Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in
Neurobiology

Toward a foundational brain model of intelligence

Xiao-Jing Wang

Foundation models, that are pre-trained by massive data then serve as the core of various systems fine-tuned to accomplish a range of specific tasks, have begun to be deployed in Neuroscience. Here I consider how this new approach may go beyond input-output predictions to yield mechanistic understanding, and enable superb mental abilities which are still poorly understood and elusive in artificial intelligence (AI) systems. I propose the development of a computational platform that incorporates connectome and cell types, captures complex dynamics of recurrent neural circuits, and learns to accomplish multiple cognitive tasks the combined performance of which measures intelligence. Illustrated by recent works in the field of NeuroAI, key requirements are outlined for a biologically-based foundational brain model of intelligence with, at its core, the prefrontal cortex that underlies cognitive capabilities ranging from learning-to-learn, reasoning, planning to executive control of behavior. Resulting novel algorithms could inspire the advancement of general human-like AI.

Addresses
Center for Neural Science, New York University, 4 Washington Place,
New York, NY 10003, USA

Corresponding author: Wang, Xiao-Jing (xjwang@nyu.edu)

Current Opinion in Neurobiology 2026, **98:**103198

This review comes from a themed issue on **Special celebrating issue**

Edited by **Eunjoon Kim** and **Piali Sengupta**

For a complete overview see the Issue and the Editorial

Available online xxx

https://doi.org/10.1016/j.conb.2026.103198

Introduction

New technological tools and the resulting massive data offer unprecedented opportunities in neuroscience. One of the current trends spurred by these advances is to leverage artificial intelligence (AI) for data analysis and computational modeling. For instance, Wang et al. (2025) used machine learning to build a neural network model for the mouse visual cortex [1]; Binz et al. (2025) reported a large language model that performed cognitive tasks of a behavioral dataset called Psych-101 that includes more than 100 experiments [2]. This type of models is called foundation models in AI, which are first pretrained by multimodal datasets then serve as the core of various systems fine-tuned to accomplish a variety of specific tasks [3].

These recent works represent an innovative approach that holds promise for tackling data analysis of different modalities. However, the potential for success of this perspective in neuroscience depends on whether such models can go beyond prediction to mechanistic understanding [4,5]. Here, I will highlight main challenges to be overcome in order to achieve that goal.

Connectome-based modeling for model-brain: alignment and much more

One salient problem of AI modeling is to ‘align’ parts of an artificial neural network model with those of the brain. In a deep neural network (DNN) for human vision consisting of a series of layers, alignment of an area in the model was determined by the best fit of its activity with blood-oxygen-level-dependent (BOLD) signal of functional Magnetic Resonance Imaging in one of the cortical visual areas. The resulting alignment turned out to be positively correlated with the anatomically defined cortical hierarchy [6]. However, the correlation is modest ($R \sim 0.6$), and this approach is at best descriptive. By contrast, in a connectome-based model, parcellated areas (primary visual cortex V1, etc.) are well defined; thus, there is one-to-one correspondence between areas in the model and the brain, avoiding the problem of alignment altogether. Furthermore, today’s DNNs do not capture neural interactions cross-brain regions, not only because of a lack of anatomical constraints but also due to the commonly used feed-forward architecture in sharp contrast to the cortex which is endowed with an abundance of local recurrent connections [7] and feedback connections between areas [8]. For these reasons, a connectome-based computational framework would represent an important step forward for circuit mechanistic understanding.

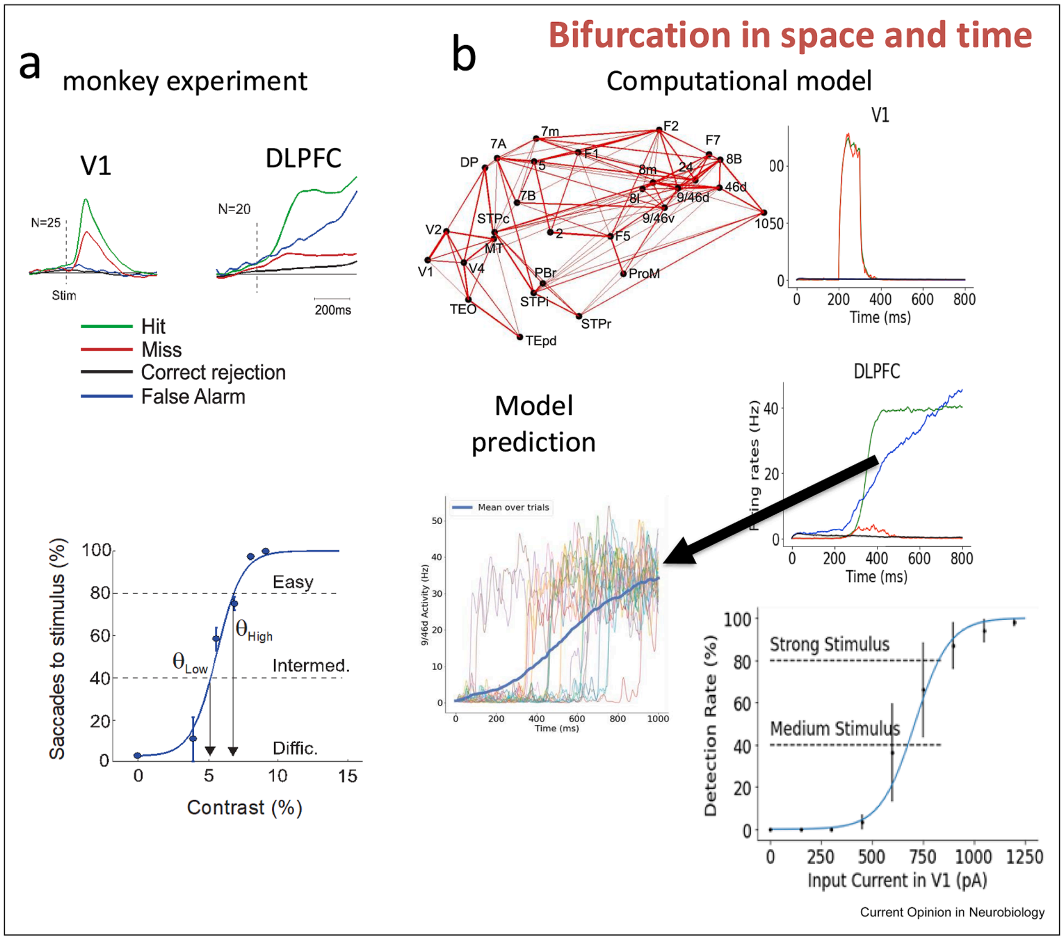
In recent years, connectome-based models of a large-scale multiregional cortex has been developed for human [9,10], mouse [11], macaque [12–14], and marmoset [15] monkeys. This line of research begins to yield surprising model predictions such as a hierarchy of time constants [12] which is supported by single-neuron data from monkeys [15,16] and mice [17]. The connectome-based computational framework is timely in parallel with experimental brain-wide studies beyond local-area considerations. Neural recordings in behaving mice found that neural representations of sensory stimuli, decisions, and motor responses are widely distributed in the brain [18–25]. At the first sight, these observations appear to contradict the notion that different brain areas are specialized in subserving distinct aspects of cognition. Instead, new experimental evidence motivated a more rigorous definition of functional specialization compatible with *selectively* distributed processing, where a functional module composed of a number of areas (rather than a single local area) is dedicated to a function, such as working memory [26] or decision-making (Figure 1) [27].

The work of Klatzmann et al. (2025) represents the first realistic primate cortex model of the global neuronal workspace theory of consciousness [28], in support of a central role of the prefrontal cortex (PFC). The PFC is a critical brain structure for a wide range of cognitive functions [29–33]; its abnormalities are at the root of major psychiatric disorders [34]. Indeed, fMRI studies found that performing tests to assess intelligence consistently activates the PFC [35,36]. It thus is imperative to understand the PFC underlying intelligence.

Leveraging AI tools to build neural circuit models

Training neural networks by machine learning has become a central theme in NeuroAI, at the interface

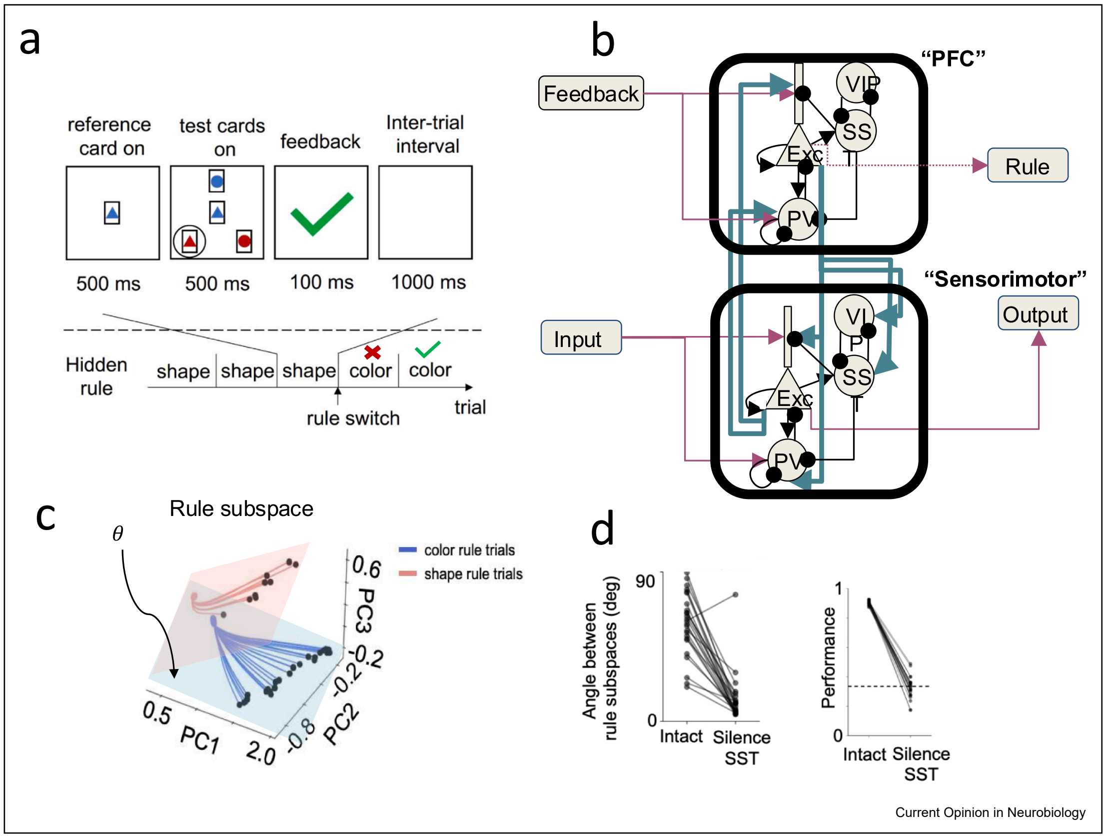
Figure 1



A connectome-based cortex model simulates simple decision-making of a monkey experiment [73]. (**a**) Top: When a visual stimulus with a near-detection threshold contrast is shown, subjects either are aware of it (a hit, green) or not (a miss, red). When no stimulus is shown, a subjective judgment is either its absence (a correct rejection, black) or presence (a false alarm, blue). Neural spiking activity in the primary visual cortex V1 encodes the physical stimulation (green and red), whereas that in the dorsolateral prefrontal cortex reflects subjective decision (green and blue). Bottom: psychometric function (probability of detection) as a function of stimulus contrast. (**b**) The connectome-based model reproduces the salient physiological observations (top) and behavioral performance (bottom, where contrast is implemented by the strength of input to V1). Moreover, the model provides an experimentally testable prediction that the gradual ramping activity in dorsolateral prefrontal cortex (DLPFC) in false-alarm trials is an artifact. In single trials, neural activity of DLPFC exhibits a sudden jump, with the jumping time-varying stochastically from trial to trial, resulting in a smooth ramping time course of the trial average (insert) [27].

between computational neuroscience and AI [37—42]. Generally, input-driven information processes like object recognition are modeled by feedforward DNNs, whereas internally driven mental processes such as subjective decision-making, working memory, and planning, require recurrent neural networks (RNNs). A variety of RNNs have been developed: some use conventional 'vanilla' units akin to biological neurons [39,43], and others incorporate gating units [44,45]; RNNs can be trained using different learning algorithms including supervised learning [39], reinforcement learning [41,46] or nonstandard clone-structured causal graph [47]. Leveraging AI, models have been developed for core building blocks of cognition that depend on the PFC, ranging from task set representation underlying multitasking [48,49] to schema formation for learning-to-learn [50,51]. Figure 2 shows an example of this kind designed to elucidate the critical role of the PFC in behavioral flexibility [52]. A classic assessment of frontal lobe function is the Wisconsin Card Sorting Test (WCST) in which a behavioral rule is not cued by an input but must be inferred and maintained internally over trials to guide sensorimotor mapping [53]. A WCST variant is illustrated in Figure 2a [54]. When the behavioral rule suddenly changes (e.g. from shape to color), the brain must infer the new rule on the basis of feedback to behavioral responses and flexibly switch to a new rule. To test the hypothesis that rule-guided task switching is accomplished in the brain through gating mediated by a specific subtype of GABAergic inhibitory neurons, this RNN incorporated excitatory pyramidal cells and three inhibitory cell types that form a disinhibitory motif composed of parvalbumin, somatostatin (SST), and vasoactive intestinal peptide (VIP)—-positive interneurons. This motif was first proposed theoretically [55], widely supported by later experiments [56,57]. The idea is that a rule-coding circuit of the PFC module projects to both SST+ and VIP+ cells in the sensorimotor module, their relative activities 'gate in' behaviorally relevant feature (e.g. color) in some pyramidal neuron groups and 'gate out' the irrelevant one (shape) in the sensorimotor module. The trained

Figure 2



A trained RNN model for WCST. (**a**) The task is to choose one of 3 options that matches the reference card at the center in either shape or color, depending on a hidden rule that switches after a number of trials. (**b**) The model schema has two modules: the 'PFC' module that learns to encode rules; the 'sensorimotor' module that is trained to generate a choice. (**c**) Neural representations under two different rules correspond to nearly orthogonal low-dimensional subspaces (defined by a principal component analysis of population activity). (**d**) When the dendrite-targeting SST-positive inhibitory neurons are inactivated, the principal angle between the two subspaces is reduced (left) and behavioral performance is at chance level (right). Adapted from Ref. [52]. PFC, prefrontal cortex; RNN, recurrent neural network; SST, somatostatin; WCST, Wisconsin Card Sorting Test.

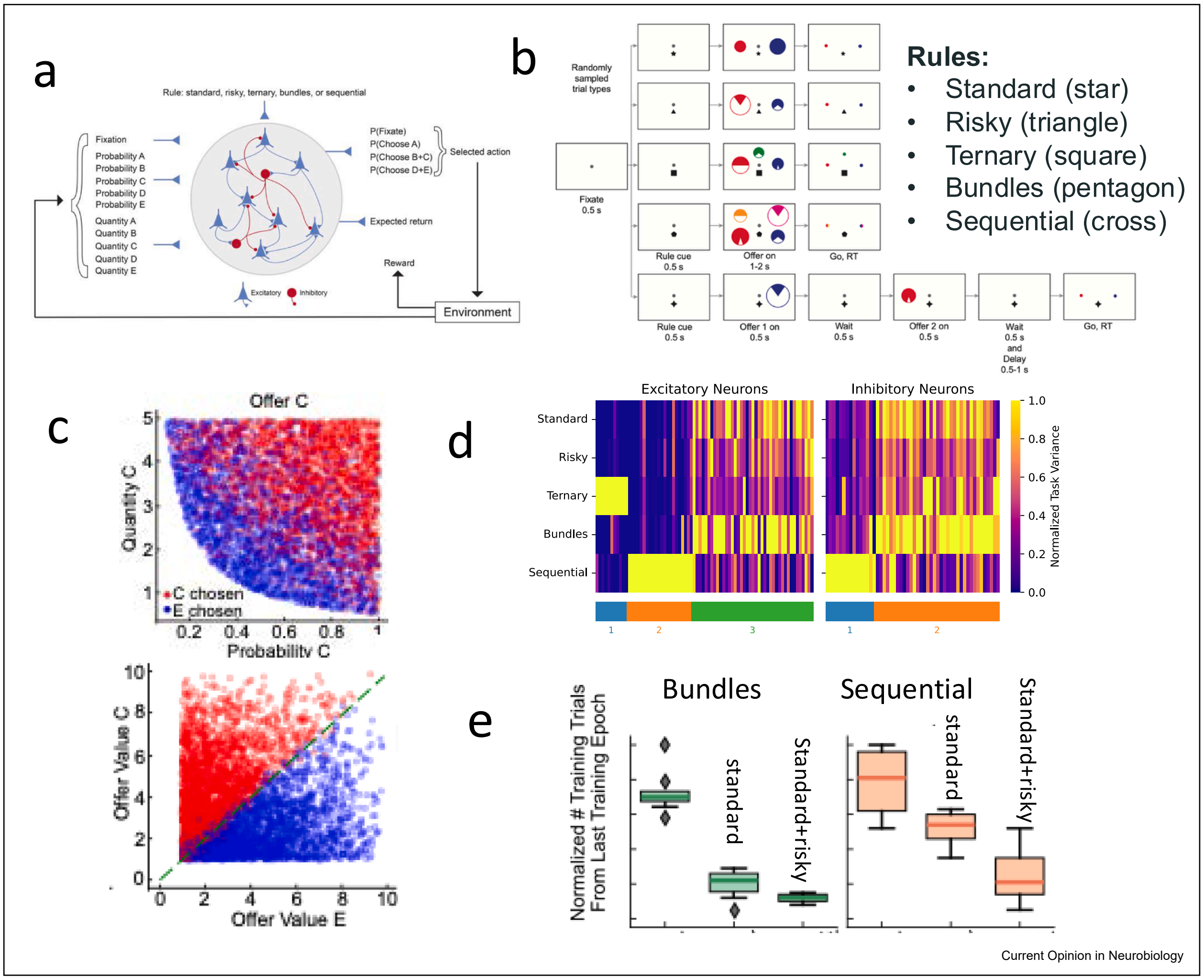
RNN shows that this is indeed the case with the SST+ cells playing a critical role in gating.

The work is noteworthy in two respects. First, it is an RNN trained by machine learning that incorporates cell types and pyramidal dendrites, demonstrating the feasibility of integrating biological details in an AI-type framework. Second, in recent years the dynamical systems approach is commonly applied to describe population activity in the state space, using dimensionality reduction, subspace, or manifold discovery [58–60]. Liu and Wang (2024) showed that neural representations correspond to approximately orthogonal low-dimensional subspaces according to the two behavioral rules. When the SST+ cells were inactivated, the angle between the two subspaces was dramatically reduced and performance is at chance level (Figure 2). This was the first paper to connect the descriptive subspace/manifold approach with a cell-type specific neural circuit mechanism [52].

Compositionality and generalization

A computational model would not be qualified as of the foundation kind unless it can be generalized to novel stimuli and tasks. The hypothesis that an infinite

Figure 3



Current Opinion in Neurobiology

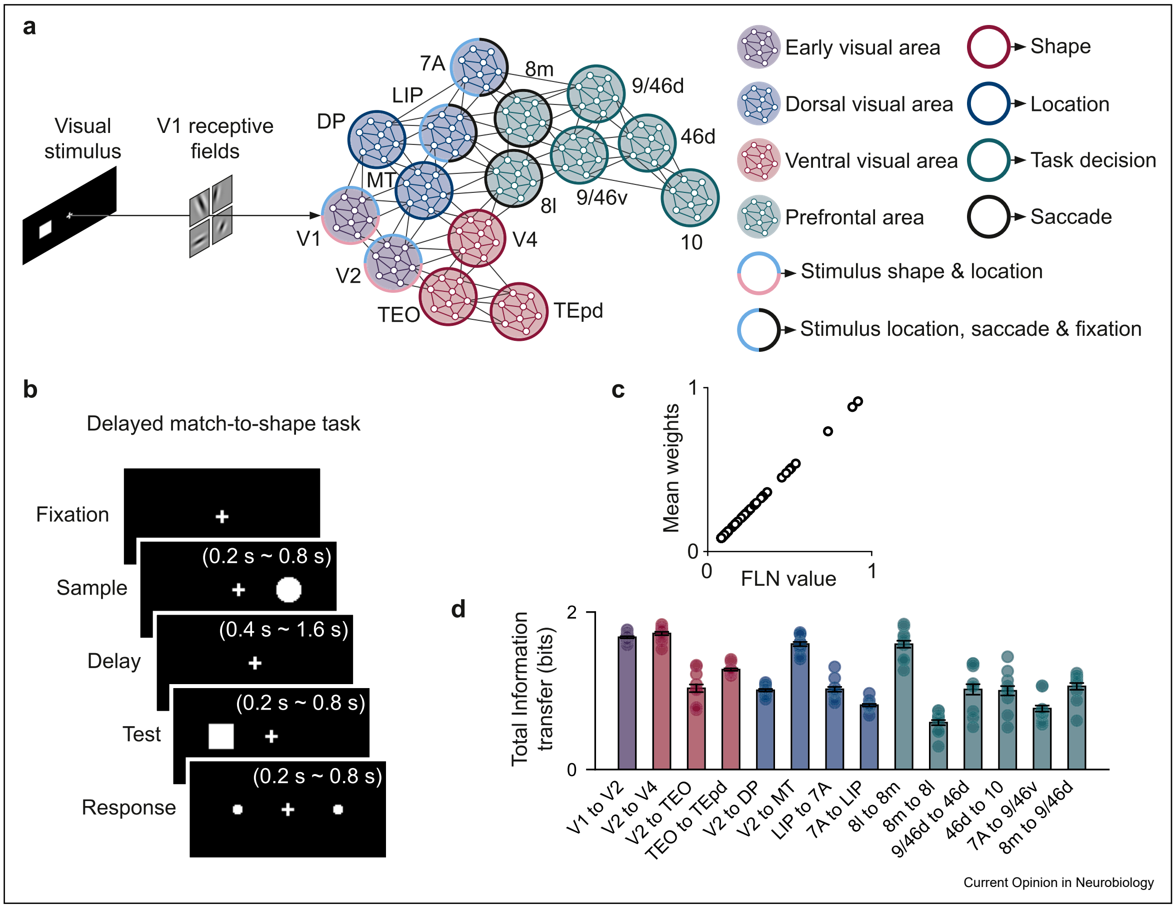
A trained RNN model for probabilistic reward-dependent decision making. (**a**) Architecture of a recurrent neural network for economic choice that obeys Dale's law. (**b**) Five different tasks of choosing between two options associated with different reward amounts ('standard'); two options of reward probabilities and amounts ('risky'); three options ('ternary'); two options, each yielding a mixture of reward types and amounts ('bundles'); and two options the presentations of which are separated by a mnemonic delay period ('sequential'). (**c**) Analysis of post-training network behavior. Each point, indicating a trial decision, is color-coded to represent the chosen offer (C: red; E: blue). (**d**) Analysis of functional cell types revealed 3 clusters of excitatory neurons (left), with the first and second dedicated to three-choice task and sequential task, respectively. Inhibitory neurons show 2 clusters; one of them is specialized in the sequential task (right). (**e**) Learning benefits from curriculum training. Compared to initialization with a random network (a 'blank slate,' the leftmost data point), learning to perform the bundles task (left) and sequential task (right) is faster if the model is first trained to perform the standard task (middle point) and further accelerated with the model also trained beforehand to perform the risky task (right point). Adopted from Ref. [46]. RNN, recurrent neural network.

number of problems can be solved by the use of a limited number of building blocks of cognition that are combined on-demand according to some syntactic rules [61] can be tested in human studies [62,63] and nonhuman animal experiments [64]. An RNN showed how functional clusters of neurons dedicated to different functions can be flexibly combined according to a behavioral rule to carry out 20 cognitive tasks of working memory, decision-making, categorization, and inhibitory control of actions [49,65,66]. Similarly, an RNN trained by reinforcement learning was reported to perform multiple reward-dependent economic decisions [46].

Figure 3a—b shows the model that accomplishes 5 rule-guided tasks. This framework enabled investigations of several questions. For instance, given that the inputs to neurons are additive by excitation or subtractive by inhibition, how is multiplication of a reward size A and its probability p computed to estimate expected value V = Ap in a neural circuit (Figure 3c)? Unlike the previous works, the new model is endowed with Dale's law. It showed that both excitatory and inhibitory neurons are selective for behavioral attributes (Figure 3d), as observed experimentally [67]. Compositionality emerges from training that facilitates acceleration of learning when training occurs through an appropriate curriculum (Figure 3e). Finally, the model exhibits generalization, when training takes places with a small subset of stimuli then tested using previously unseen stimuli, as observed in a monkey experiment [68].

These recent works are a proof of principle for computational models of internally driven cognitive functions,

Figure 4



Connectome-based model of macaque cortex trained to perform cognitive tasks. (**a**) Model setup. Visual stimulus is presented to the multiarea recurrent neural network (RNN) through Gabor filters unique to each excitatory neuron in V1. Stimulus shape and position are trained to be decoded from the ventral and dorsal visual areas, respectively. (**b**) The multiregional RNN trained to perform a delayed match-to-shape working memory task. Two stimuli are presented with a delay between them and one of two saccades is performed based on whether the stimuli have the same shape. Decision to saccade is read out from the frontal eye field. (**c**) Connectomic constraints are preserved through learning, with the average interareal connection weights plotted against anatomically measured data for macaque monkey cortex (FLN stands for fractional labeled neurons) from Ref. [74]. (**d**) Information flow analysis. Information gain about a stimulus feature (e.g. position or shape) from one area to another, computed using information theory, provides an quantification and an intuitive understanding of how visual information flows within the network. Adapted from Soo, W. & X.-J. Wang (2023) Connectome-constrained multi-area recurrent neural networks performing working memory tasks. Society for Neurosci. poster #PSTR 301.10/VV16.

characterized by compositionality and generalization capabilities.

Concluding remarks

So far, two lines of progress have been made independently: connectome-based modeling of the multiregional brain and training RNNs. What is needed now is to integrate the two methodologies. A baby step in that direction is shown in Figure 4, where a cognitive task requiring working memory and decision-making is simulated by a connectomically constrained multiregional RNN, where information flows between parcellated areas are quantified (see also [69]). These preliminary results demonstrate the feasibility of building a connectome-based large-scale foundational brain model capable of performing cognitive tasks. Arguably, this research direction warrants as much efforts as that of computational predictive models of real physical world, implementable in automated robots [70,71]. Given that the PFC abnormalities lead to harmful behavior [72], lessons from learning computational algorithms of how the PFC achieves executive control of thoughts and actions will inspire innovations in AI safety. A computational program with the PFC at its core may well transform neuroscience and pave a path to safe artificial general intelligence.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

I would like to thank Steve Baccus and Kang Shen for the hospitality of my sabbatical year in the Department of Neurobiology and the Wu Tsai Neuroscience Institute at Stanford University, where the paper was written.

Data availability

No data was used for the research described in the article.

References

Papers of particular interest, published within the period of review, have been highlighted as:

* of special interest
** of outstanding interest

1. ** Wang EY, *et al*.: **Foundation model of neural activity predicts response to new stimulus types**. *Nature* 2025, **640**: 470–477.

This work reports an AI-type computational model of the mouse visual cortex, in close relationship with the experimental dataset of the Machine Intelligence from Cortical Networks (MICrONS).

2. * Binz M, *et al*.: **A foundation model to predict and capture human cognition**. *Nature* 2025, **644**:1002–1009.

A foundation model of human behavior, based on a large language model archi-tecture, is shown to perform better than domain-specific cognitive models in many tasks used over 100 experiments.

3. Bommasani R, *et al*.: *On the opportunities and risks of foundation models*. 2021. *arXiv preprint arXiv:2108.07258*.
4. Serre T, Pavlick E: **From prediction to understanding: will ai foundation models transform brain science?** *Neuron* 2025, https://doi.org/10.1016/j.neuron.2025.09.039.
5. Linsley D, Feng P, Serre T: **Better artificial intelligence does not mean better models of biology**. *Trends Cognit Sci* 2025, https://doi.org/10.1016/j.tics.2025.11.016.
6. Tang Y, Gokce A, Al-Karkari KJ, Yamins D, Schrimpf M: **Diverse per-ceptual representations across visual pathways emerge from a single objective**. *bioRxiv* 2025, **2025–07**.
7. Douglas RJ, Martin KAC: **Neuronal circuits of the neocortex**. *Annu Rev Neurosci* 2004, **27**:419–451.
8. Gilbert CD, Li W: **Top-down influences on visual processing**. *Nat Rev Neurosci* 2013, **14**:350–363.
9. Izhikevich EM, Edelman GM: **Large-scale model of mammalian thalam-ocortical systems**. *Proc Natl Acad Sci USA* 2008, **105**: 3593–3598.
10. Deco G, Jirsa VK: **Ongoing cortical activity at rest: criticality, multista-bility, and ghost attractors**. *J Neurosci* 2012, **32**: 3366–3375.
11. * Ding X, Froudist-Walsh S, Jaramillo J, Jiang J, Wang X-J: **Predicting distributed working memory activity in a large-scale mouse brain: the impor-tance of the cell type-specific connectome**. *eLife* 2024, **12**, e85442.

The authors built a connectome-based model of the whole mouse cortico-thalamic system for distributed working memory. Their findings highlight the need of cell-type specific connectomic analysis, and propose causality analysis of neural circuit mechanism of a cognitive function at the level of a functional module of interacting areas rather than a single brain region.

12. Chaudhuri R, Knoblauch K, Gariel MA, Kennedy H, Wang X-J: **A large-scale circuit mechanism for hierarchical dynamical processing in the primate cortex**. *Neuron* 2015, **88**: 419–431.
13. Froudist-Walsh S, *et al*.: **A dopamine gradient controls access to distributed working memory in the large-scale monkey cortex**. *Neuron* 2021, **109**:3500–3520.
14. Mejias JF, Wang X-J: **Mechanisms of distributed working memory in a large-scale model of the macaque neocortex**. *eLife* 2022, **11**, e72136.
15. Li G, Li S, Wang X-J: **A hierarchy of time constants and reliable signal propagation in the marmoset cerebral cortex**. *Nat Commun* 2025, **16**, 11640.
16. Murray JD, *et al*.: **A hierarchy of intrinsic timescales across primate cortex**. *Nat Neurosci* 2014, **17**:1661–1663.
17. Song M, *et al*.: **Hierarchical gradients of multiple timescales in the mammalian forebrain**. *Proc Natl Acad Sci* 2024, **121**, e2415695121.
18. Siegel M, Buschman TJ, Miller EK: **Cortical information flow during flexible sensorimotor decisions**. *Science* 2015, **348**: 1352–1355.
19. Dotson NM, Hoffman SJ, Goodell B, Gray CM: **A large-scale semi-chronic microdrive recording system for non-human primates**. *Neuron* 2017, **96**:769–782.
20. Stringer C, *et al*.: **Spontaneous behaviors drive multidimensional, brainwide activity**. *Science* 2019, **364**:255.
21. Steinmetz NA, *et al*.: **Neuropixels 2.0: a miniaturized high-density probe for stable, long-term brain recordings**. *Science* 2021, **372**:eabf4588.
22. Chen S, *et al*.: **Brain-wide neural activity underlying memory-guided movement**. *Cell* 2024, **187**:676–691.
23. Khilkevich A, *et al*.: **Brain-wide dynamics linking sensation to action during decision-making**. *Nature* 2024, **634**:890–900.
24. ** International Brain Laboratory, *et al*.: **A brain-wide map of neural activity during complex behaviour**. *Nature* 2025, **645**, 177191.

This paper summarizes results of the International brain Laboratory. The con-sortium made of 11 labs recorded from 267 brain regions in 115 mice performing a vi- sual decision task, showing that neural representations and processes are widespread across brain regions.

25. Findling C, *et al.*: **Brain-wide representations of prior information in mouse decision-making**. *Nature* 2025, **645**:192–200.

26. Wang X-J, *et al.*: **Bifurcation in space: emergence of functional modularity in the neocortex**. *bioRxiv* 2025, **2023–06**.

27. Klatzmann U, *et al.*: **A dynamic bifurcation mechanism explains cortex-wide neural correlates of conscious access**. *Cell Rep* 2025, **44**, 115372.
* *
A connectome-based large-scale model of the primate cortex was developed to simulate a detection task. It accounts for the observations in a monkey physiological experiment that the primary sensory cortex encodes external stimulation whereas higher association areas including the prefrontal cortex reflect subjective decision (which is opposite to veridical evidence in miss trials and false alarm trials), in support of the global neuronal workspace theory of consciousness.

28. Dehaene S, Changeux J-P: **Experimental and theoretical approaches to conscious processing**. *Neuron* 2011, **70**: 200–227.

29. Goldman-Rakic PS: **Circuitry of primate prefrontal cortex and regulation of behavior by representational memory**. In *Handbook of Physiology – the nervous system V*. Edited by Plum F, Mountcastle V, Bethesda, Maryland: American Physiological Society; 1987:373–417. chap. 9.

30. Miller EK, Cohen JD: **An integrative theory of prefrontal cortex function**. *Annu Rev Neurosci* 2001, **24**:167–202.

31. Stuss DT, Knight RT: *Principles of frontal lobe function*. New York: Oxford University Press; 2002.

32. Fuster JM: *The prefrontal cortex*. 2nd ed. New York: Academic Press; 1989.

33. Wang X-J: **The prefrontal cortex as a quintessential 'cognitive-type' neural circuit: working memory and decision making**. In *Principles of frontal lobe function*. Edited by Stuss DT, Knight RT. 2nd ed., New York: Cambridge University Press; 2013:226–248.

34. Wang X-J, Krystal JH: **Computational psychiatry**. *Neuron* 2014, **84**:638–654.

35. Duncan J, *et al.*: **A neural basis for general intelligence**. *Science* 2000, **289**:457–460.

36. Hampshire A, Highfield RR, Parkin BL, Owen AM: **Fractionating human intelligence**. *Neuron* 2012, **76**:1225–1237.

37. Yamins DL, *et al.*: **Performance-optimized hierarchical models predict neural responses in higher visual cortex**. *Proc Natl Acad Sci USA* 2014, **111**:8619–8624.

38. Mante V, Sussillo D, Shenoy KV, Newsome WT: **Context-dependent computation by recurrent dynamics in prefrontal cortex**. *Nature* 2013, **503**:78–84.

39. Song HF, Yang GR, Wang X-J: **Training excitatory-inhibitory recurrent neural networks for cognitive tasks: a simple and flexible framework**. *PLoS Comput Biol* 2016, **12**, e1004792.

40. Rajan K, Harvey CD, Tank DW: **Recurrent network models of sequence generation and memory**. *Neuron* 2016, **90**: 128–142.

41. Song HF, Yang GR, Wang X-J: **Reward-based training of recurrent neural networks for cognitive and value-based tasks**. *eLife* 2017, **6**, https://doi.org/10.7554/eLife.21492.

42. Yang GR, Wang X-J: **Artificial neural networks for neuroscientists: a primer**. *Neuron* 2020, **107**:1048–1070.

43. Ehrlich DB, Murray JD: **Geometry of neural computation unifies working memory and planning**. *Proc Natl Acad Sci* 2022, **119**, e2115610119.

44. Jensen KT, Hennequin G, Mattar MG: **A recurrent network model of planning explains hippocampal replay and human behavior**. *Nat Neurosci* 2024, **27**:1340–1348.

45. Dorrell W, Latham PE, Behrens TE, Whittington J: **An efficient computing theory of prefrontal structured working memory representations**. *bioRxiv* 2026, **2026–02**.

46. Battista A, Padoa-Schioppa C, Wang X-J: **A neural circuit framework for economic choice: from building blocks of valuation to compositionality in mul-titasking**. *Neuron* 2026, **114**, https://doi.org/10.1016/j.neuron.2025.12.010.
* *
A recurrent neural network model endowed with Dale's law is trained by re-inforcement learning to perform 5 different value-based decision tasks. It shows compositionality, learning-to-learn, and generalization to previously unseen stimuli.

47. Sun W, *et al.*: **Learning produces an orthogonalized state machine in the hip-pocampus**. *Nature* 2025, **640**:165–175.

48. Sakai K: **Task set and prefrontal cortex**. *Annu Rev Neurosci* 2008, **31**:219–245.

49. Yang GR, Joglekar MR, Song HF, Newsome WT, Wang X-J: **Task representations in neural networks trained to perform many cognitive tasks**. *Nat Neurosci* 2019, **22**:297–306.

50. Zhou J, *et al.*: **Evolving schema representations in orbitofrontal ensembles during learning**. *Nature* 2021, **590**:606–611.

51. Goudar V, Peysakhovich B, Freedman DJ, Buffalo EA, Wang X-J: **Schema formation in a neural population subspace underlies learning-to-learn in flexible sensorimotor problem-solving**. *Nat Neurosci* 2023, **26**:879–890.

52. Liu Y, Wang X-J: **Flexible gating between subspaces in a neural network model of internally guided task switching**. *Nat Commun* 2024, **15**:6497.
*
This is the first RNN model trained by machine learning that incorporates four cell types. It explains mechanistically neural representation subspaces in terms of a specific circuit motif mediated by inhibitory neurons that target dendrites of excitatory neurons.

53. Milner B: **Effects of different brain lesions on card sorting**. *Arch Neurol* 1963, **9**:90–100.

54. Mansouri FA, Matsumoto K, Tanaka K: **Prefrontal cell activities related to monkeys' success and failure in adapting to rule changes in a Wisconsin Card Sorting Test analog**. *J Neurosci* 2006, **26**:2745–2756.

55. Wang X-J, Tegnér J, Constantinidis C, Goldman-Rakic PS: **Division of labor among distinct subtypes of inhibitory neurons in a cortical microcircuit of working memory**. *Proc Natl Acad Sci U S A* 2004, **101**:1368–1373.

56. Tremblay R, Lee S, Rudy B: **GABAergic interneurons in the neocortex: from cellular properties to circuits**. *Neuron* 2016, **91**:260–292.

57. Wang X-J, Yang GR: **A disinhibitory circuit motif and flexible information routing in the brain**. *Curr Opin Neurobiol* 2018, **49**: 75–83.

58. Cunningham JP, Byron MY: **Dimensionality reduction for large-scale neural recordings**. *Nat Neurosci* 2014, **17**: 1500–1509.

59. Gallego JA, Perich MG, Miller LE, Solla SA: **Neural manifolds for the control of movement**. *Neuron* 2017, **94**:978–984.

60. Jazayeri M, Ostojic S: **Interpreting neural computations by examining intrinsic and embedding dimensionality of neural activity**. *Curr Opin Neurobiol* 2021, **70**:113–120.

61. Wang X-J: *Theoretical neuroscience: understanding cognition*. Tayolr & Francis/CRC press; 2025.

62. Cole MW, *et al.*: **Multi-task connectivity reveals flexible hubs for adaptive task control**. *Nat Neurosci* 2013, **16**: 1348–1355.

63. Mill RD, Cole MW: **Dynamically shifting from compositional to con-junctive brain representations supports cognitive task learning**. *Nat Commun* 2025, **16**, 10084.

64. Tafazoli S, *et al.*: **Building compositional tasks with shared neural subspaces**. *Nature* 2025, **650**:164–172.
*
The authors trained monkey subjects to perform three tasks while single-neron recordings were carried out from five cortical and subcortical regions, demonstrating compositionality in the primate brain.

65. * Driscoll LN, Shenoy K, Sussillo D: **Flexible multitask computation in recurrent networks utilizes shared dynamical motifs**. *Nat Neurosci* 2024, **27**:1349–1363.

The authors analyzed a previous RNN model trained to perform 20 rule-guided cognitive tasks [49], revealing that dynamical motifs shared across tasks underlie flexible multitasking.

66. Riveland R, Pouget A: **Natural language instructions induce compositional generalization in networks of neurons**. *Nat Neurosci* 2024, **27**:988–999.

67. Najafi F, *et al.*: **Excitatory and inhibitory subnetworks are equally selective during decision-making and emerge simultaneously during learning**. *Neuron* 2020, **105**:165–179.

68. Bongioanni A, *et al.*: **Activation and disruption of a neural mechanism for novel choice in monkeys**. *Nature* 2021, https://doi.org/10.1038/s41586–020–03115–5.

69. Perich MG, *et al.*: **Inferring brain-wide interactions using data-constrained recurrent neural network models**. *bioRxiv* 2021, https://doi.org/10.1101/2020.12.18.423348.

70. LeCun Y: *A path towards autonomous machine intelligence*. 2022. https://openreview.net/pdf.

71. Li F-F: *From words to worlds: spatial intelligence is AI's next fron-tier*. 2025. *Substack. 2*, https://drfeifei.substack.com/p/from–words–to–worlds–spatial–intelligence.

72. Sapolsky RM: Behave: the biology of humans at our best and worst *(Lon-don*. Penguin Press; 2017.

73. van Vugt B, *et al.*: **The threshold for conscious report: signal loss and response bias in visual and frontal cortex**. *Science* 2018, **360**:537–542.

74. Markov NT, *et al.*: **A weighted and directed interareal connectivity matrix for macaque cerebral cortex**. *Cerebr Cortex* 2014, **24**:17–36.