* **10**, 2008–2017 (1990).
207. Wilson, M. A. & McNaughton, B. L. Reactivation of hippocampal ensemble memories during sleep. *Science* **265**, 676–679 (1994).
208. Skaggs, W. E. & McNaughton, B. L. Replay of neuronal firing sequences in rat hippocampus during sleep following spatial experience. *Science* **271**, 1870–1873 (1996).
209. Tsodyks, M. & Sejnowski, T. Associative memory and hippocampal place cells. *Int. Neural Syst.* **6**, 81–86 (1995).
210. Samsonovich, A. & McNaughton, B. L. Path integration and cognitive mapping in a continuous attractor neural network model. *J. Neurosci.* **17**, 5900–5920 (1997).
211. Samsonovich, A. V. *Attractor Map Theory of the Hippocampal Representation of Space*. Ph.D. thesis (Univ. Arizona, 1997).
212. Battista, A. & Monasson, R. Capacity-resolution trade-off in the optimal learning of multiple low-dimensional manifolds by attractor neural networks. *Phys. Rev. Lett.* **124**, 048302 (2020).
213. Yim, M. Y., Sadun, L. A., Fiete, I. R. & Taillefumier, T. Place-cell capacity and volatility with grid-like inputs. *eLife* **10**, e62702 (2021).
214. Colgin, L. L., Moser, E. I. & Moser, M.-B. Understanding memory through hippocampal remapping. *Trends Neurosci.* **31**, 469–477 (2008).
215. Alme, C. B. et al. Place cells in the hippocampus: eleven maps for eleven rooms. *Proc. Natl Acad. Sci. USA* **111**, 18428–18435 (2014).
216. Solstad, T., Moser, E. I. & Einevoll, G. T. From grid cells to place cells: a mathematical model. *Hippocampus* **16**, 1026–1031 (2006).
217. Barry, C. et al. The boundary vector cell model of place cell firing and spatial memory. *Rev. Neurosci.* **17**, 71–98 (2006).
218. Whittington, J. C. R. et al. The Tolman–Eichenbaum machine: unifying space and relational memory through generalization in the hippocampal formation. *Cell* **183**, 1249–1263.e23 (2020).
219. Kudrimoti, H., Barnes, C. & McNaughton, B. Reactivation of hippocampal cell assemblies: effects of behavioral state, experience, and EEG dynamics. *J. Neurosci.* **19**, 4090–4101 (1999).
220. Davidson, T. J., Kloosterman, F. & Wilson, M. A. Hippocampal replay of extended experience. *Neuron* **63**, 497–507 (2009).
221. Pfeiffer, B. E. & Foster, D. J. Hippocampal place-cell sequences depict future paths to remembered goals. *Nature* **497**, 74–79 (2013).
222. Agmon, H. & Burak, Y. A theory of joint attractor dynamics in the hippocampus and the entorhinal cortex accounts for artificial remapping and grid cell field-to-field variability. *eLife* **9**, e56894 (2020).
223. Moran, D. W. & Schwartz, A. B. Motor cortical representation of speed and direction during reaching. *J. Neurophysiol.* **82**, 2676–2692 (1999).
224. Shenoy, K. V., Sahani, M. & Churchland, M. M. Cortical control of arm movements: a dynamical systems perspective. *Annu. Rev. Neurosci.* **36**, 337–359 (2013).
225. Gallego, J. A. et al. Cortical population activity within a preserved neural manifold underlies multiple motor behaviors. *Nat. Commun.* **9**, 4233 (2018).
226. Gallego, J. A., Perich, M. G., Chowdhury, R. H., Solla, S. A. & Miller, L. E. Long-term stability of cortical population dynamics underlying consistent behavior. *Nat. Neurosci.* **23**, 260–270 (2020).
227. Wehr, M. & Laurent, G. Odour encoding by temporal sequences of firing in oscillating neural assemblies. *Nature* **384**, 162–166 (1996).
228. Rokni, U. & Sompolinsky, H. How the brain generates movement. *Neural Comput.* **24**, 289–331 (2012).
229. Kobak, D. et al. Demixed principal component analysis of neural population data. *eLife* **5**, e10989 (2016).
230. Klukas, M., Lewis, M. & Fiete, I. Efficient and flexible representation of higher-dimensional cognitive variables with grid cells. *PLoS Comput. Biol.* **16**, e1007796 (2020).
231. Banino, A. et al. Vector-based navigation using grid-like representations in artificial agents. *Nature* **557**, 429–433 (2018).
232. Sanders, H., Wilson, M. A. & Gershman, S. J. Hippocampal remapping as hidden state inference. *eLife* **9**, e51140 (2020).
233. Killian, N. J., Jutras, M. J. & Buffalo, E. A. A map of visual space in the primate entorhinal cortex. *Nature* **491**, 761–764 (2012).
234. Aronov, D., Nevers, R. & Tank, D. W. Mapping of a non-spatial dimension by the hippocampal–entorhinal circuit. *Nature* **543**, 719–722 (2017).
235. Constantinescu, A. O., O'Reilly, J. X. & Behrens, T. E. Organizing conceptual knowledge in humans with a gridlike code. *Science* **352**, 1464–1468 (2016).
236. Hillar, C. J. & Tran, N. M. Robust exponential memory in Hopfield networks. *Math. Neurosci.* **8**, 1 (2018).
237. Fiete, I., Schwab, D. & Tran, N. M. in *Proc. 2nd Workshop on Biological Distributed Algorithms* https://fietelabmit.files.wordpress.com/2018/12/Ngoc_BDA_2014.pdf (2014).
238. Mosheiff, N. & Burak, Y. Velocity coupling of grid cell modules enables stable embedding of a low dimensional variable in a high dimensional neural attractor. *eLife* **8**, 48494 (2019).
239. Muscinelli, S. P., Gerstner, W. & Brea, J. Exponentially long orbits in Hopfield neural networks. *Neural Comput.* **29**, 458–484 (2017).
240. Mathis, A., Herz, A. & Stemmler, M. Optimal population codes for space: grid cells outperform place cells. *Neural Comput.* **24**, 2280–2317 (2012).
241. Gardner, E. The space of interactions in neural network models. *J. Phys. A Math. Gen.* **21**, 257 (1988).
242. Gripon, V. & Berrou, C. Sparse neural networks with large learning diversity. *IEEE Trans. Neural Netw.* **22**, 1087–1096 (2011).
243. Abu-Mostafa, Y. S. & St Jacques, J. Information capacity of the Hopfield model. *IEEE Trans. Inf. Theory* **31**, 461–464 (1985).
244. McEliece, R. J., Posner, E. C., Rodemich, E. R. & Venkatesh, S. S. The capacity of the Hopfield associative memory. *IEEE Trans. Inf. Theory* **33**, 461–482 (1987).
245. Amit, D. J., Gutfreund, H. & Sompolinsky, H. Statistical mechanics of neural networks near saturation. *Ann. Phys.* **173**, 30–67 (1987).
246. Maheswaranathan, N., Williams, A. H., Golub, M. D., Ganguli, S. & Sussillo, D. Reverse engineering recurrent networks for sentiment classification reveals line attractor dynamics. In *Advances in Neural Information Processing Systems* 15696–15705 (NeurIPS, 2019).
247. Kanitscheider, I. & Fiete, I. Emergence of dynamically reconfigurable hippocampal responses by learning to perform probabilistic spatial reasoning. Preprint at *bioRxiv* https://doi.org/10.1101/231159 (2017).
248. Schaeffer, R., Khona, M., Meshulam, L., International Brain Laboratory & Fiete, I. Reverse-engineering recurrent neural network solutions to a hierarchical inference task for mice. In *Advances in Neural Information Processing Systems* 4584–4596 (NeurIPS 2020).
249. Kanitscheider, I. & Fiete, I. R. Training recurrent networks to generate hypotheses about how the brain solves hard navigation problems. In *Advances in Neural Information Processing Systems* 4529–4538 (NeurIPS, 2017).
250. Wang, X.-J. Synaptic basis of cortical persistent activity: the importance of NMDA receptors to working memory. *J. Neurosci.* **19**, 9587–9603 (1999).
251. Ermentrout, G. B. & Kopell, N. Multiple pulse interactions and averaging in systems of coupled neural oscillators. *J. Math. Biol.* **29**, 195–217 (1991).
252. Boerlin, M., Machens, C. K. & Denève, S. Predictive coding of dynamical variables in balanced spiking networks. *PLoS Comput. Biol.* **9**, e1003258 (2013).
253. Frady, E. P. & Sommer, F. T. Robust computation with rhythmic spike patterns. *Proc. Natl Acad. Sci. USA* **116**, 18050–18059 (2019).
254. Darshan, R. & Rivkind, A. Learning to represent continuous variables in heterogeneous neural networks. *Cell Rep.* **39**, 110612 (2021).
255. Arnold, D. B. & Robinson, D. A. A learning network model of the neural integrator of the oculomotor system. *Biol. Cybern.* **64**, 447–454 (1991).
256. Hahnloser, R. H. R., Seung, H. S. & Slotine, J.-J. Permitted and forbidden sets in symmetric threshold-linear networks. *Neural Comput.* **15**, 621–638 (2003).
257. Seung, H. S. Learning continuous attractors in recurrent networks. In *Advances in Neural Information Processing Systems* 654–660 (NeurIPS, 1998).
258. Mante, V., Sussillo, D., Shenoy, K. V. & Newsome, W. T. Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature* **503**, 78–84 (2013).
259. Cueva, C. J. & Wei, X.-X. Emergence of functional and structural properties of the head direction system by optimization of recurrent neural networks. In *Intl Conf. on Learning Representations 2020* https://openreview.net/forum?id=HklSeREtPB (2020).
260. Schaeffer, R., Khona, M. & Fiete, I. R. in *ICML 2022 2nd AI for Science Workshop* https://openreview.net/forum?id=mxi1xKzNFrb (2022).
261. Grosenick, L., Marshel, J. H. & Deisseroth, K. Closed-loop and activity-guided optogenetic control. *Neuron* **86**, 106–139 (2015).
262. Latham, P. E., Deneve, S. & Pouget, A. Optimal computation with attractor networks. *J. Physiol.* **97**, 683–694 (2003).
263. Bouchacourt, F. & Buschman, T. J. A flexible model of working memory. *Neuron* **103**, 147–160 (2019).
264. Hasenstaub, A., Sachdev, R. N. S. & McCormick, D. A. State changes rapidly modulate cortical neuronal responsiveness. *J. Neurosci.* **27**, 9607–9622 (2007).
265. Aksay, E., Baker, R., Seung, H. S. & Tank, D. W. Anatomy and discharge properties of pre-motor neurons in the goldfish medulla that have eye-position signals during fixations. *J. Neurophysiol.* **84**, 1035–1049 (2000).
266. Godaux, E., Mettens, P. & Chéron, G. Differential effect of injections of kainic acid into the prepositus and the vestibular nuclei of the cat. *J. Physiol.* **472**, 459–482 (1993).
267. Hulse, B. K. et al. A connectome of the *Drosophila* central complex reveals network motifs suitable for flexible navigation and context-dependent action selection. *eLife* **10**, e66039 (2021).
268. Gu, Y. et al. A map-like micro-organization of grid cells in the medial entorhinal cortex. *Cell* **175**, 736–750.e30 (2018).
269. Koulakov, A. A. & Chklovskii, D. B. Orientation preference patterns in mammalian visual cortex: a wire length minimization approach. *Neuron* **29**, 519–527 (2001).
270. Wills, T. J., Cacucci, F., Burgess, N. & O'Keefe, J. Development of the hippocampal cognitive map in preweanling rats. *Science* **328**, 1573–1576 (2010).
271. Langston, R. F. et al. Development of the spatial representation system in the rat. *Science* **328**, 1576–1580 (2010).

Acknowledgements
I.R.F. acknowledges funding from the Simons Foundation, the Office of Naval Research, the Howard Hughes Medical Institute (HHMI) through the Faculty Scholars Program, the Department of Brain and Cognitive Sciences, MIT, and the McGovern Institute, MIT. M.K. is supported by a Friends of the McGovern Institute Fellowship, a MathWorks Fellowship and the Department of Physics, MIT. The authors thank X. J. Wang for helpful discussion on short-term memory and persistent activity, and K. Daie, the anonymous reviewers, S. Chandra and other members of the Fiete laboratory for helpful comments on the manuscript.

Author contributions
The authors contributed equally to all aspects of the article.

Competing interests
The authors declare no competing interests.

Additional information
Correspondence should be addressed to Ila R. Fiete.

Peer review information *Nature Reviews Neuroscience* thanks A. Compte, who co-reviewed with J. Barbosa, and the other, anonymous, referee(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© Springer Nature Limited 2022