Essay

Deciphering the mysteries of the neural code

Haim Sompolinsky

Center for Brain Science, Harvard University, Cambridge, MA, USA. Edmond and Lily Safra Center for Brain Sciences, Hebrew University, Jerusalem, Israel.

This essay is dedicated to the memory of my father David Sompolinsky. As a medical student in Veterinary Medicine in Copenhagen, with the support of his professors and the Danish Resistance, David organised the rescue of 700 Danish Jews in October 1943, helping them escape Nazi persecution and find safety in Sweden.

Brain theory, like any scientific theory, starts by posing high-level hypotheses and grand ideas. To make progress, however, these ideas must be formalised as concrete mathematical models describing the operations of networks of nerve cells and their synaptic interactions. These models can then be analysed using mathematical and computational techniques. Ultimately, predictions from these calculations should be tested against experimental data, a process that guides both new experiments and new theoretical work, advancing our understanding of brain function.

What should a brain theory aim to explain? Many of the challenges of theoretical neuroscience may be characterised as aiming to decipher the neural code. Whenever we perceive, think or act, large populations of neurons fire in complex spatio-temporal patterns. The neural code is the mapping between brain activity on one side, and cognition and behaviour on the other. Understanding the neural code involves comprehending the mechanisms that underlie the relationship between neuronal activity patterns and brain function. Such a theory not only illuminates normal brain function but also sheds light on the mechanisms underlying neurological, developmental and psychiatric disorders.

The concept of neural code invites an analogy to the genetic code, the mapping between DNA sequences and proteins. However, the neural code is vastly more complex, and deciphering it poses immense challenges. Firstly, activity patterns dynamically unfold over time, rather than being static. Secondly, they are generated by enormously complex anatomy and biochemistry. Lastly, they depend on an almost infinite number of variables in the sensory environment and internal physiological signalling.

A detailed mapping of neuronal activity patterns is neither sufficient nor necessary for understanding brain function. The sheer microscopic complexity of brain circuits would overwhelm understanding. Instead, neuroscience theory constructs mathematical models of neuronal circuits that abstract away or 'coarse grain' much of the underlying complexity, retaining only key aspects of the structure and biophysics (**Figure 1**). The validity of these 'simplified' models is tested through systematic computer simulations of more detailed models and comparisons with experimental data. As computational and experimental neuroscience advance, we gain a better understanding of which 'simplifications' are valid and which need revision, leading to more biologically faithful models with improved predictive power.

FIGURE 1 Coarse graining in neural circuit models. (**A** & **B**) Complex dendritic and axonal morphologies and synaptic coupling between two cortical neurons are coarse grained as two nodes on a network connects by a directed edge. (**C** & **D**) Millisecond dynamics of membrane potential are coarse grained as a sequence of binary spikes and further coarse grained as firing rate dynamics with time scales of 10 msecs.
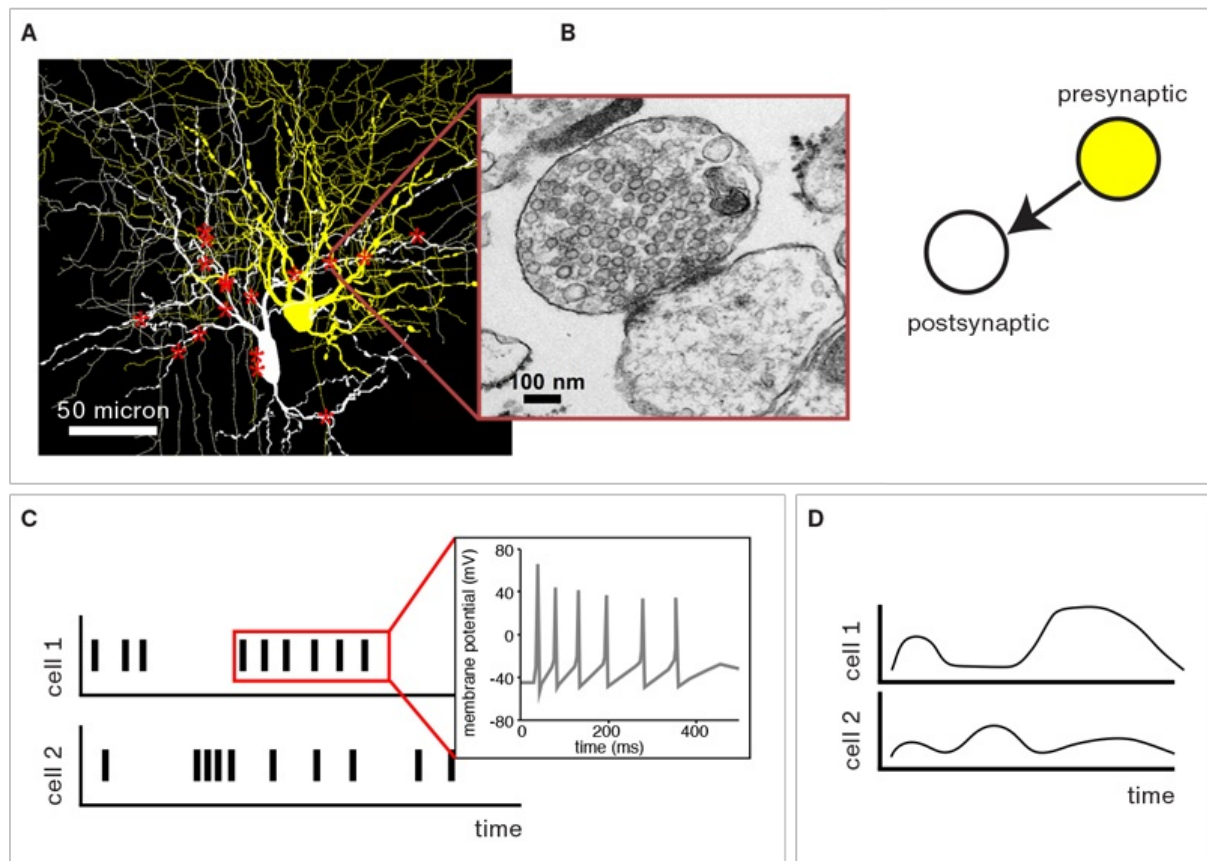
A) Gupta A, Wang Y, Markram H. Organizing principles for a diversity of GABAergic interneurons and synapses in the neocortex. Science. 2000 Jan 14;287(5451):273-8. doi: 10.1126/science.287.5451.273.
B) Copyright OIST (Okinawa Institute of Science and Technology Graduate University).
Creative Commons Attribution 4.0 International License (CC BY 4.0).
C) Cummins TR, Rush AM, Estacion M, Dib-Hajj SD, Waxman SG. Voltage-clamp and current-clamp recordings from mammalian DRG neurons. Nat Protoc. 2009;4(8):1103-12. doi: 10.1038/nprot.2009.91.

The study of local neuronal circuits has proven particularly productive for deciphering the principles linking brain activity to cognition and behaviour [1]. These circuits consist of populations from hundreds to millions of neurons and their synaptic interconnections, all assumed to 'code' for the same set of features of the environment. Theories of these circuits focus on the collective, system-level behaviour emerging from the orchestrated activity of these neuronal populations, metaphorically seen as an orchestra without a conductor, the missing 'homunculus'. These emergent collective computations are reminiscent of phases and phase transitions in condensed matter and the spontaneous pattern formation in physical and biological systems [2]. Just as solidity emerges from the properties of a large collection of interacting atoms, cognition is an emergent property of the activities of large neuronal populations [3].

Understanding neuronal circuits also informs brain disorders traditionally attributed to genetic and molecular anomalies. Research shows that behavioural and cognitive dysfunctions often involve multiple molecular and

genetic pathways within a single disease and, conversely, overlapping across diseases. Studying circuit-level dysfunctions provides a crucial link between molecular mechanisms and disease symptoms and may guide new therapeutic insights.

In the following, I will describe several milestones in theoretical neuroscience, demonstrating the unique power of theory to delineate mechanisms underlying neuronal collective computations and the emergent neural code. These examples are related to my research and do not amount to an exhaustive review of this broad and rich field.

Associative memory storage and retrieval

Early physics-inspired theoretical neuroscience work investigated associative memory in neural circuits. Pioneering physicist John Hopfield proposed a neural network model for associative memory, drawing a clear analogy with magnetic systems in statistical mechanics [4]. In this model, the neural code of each memory is an activity pattern distributed among the N neurons in the network. Memory is encoded by the Hebb plasticity rule ('fire together, wire together'). Retrieval, triggered by a cuing signal, consists of the dynamic evolution of the network state that ultimately converges to a persistent state (a 'fixed point attractor') that corresponds to one of the encoded memories (**Figure 2**).

FIGURE 2 Storage and retrieval of associative memory. (**A**) Memory states are stored as attractors of the Hopfield model. Evolution of the network converges to different fixed points determined by the initialisations ("cuing signals"), corresponding to the memory retrieval process. (**B**) The phase diagram of the Hopfield model exhibits three phases depending on the number of memories and the level of stochasticity: the memory, the spurious phase and the no-retrieval phase.
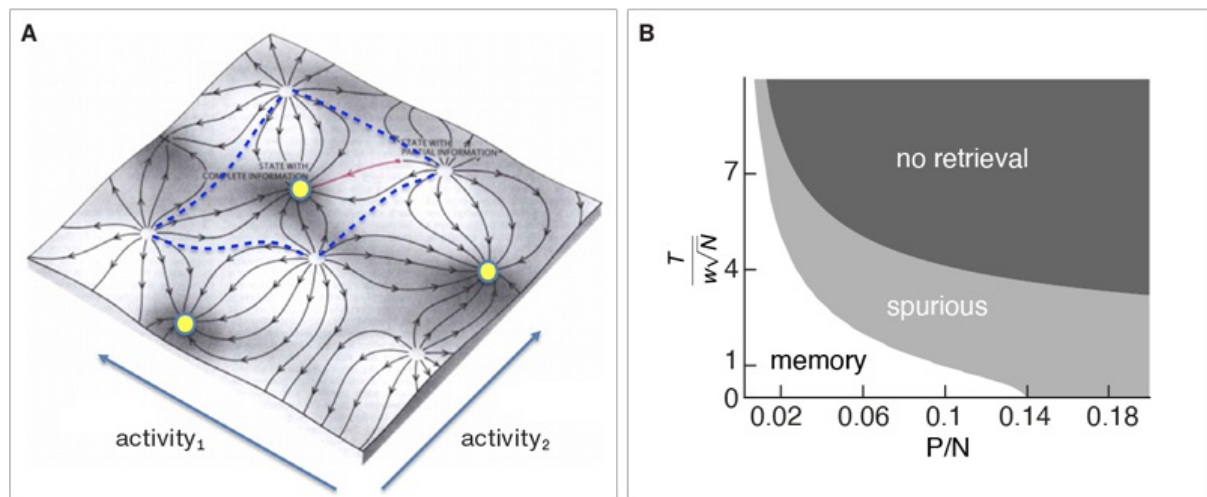
A) Tank DW, Hopfield JJ. Collective Computation in Neuronlike Circuits. Scientific American, 1987 - 257(6), 104–115. http://www.jstor.org/stable/24979583

Working with Hebrew University colleagues Hanoch Gutfreund and the late Daniel Amit, we leveraged sophisticated statistical physics methods to fully analyse the memory properties of the Hopfield model [5, 6]. Our theory was one of the first rigorous, quantitative theories of collective computation in large neural networks. The results of our theory are summarised in the phase diagram of Figure 2. Key parameters affecting the memory functionality include the number of neurons, N, the number of encoded memories (each consisting of a

pattern of activity of the N neurons), denoted P, the level of neuronal stochastic firing (referring to spontaneous activity unrelated to the memory retrieval), denoted as T, and the mean strength of synaptic connections, w. The diagram depicts three collective states ('phases'). In the low T, low load region, the retrieval process successfully constructs the target memory almost perfectly. In particular, at zero T, this occurs as long as the load per neuron is smaller than a critical value ~0.14, known as the memory capacity. In the intermediate regime, the network dynamics execute a retrieval process that culminates in a spurious memory (analogous to 'confabulation'). Finally, in the high noise, high load regime, the network fails to execute a retrieval process altogether – the stochastic dynamics of the network wander between states without converging on a particular fixed point.

The phase diagram of Figure 2 shows how the performance of the system scales with the number of neurons in the memory networks. For instance, consider a degenerative disease that results in the death of half of the neurons after the encoding of P memories. The information represented by the missing neurons is irretrievable. However, because the memory retrieval is associative, the shrinking size may affect the quality of retrieval of the information in the remaining neurons. Indeed, according to Figure 2, this occurs because both the load per neuron and the normalised neuronal stochasticity are increasing following the degeneration. Interestingly, the latter effect can be compensated for by increasing the overall strength of the synaptic connections [7]. These examples demonstrate the richness of the predictions that can be drawn from a rigorous theory of computation in a neural circuit. Our analysis also highlighted the deficiency of the model; primarily that it exhibits catastrophic forgetting: when the load increases above the capacity, all memories, not only new (or old) ones, are lost. Subsequent studies have studied mechanisms that allow for more graceful forgetting, including memory rehearsal (e.g., during sleep) and memory consolidation [8]. Attractor networks have been extensively studied as models of hippocampus-cortex memory systems [9-11].

The generative power of the brain: the ring attractor

The brain is not a passive device reacting to external stimuli. The active, generative power of the brain is evident from phenomena like planning, prediction, exploration, imagery and creativity. The repertoire of neural codes must therefore span intrinsically generated activity patterns that correspond to hypotheses about present or future states of the world. What are the neural mechanisms underlying these abilities?

Physics studies the spontaneous emergence of macroscopic structures in nonlinear dynamic systems operating under simple local rules. Examples include the shape of snowflakes, the patterns of clouds, the stripes in zebra skin and the hexagonal shape of the honeycomb. These ideas inspire neural models of spontaneous generation of spatial activity patterns in the brain and their relation to the neural code.

Accumulating information about the structure of cortical circuits revealed extensive connectivity between neurons in the cortex, surpassing in magnitude the drive from the upstream sensory pathways. This puzzle led me to hypothesise that recurrently connected cortical circuits generate a manifold of intrinsic activity patterns, each coding for a potential value of an external variable. When an external input arrives, it selects the cortical activity pattern most consistent with it. Through the intrinsic activity patterns and their modulation by stimuli, cortical circuits generate predictions about the external world, which are revised and updated according to incoming evidence from sensory signals [12].

I constructed a minimal neural circuit model that exhibits such capability. In this circuit, neurons are arranged in a ring geometry, where the location of each neuron codes for its preferred angle, such as the angle of a visual stimulus in sensory areas or the direction of movement of the animal's arm in motor areas. Neighbouring neurons exert excitatory influence on each other, whereas distal neurons inhibit each other. I have shown that this circuit exhibits a ring attractor, spontaneously forming persistent bumps of activity that can be located

anywhere on the ring (**Figure 3**). The model predicts the conditions for this pattern formation and describes how external stimuli interact with them, selecting or moving the bump from a present location to one which best fits the input spatial profile [12]. Ring attractors were proposed for multiple functions including visual working memory, voluntary movements and the head direction system where a neural circuit codes for the direction of the animal's heading [13].

FIGURE 3 The ring attractor and the toroidal attractor. (**A**) Connectivity in the ring attractor model. Neighbouring neurons on a ring excite each other, whereas distal neurons inhibit each other. The steady states of the activity profiles of the model lie on a ring manifold attractor. (**B**) The model predicts persistent activity bumps that can be located anywhere on a ring. Similar persistent activity bumps are observed in fly head-direction cells. (**C**) A 2-D extension of the ring attractor model generates a toroidal attractor. Experiments have shown that grid cell activities lie in a toroidal manifold.
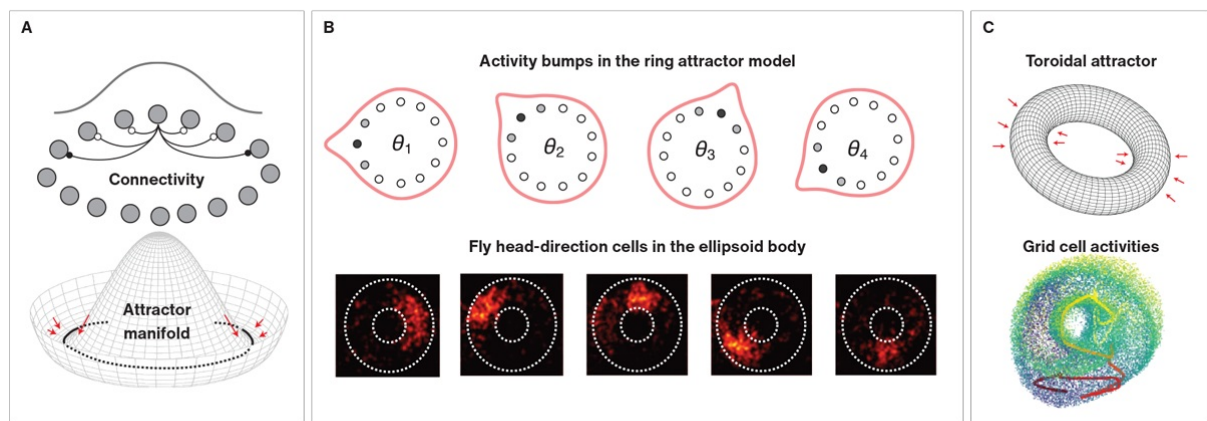

B) bottom [14]. C) bottom [19]

Direct experimental evidence of the ring attractor has emerged in recent years. In a beautiful set of experiments, neuroscientists imaged the activity of neurons in the fly 'compass' system and visualised bumps of activity moving around a ring of neurons, constantly aligning their position with visual or self-motion cues. Bumps appear spontaneously in the dark and during sleep [14]. Remarkably, the synaptic connections between the neurons were fully mapped at the nanoscale and were found to exhibit the same shape of synaptic connections predicted by the ring model [15]. Similar ring networks were recently observed in the head direction system of fish [16]. Experimental evidence of ring attractors was also found in the rodent head direction system [17], although connectivity information is still lacking.

A recent discovery in the entorhinal cortex found a brain region in which neurons code the position of the animal in the environment not by single bumps of activity but by multiple bumps arranged in a beautiful hexagonal lattice [18]. Such a regular pattern is an archetypal example of pattern formation. However, the original observations of these patterns were limited to the response field of single 'grid cells', leaving open the intriguing question of what the underlying circuit mechanism is. Theoretical work proposed that underlying the grid system is a neural circuit with an architecture characterised by two angles, a toroid, extending the ring architecture in the ring network. This network exhibits a 'toroidal attractor' where persistent activity bumps can be anywhere on the surface of the torus, with their location updated by self-motion or sensory cues. Indeed, recent experiments showed evidence of the existence of a toroidal manifold that persists even in the dark or during sleep (**Figure 4**) [19]. Direct evidence for the underlying connectivity awaits the mapping of the synaptic wiring of this system.

FIGURE 4 Connectivity and synaptic inputs in E-I balanced networks. [21] (**A**) Connectivity of an E-I network. (**B**) Both excitatory and inhibitory input currents are strong, but they balance out in the total input current. The total current fluctuates around the spiking threshold, generating irregular spikes.
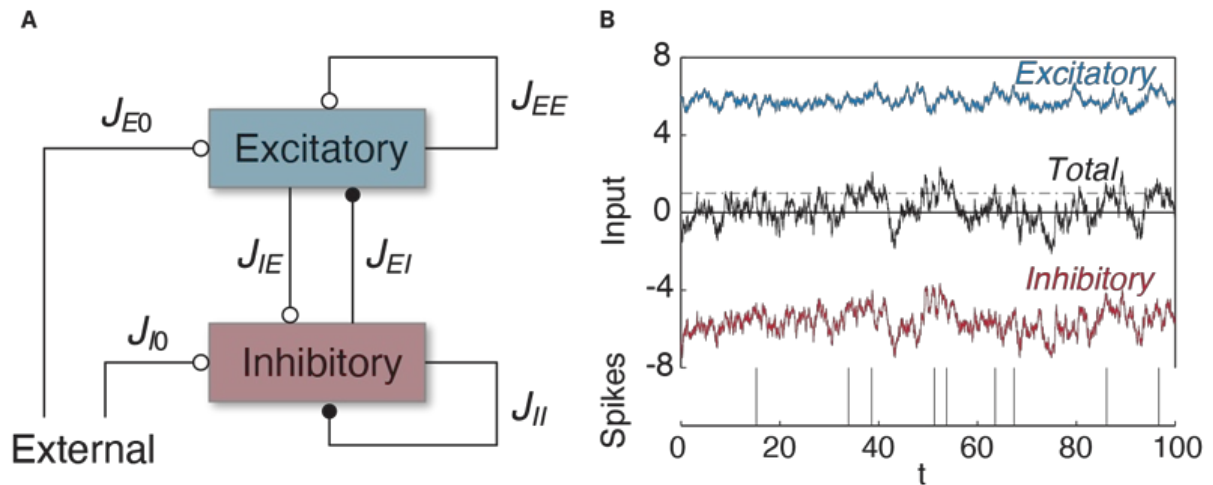

Overall, the experiments in the navigational system have been a remarkably fruitful testing ground for neural circuit theories. In this case, these theories link the collective neuronal activity patterns to the emergent neural code underlying the ongoing location and heading of animals, which is an essential ingredient of spatial navigation function. The remarkable experimental validation of key theoretical predictions demonstrates the pivotal role of theoretical neuroscience in unravelling brain computational principles through abstractions of the underlying microscopic complexity and by engaging in a fruitful dialogue with experimental studies of the brain.

Stability and stochasticity of brain dynamics: excitation-inhibition balance

Not all neuronal dynamics are characterised by regular, structured patterns coding for actual or predicted variables such as external stimuli or the position of the animal. It has long been empirically established that structured neuronal patterns are superimposed on a background of irregular activity, indicating that even in the absence of cognitive processes, the brain is not silent but exhibits spontaneous irregular activity. Furthermore, when engaged in cognitive functions, the induced activity patterns are embedded in a background of irregular neuronal firing. The origins and functions of these irregular activity patterns have remained elusive. Early theoretical work failed to account for this 'neural noise', as the dynamics of the proposed circuit models invariably exhibited regular deterministic activity, so the noise had to be injected ad-hoc into the network model, as was done with respect to neural noise in Figure 2 above.

The resolution to this enigma emerged when my then-postdoc the late Carl van Vreeswijk and I explored neural circuit models with synaptic connections that were substantially stronger than those assumed in previous models. Surprisingly, we discovered that networks with strong excitatory and inhibitory synaptic currents exhibit intrinsically generated irregular patterns akin to chaotic dynamics observed in physical systems. Our theoretical analysis revealed that the crucial mechanism is competition between strong, opposing forces – excitatory and inhibitory – within the circuit. For such neuronal circuits to stabilise, these opposing forces must dynamically balance themselves, allowing the network to operate in a healthy dynamic range, a state known as excitation-inhibition balance (**Figure 5**) [20, 21].

FIGURE 5 Neural manifolds in artificial neural networks and the brain. (**A**) Neural manifolds of different categories (cats and dogs). Each manifold represents a single category, variations within the manifold represent variables that do not change the category identity (such as different locations, orientations or sizes). (**B**) The radius of the object manifolds becomes smaller in deeper layers, both in artificial neural networks and along the visual hierarchy, making classification of different objects easier.
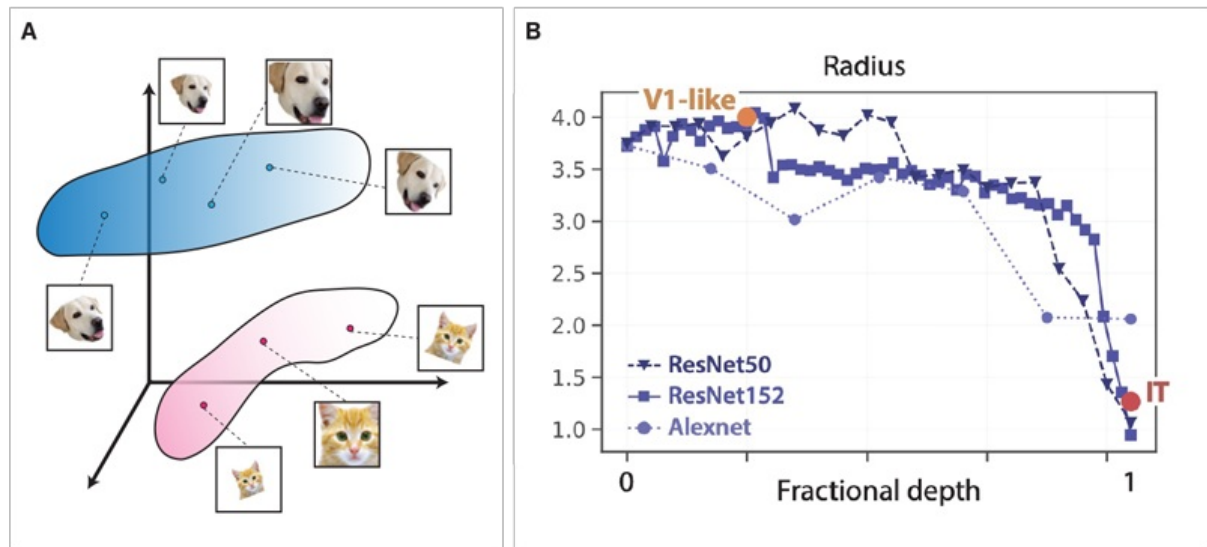
A & B)Chung S, Lee DD, Sompolinsky H-. Classification and Geometry of General Perceptual Manifolds Phys. Rev. X 8, 031003 – Published 5 July 2018

The mechanism of excitation-inhibition balance elucidates many aspects of cortical activity, including the observation that increases in neuronal firing are often accompanied by increases in both excitatory and inhibitory drives. Furthermore, the theory highlights the functional benefits of maintaining an active balance. Neural networks that remain quiescent at rest require a considerable delay to activate in response to a stimulus. In contrast, neurons in an active balanced state are often close to their firing threshold and can thus respond to stimuli or changes in conditions with very short reaction time. Can E-I balanced networks with their fast responses accommodate the slow dynamics of attractor networks? Interesting ongoing research is studying neural circuit models which combine E-I balance with attractor dynamics [22, 23].

Subsequent numerous computational and experimental studies have demonstrated that the balance between excitation and inhibition is a fundamental principle of brain dynamics and architecture. Complementary mechanisms for maintaining a robust E-I balance have been discovered including synaptic plasticity, adaptation and homeostatics [24-26]. Importantly, the theory predicts that deviations in the strength of excitatory and inhibitory connections beyond a certain range can destabilise the system, leading to either unstable bursts of activity akin to epileptic seizures due to unchecked excitation, or to excessively suppressed activity overwhelmed by strong inhibition, thereby shedding light on the mechanisms underlying circuit dysfunctions observed in neurological and psychiatric disorders [27-29].

Bridging neuroscience and artificial intelligence: neural manifolds of categories

Theoretical and experimental neuroscience research over the past decades has deepened our understanding of how neuronal dynamics translate into computation, primarily within local neural circuits. To understand

complex cognitive functions, we must investigate larger-scale networks, spanning multiple brain regions. However, in the past, large-scale (or 'deep') neural network models failed to exhibit cognitive capabilities close to animal and human performance.

The past decade has marked a significant turning point. Advances in digital databases and compute power have led to the construction of brain-inspired deep neural networks with unparalleled abilities, some at par or surpassing human cognitive performances, ushering in the artificial intelligence (AI) revolution. Overall, these networks operate on the same principles as those adopted in computational neuroscience. They rely on distributed computations with emergent behaviours that are the product of many massively interconnected simple units and learn from experience-dependent plasticity rather than from some top-down encoded rules or hard-wired knowledge. Interestingly, many artificial networks mirror brain-like functions such as long-term and working memory, attention, context dependence and reinforcement learning. Some generative AI models are powerful extensions of the manifolds of spontaneous, persistent patterns generated in the attractor models of computational neuroscience, described above.

The new artificial deep networks provide powerful models for the emergence of complex cognitive functions in large-scale brain circuits. This development is ushering in a new era of an extremely fruitful dialogue between AI and neuroscience.

An example of progress along this line from my recent research is the study of neural mechanisms underlying the representation of categories. While the input to the brain has the form of a continuous stream of analog sensory stimuli, humans and animals dissect the world into discrete categories, such as objects, words and concepts. Focusing on the visual cortex, it has been noted that even at the top stage, the inferior temporal (IT) cortex, neurons are selective not only to object identity in their input images, but also to such variables as location, orientation and size, which do not change the object identity. Interestingly, the same holds for the top layers of artificial deep networks trained on millions of images for object recognition tasks. This led to the hypothesis that the neural code of object identity is not an invariant response to object identity, but consists of manifolds of states, all corresponding to neural population responses to the variety of stimuli associated with the object. These manifolds are intertwined with each other ('entangled') in early stages of visual processing, but become well separated ('disentangled') in the final stages [30].

In recent work from my group [31-32], we formulated theoretically driven geometric measures of the object manifolds and derived the quantitative relation between these measures and the ability to recognise objects in an image and to rapidly learn to recognise novel objects. Surprisingly, both the visual cortex and the artificial deep network exhibit very similar manifold geometries, as demonstrated in Figure 5. This approach is now being extended to more complex tasks such as visual reasoning and language processing.

Concluding remarks

Two major developments are transforming the landscape of brain theory. One is the AI revolution, mentioned above. Another transformative development is the accumulation of massive amounts of high-quality data about brain structure and dynamics as well as animal behaviour (the evolution of 'Big Data' in neuroscience, neurology and neuroethology). For instance, the number of neurons that can be simultaneously recorded in the mouse has reached ~ 1 million (about 1 percent of the whole brain) with calcium imaging techniques [33], and ~ 10,000 with electrophysiology techniques [34]. Advances on the structural side include the complete mapping of all synaptic connections in the fly brain (the fly 'connectome'), exhaustive mapping of cell types in the mammalian retina, large-scale recordings from neurons in freely moving mice, and powerful data acquisition and algorithms for accurate estimates of body posture and limb movements. Large brain observatories are established to

aggregate molecular, structural and activity data of the nervous system, providing open access to researchers and opportunities for remote design of new experiments [35].

The availability of big data, the advances in sophisticated neural network mechanistic models, the emergence of new tools and new ideas from AI research, and contemporary brain research are poised to generate new breakthroughs into the complex neuronal machinery underlying cognitive functions and behaviours, shedding further layers away from the mysteries of the neural code.

Correspondence Haim Sompolinsky. E-mail: hsompolinsky@ mcb.harvard.edu

SUMMARY

Haim Sompolinsky

Deciphering the mysteries of the neural code

This essay is dedicated to the memory of my father David Sompolinsky. As a medical student in Veterinary Medicine in Copenhagen, with the support of his professors and the Danish Resistance, David organised the rescue of 700 Danish Jews in October 1943, helping them escape Nazi persecution and find safety in Sweden.

REFERENCES

1. Sompolinsky H. Computational neuroscience: beyond the local circuit. Curr Opin Neurobiol. 2014;25:xiii-xviii. https://doi.org/10.1016/j.conb.2014.02.002
2. Carbone A, Gromov M, Prusinkiewicz P. Pattern formation in biology, vision and dynamics. World Scientific; 2000. https://doi.org/10.1142/9789812817723
3. McKenzie RH. Condensed Matter Physics: A very short introduction. Oxford University Press; 2023. doi:10.1093/actrade/9780198845423.001.0001. https://doi.org/10.1093/actrade/9780198845423.001.0001
4. Hopfield JJ. Neural networks and physical systems with emergent collective computational abilities. Proc Natl Acad Sci. 1982;79:2554-8. https://doi.org/10.1073/pnas.79.8.2554
5. Amit DJ, Gutfreund H, Sompolinsky H. Spin-glass models of neural networks. Phys Rev A (Coll Park). 1985;32:1007. https://doi.org/10.1103/PhysRevA.32.1007
6. Amit DJ, Gutfreund H, Sompolinsky H. Storing infinite numbers of patterns in a spin-glass model of neural networks. Phys Rev Lett. 1985;55:1530. https://doi.org/10.1103/PhysRevLett.55.1530
7. Horn D, Levy N, Ruppin E. Memory maintenance via neuronal regulation. Neural Comput. 1998;10:1-18. https://doi.org/10.1162/089976698300017863
8. Shaham N, Chandra J, Kreiman G, Sompolinsky H. Stochastic consolidation of lifelong memory. Sci Rep. 2022;12:13107.

https://doi.org/10.1038/s41598-022-16407-9

9. 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. 2020;9:e56894. https://doi.org/10.7554/eLife.56894

10. Rolls ET, Kesner RP. A computational theory of hippocampal function, and empirical tests of the theory. Prog Neurobiol. 2006;79:1-48. https://doi.org/10.1016/j.pneurobio.2006.04.005

11. Kesner RP, Rolls ET. A computational theory of hippocampal function, and tests of the theory: new developments. Neurosci Biobehav Rev. 2015;48:92-147. https://doi.org/10.1016/j.neubiorev.2014.11.009

12. Ben-Yishai R, Bar-Or RL, Sompolinsky H. Theory of orientation tuning in visual cortex. Proc Natl Acad Sci. 1995;92:3844-8. https://doi.org/10.1073/pnas.92.9.3844

13. Khona M, Fiete IR. Attractor and integrator networks in the brain. Nat Rev Neurosci. 2022;23:744-66. https://doi.org/10.1038/s41583-022-00642-0

14. Seelig JD, Jayaraman V. Neural dynamics for landmark orientation and angular path integration. Nature. 2015;521:186-91. https://doi.org/10.1038/nature14446

15. Hulse BK, Haberkern H, Franconville R, et al. A connectome of the Drosophila central complex reveals network motifs suitable for flexible navigation and context-dependent action selection. Elife. 2021;10:e66039. https://doi.org/10.7554/eLife.66039

16. Petrucco L, Lavian H, Wu YK, Svara F, et al. Neural dynamics and architecture of the heading direction circuit in zebrafish. Nat Neurosci. 2023;26(5):765-73. https://doi.org/10.1038/s41593-023-01308-5

17. Chaudhuri R, Gerçek B, Pandey B, et al. The intrinsic attractor manifold and population dynamics of a canonical cognitive circuit across waking and sleep. Nat Neurosci. 2019;22:1512-20. https://doi.org/10.1038/s41593-019-0460-x

18. Hafting T, Fyhn M, Molden S, et al. Microstructure of a spatial map in the entorhinal cortex. Nature. 2005;436:801-6. https://doi.org/10.1038/nature03721

19. Gardner RJ, Hermansen E, Pachitariu M, et al. Toroidal topology of population activity in grid cells. Nature. 2022;602(7895):123-8. https://doi.org/10.1038/s41586-021-04268-7

20. Van Vreeswijk C, Sompolinsky H. Chaos in neuronal networks with balanced excitatory and inhibitory activity. Science. 1996;274:1724-6. https://doi.org/10.1126/science.274.5293.1724

21. van Vreeswijk C, Sompolinsky H. Chaotic balanced state in a model of cortical circuits. Neural Comput. 1998;10:1321-71. https://doi.org/10.1162/089976698300017214

22. Shaham N, Burak Y. Slow diffusive dynamics in a chaotic balanced neural network. PLoS Comput Biol. 2017;13:e1005505. https://doi.org/10.1371/journal.pcbi.1005505

23. Lin X, et al. Slow and weak attractor computation embedded in fast and strong EI balanced neural dynamics. Adv Neural Inf Process Syst. 2024;36.

24. Landau ID, Egger R, Dercksen VJ, et al. The impact of structural heterogeneity on excitation-inhibition balance in cortical networks. Neuron. 2016;92:1106-21. https://doi.org/10.1016/j.neuron.2016.10.027

25. Vogels TP, Abbott LF. Gating multiple signals through detailed balance of excitation and inhibition in spiking networks. Nat Neurosci. 2009;12:483-91. https://doi.org/10.1038/nn.2276

26. Rubin R, Abbott LF, Sompolinsky H. Balanced excitation and inhibition are required for high-capacity, noise-robust neuronal selectivity. Proc Natl Acad Sci. 2017;114:E9366-E9375. https://doi.org/10.1073/pnas.1705841114

27. Sohal VS, Rubenstein JLR. Excitation-inhibition balance as a framework for investigating mechanisms in neuropsychiatric disorders. Mol Psychiat. 2019;24:1248-57. https://doi.org/10.1038/s41380-019-0426-0

28. Tatti R, Haley MS, Swanson OK, et al. Neurophysiology and regulation of the balance between excitation and inhibition in neocortical circuits. Biol Psychiat. 2017;81:821-31. https://doi.org/10.1016/j.biopsych.2016.09.017

29. Dehghani N, Peyrache A, Telenczuk B, et al. Dynamic balance of excitation and inhibition in human and monkey neocortex. Sci Rep. 2016;6:23176. https://doi.org/10.1038/srep23176

30. DiCarlo, James J., and David D. Cox. Untangling invariant object recognition." *Trends in cognitive sciences* 11, no. 8 (2007): 333-341.

31. Cohen U, Chung S, Lee DD, Sompolinsky H. Separability and geometry of object manifolds in deep neural networks. Nat Commun. 2020;11(1):746. https://doi.org/10.1038/s41467-020-14578-5

32.   Sorscher B, Ganguli S, Sompolinsky H. Neural representational geometry underlies few-shot concept learning. Proc Natl Acad Sci U S A. 2022;119(43):e2200800119. https://doi.org/10.1073/pnas.2200800119

33.   Demas J, Manley J, Tejera F, et al. High-speed, cortex-wide volumetric recording of neuroactivity at cellular resolution using light beads microscopy [published correction appears in Nat Methods. 2021 Dec;18(12):1552]. Nat Methods. 2021;18(9):1103-11. doi:10.1038/s41592-021-01239-8 https://doi.org/10.1038/s41592-021-01239-8

34.   Steinmetz NA, Aydin C, Lebedeva A, et al. Neuropixels 2.0: A miniaturized high-density probe for stable, long-term brain recordings. Science. 2021;372(6539):eabf4588. doi:10.1126/science.abf4588 https://doi.org/10.1126/science.abf4588

35.   Koch C, Svoboda K, Bernard A, et al. Next-generation brain observatories. Neuron. 2022;110(22):3661-3666. https://doi.org/10.1016/j.neuron.2022.09.033