

Emergence of Brains

William Bialek 

Joseph Henry Laboratories of Physics, Princeton University, Princeton, New Jersey 08544, USA



(Received 22 May 2025; accepted 15 July 2025; published 7 August 2025)

As with many creative human activities, physics resists definition. Physics, as the saying goes, is what physicists do. But we can recognize major themes. One theme is the reductionist search for the building blocks of matter. In another direction is the exploration of the origins and history of the universe writ large. At intermediate scales, we explore the emergence of macroscopic behaviors that are more than the sum of their parts. An important lesson of the late 20th century is that these seemingly distinct themes are intertwined, so that theoretical physics today is more unified while at the same time reaching ambitiously across an even broader range of topics. The 2024 Nobel Prize in Physics, awarded jointly to John Hopfield and Geoffrey Hinton, celebrates the discovery of new classes of emergent phenomena, inspired by the function of brains and minds.

DOI: [10.1103/62p9-wn32](https://doi.org/10.1103/62p9-wn32)

On October 8, 2024, the Nobel Prize in Physics was awarded to John Hopfield and Geoffrey Hinton. In describing the scientific background, the Nobel committee began by saying that “with [this work] the boundaries of physics are extended to host phenomena of life as well as computation” [1]. This is a moment for the biological physics community to celebrate. It also is a moment to reflect on how we got here and where we are going.

I. FOUNDATIONS

In his landmark 1982 paper, Hopfield argued that computation should be seen as the global dynamics of a many-body system [2]. The system arrives at the result of the computation when it reaches a fixed point of these dynamics. An important example is associative or content-addressable memory, where the initial condition provides a hint or partial version of the pattern to be recalled, and the fixed point reflects the complete pattern.

Hopfield studied a simple model for dynamics in a network of neurons and found conditions under which these dynamics are equivalent to relaxation or sliding downhill on an energy surface; fixed points then are local minima of the energy. Many familiar systems have a single well-defined ground state and thus would be useless as computing devices because all initial conditions lead to the same answer. In the example of associative memory, functional behavior is possible only if the energy surface has many local minima corresponding to storing many memories. Further, if we want the system to store memories of particular things, then there must be mechanisms to sculpt the energy surface so that the local minima are found at states that represent these particular things.

Before Hopfield, we knew about systems that break a discrete symmetry and choose one out of a small number of possible ground states, as with many familiar magnets. We also knew about systems that break a continuous symmetry, which we can think of as storing a memory for one continuous variable. These classic categories of symmetry breaking are textbook topics in condensed matter physics [3–5]. Not long before Hopfield’s work, we came to understand models of spin glasses in which random interactions among binary variables lead to many locally stable states, but apparently there is no way of “programming” these states or storing particular memories [6]. The Hopfield model provides a simple example of a system with N degrees of freedom that has αN equivalent ground states and a rule for how the interaction parameters should become dynamical variables so as to specify the positions of these multiple ground states. Remarkably this learning rule captured at least schematically what was known to occur at real synapses in the brain.

As with many great ideas, Hopfield’s work had precursors and parallels. The idea of using statistical physics to describe emergent phenomena in networks of neurons can be traced back to metaphors from the 1950s [7]. More explicit statistical mechanics models for neural networks that learn and recall memories appeared in the 1970s in the work of Cooper, Little, Shaw, and others [8–10]. Also in the 1970s, Kohonen emphasized the matrix structure of associative memories [11], and Amari studied networks that perform associations, have stable states, and learn through algorithms quite similar to Hopfield’s [12,13].

Hopfield’s ideas about neural networks are part of a series of papers that set the agenda for the modern theoretical physicist’s engagement with the phenomena of life [14]. His work on hemoglobin elucidated the phenomenon of allostery in biological molecules, in ways that still resonate today [15,16]. His thinking about electron transfer in photosynthesis clarified rampant confusion about the distances over which the transfer occurs and the interplay of classical and quantum dynamics [17]. Most dramatically, his work on kinetic proofreading

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article’s title, journal citation, and DOI.

focused attention on the problem of precision in the transmission of information by copying from molecular templates and provided a mechanism to solve this problem [18]; closely related ideas were presented by Ninio at almost the same time [19]. In addition to its impact on our thinking about life, kinetic proofreading provided significant inspiration for a whole field of nonequilibrium statistical physics where information meets thermodynamics [20,21].

The kinetic proofreading idea is that cells build Maxwell demons, dissipating energy to sort molecular components with a precision beyond that which would be accessible in equilibrium. As a result, the arrangement of these components—the bases along DNA or messenger RNA, the pairings of transfer RNA molecules with cognate amino acids, the amino acids along the sequence of a protein—is highly atypical in the equilibrium ensemble.

In a similar spirit, learning drives neural networks toward parameters that are highly atypical for networks more generally. A compelling experimental demonstration of this is in the network that integrates vestibular inputs to control eye movements, where the discrete stored memories of the original Hopfield model are thought to merge into a “line attractor” [22]. A line attractor is a continuous range of network states that all are locally stable, and movement along the line serves to integrate inputs over time and hold a memory of this continuous variable. In this case, the position along the line also represents the position of the eyes. If we perturb the parameters of the system, either the memory will decay or the system will become unstable; therefore, the line attractor is a kind of critical point separating these qualitatively different behaviors. A simple version of virtual reality for goldfish allows manipulating the relationship between eye movements and image motion on the retina. If the brain learns to hold the relevant network near its critical point using visual feedback, then altering this feedback should drive the network away from criticality, toward memory loss or instability depending on the gain of the feedback loop. This works [23,24].

The original Hopfield model presented the problem of recalling a memory as minimizing an energy function, and the dynamics of the network carried out this minimization. Hopfield and Tank emphasized that many computational problems—both in computer science and in brain function—could be given this variational formulation [25], opening a path to think about neural computation more generally in terms of network dynamics [26,27]. The same idea of computation as relaxation on an energy function emerged with “simulated annealing” as an algorithm for practical optimization problems, such as component placement on a chip [28]. This triggered a more general statistical mechanics approach to optimization [29], which in turn generated insights into computational complexity [30]. This period also saw the emergence of statistical physics approaches to learning, using neural networks as examples [31,32]. Early exploration of models with higher-order interactions presaged the dense associative memories discussed below [33].

This foment connecting neuroscience, computation, and statistical physics continued for some time. In the midst of this, Ackley, Hinton, and Sejnowski introduced the Boltzmann machine [34]. Rather than having particular states of the network represent discrete memories as in the Hopfield

model, their idea was that all states of the network would represent possible states of the world. The goal of learning then is to drive the distribution of network states to match the distribution of world states, e.g., the distribution of images falling on the retina, and they gave simple algorithms that could do this.

As the name suggests, the dynamics of the Boltzmann machine are such that the distribution of network states is exactly the Boltzmann distribution for binary or Ising variables that represent the active/silent states of the individual neurons. The energy function in the Boltzmann distribution has the form of interactions between pairs of neurons, as in the Hopfield model or the more traditional Ising models that we study in our statistical mechanics courses [4,5]. But pairwise interactions are not enough to generate—for example—the distributions of images or words that we see in the real world. The Boltzmann machine can represent more general distributions by having both “visible” neurons that represent variables in the world and “hidden” neurons that are purely internal variables. Averaging over the states of the hidden units generates higher-order interactions among the visible units, allowing the network as a whole to match much more complicated distributions.

Boltzmann machines could be used in an analytic mode, where visible units are driven by external signals and the states of hidden units represent inferences drawn from these signals. This is even clearer in the restricted Boltzmann machine [35], where hidden and visible neurons interact with one another but not among themselves. Alternatively the network could be allowed to equilibrate, and states of visible units generate possible signals in the surrounding world. This dual nature of networks as analyzers and generators resonates across decades of development in artificial intelligence, up to the present.

II. FAST FORWARD TO THE PRESENT

The path from Hopfield networks to the Boltzmann machine is clear. The path from the Boltzmann machine to the current excitement in artificial intelligence (AI) is more complex. Hinton described the early work as a “historical enzyme,” which seems apt [36]. Most of modern AI is based on neural networks with a feed-forward architecture, more like the perceptron of the early 1960s than the Hopfield model of the 1980s. Foundational papers about perceptrons appeared in *Reviews of Modern Physics* [37,38], launching a separate thread of physicists’ engagement with AI. Those early systems had just one or two layers of processing, while modern “deep” networks have many layers. Large increases in depth, scale of training data, and computing power have led to enormous progress, culminating in human-level performance at image classification and other tasks [39]. It is remarkable, however, how much conceptually continuity we can see across the decades.

Hinton *et al.* discovered that the patterns of connections learned by a restricted Boltzmann machine as it builds a model for the distribution of inputs could be used in feed-forward networks that classify these inputs [40], and this was a major step in making the training of deep networks more efficient. Where the original models had interactions only between pairs of neurons, modern AI systems use a transformer

architecture with explicit higher-order interactions. But then there is a combinatorial explosion of possible interactions, and in transformers this is controlled by having individual neurons “paying attention” only to certain combinations of other neurons [41]. Transformers and attention are at the heart of current large language models and humanlike generative AI.

There are “modern Hopfield networks” that also allow for higher-order interactions. In these systems the combinatorial explosion is tamed by restricting the effective energy of N neurons to depend only on K projections of the N -dimensional network state; this projection structure actually applies to the original Hopfield model as well, where K is the number of memories being stored. The generalized models can store $K \propto N^k$ memories, with $k > 1$, or even $K \propto e^{\gamma N}$ [42]. The dynamics of these dense associative memories are connected to the attention mechanism in transformers [43]. Perhaps we have come full circle.

III. THE BOUNDARIES OF PHYSICS

Physics and biology have been intertwined since the start of modern science:¹ in the 17th century, microorganisms were among the first objects to be visualized with newly invented microscopes; in the 18th century, experiments on “animal electricity” proved foundational for our modern understanding of electricity more generally; in the 19th century, developments in optics and acoustics were continuous with the exploration of vision and hearing, and the laws of thermodynamics were not considered established until one could balance the energy budget for animal movements. Even with the spectacular successes of collaboration between physics and biology in the middle of the 20th century—the structure of DNA and proteins, the idea of the genetic code, and more—the status of this enterprise was unclear: Is it the application of methods from physics to the problems of biology? Is it some sort of interdisciplinary amalgam? Or can we talk about the physics of life the same way that we talk about the physics of the universe or the physics of correlated electrons?

Physics, at least in part, is a search for principles that are simple and universal. Biology, at least in part, is a celebration of the complexity and diversity of life. Hopfield’s work on kinetic proofreading [18], for example, showed how to engage with the complexity and diversity of particular biological systems and uncover simple and general issues that have the unambiguous look and feel of physics problems. If you read neuroscience or computer science papers from the early 1980s, it is impossible to read the introduction of the Hopfield model or the Boltzmann machine as being part of those disciplines.² They certainly are not applying the methods of

physics to solve the problems of biology. Instead they ask new questions in the intellectual tradition of physics, motivated by phenomena that traditionally are studied by biologists.³

By the turn of the 21st century, many more physicists were actively balancing their drive for simplicity and universality with the biologists’ appreciation of complexity and diversity. Experiments revealed quantitative regularities in examples ranging from single molecules to flocks of birds. Theorists connected formal, general principles with the details of these experiments. Progress required each physicist engaged in this work to connect with the community of biologists who studied the same system, but there was a growing ambition to connect with one another. The physics community more broadly began to see how the phenomena of life (once again) could provide challenges to physics itself, and biological physics emerged as a proper branch of physics [45].

In parallel with developments in biological physics came the realization that inference and computation are rich sources of problems in statistical mechanics, some of which have been mentioned above. A snapshot of these developments can be seen in the proceedings of a 1992 Cargèse workshop [46]. This is contemporary with the introduction of the physics e-print arXiv, which now has a category for “disordered systems and neural networks” as part of condensed matter physics, marking a literal redrawing of the physics landscape.

As we think about how the boundaries of physics have been extended, it is important that physicists are not colonialists or imperialists. The goal is not to displace or subjugate the practitioners of other disciplines. Rather we look to phenomena often studied by other communities and ask new questions, questions that pique our peculiar interests, and the pioneering work of Hopfield and Hinton was in this mode. As Hopfield titled his Nobel lecture: Physics is a point of view [47].

The physicists’ point of view is being brought to bear on more and different corners of the natural and engineered world, and the boundaries of physics are being extended further. Another lesson from the work recognized by the 2024 Nobel Prize is that it can take a full generation for the fruits of these explorations to ripen fully. Especially in a time of attacks on the scientific enterprise, it is important to support one another.

IV. WHERE ARE WE HEADED?

Prediction is difficult, especially about the future, but I have been asked to try. Throughout this paper references are meant to be illustrative rather than exhaustive, and this is especially true in this section.

A. Science enabled by AI

The day after the award to Hopfield and Hinton, half of the Nobel Prize in Chemistry was awarded to Demis Hassabis and

¹These developments deserve a full history, which as far as I know has not been written. Some perspectives can be found in the introductory sections of Refs. [44,45].

²The physics style of Hopfield’s work [2] may seem unsurprising, since he already was well known for his accomplishments in the traditional core of physics. It is perhaps more surprising how central physics ideas are to the Boltzmann machine paper [34], not least since it was published in *Cognitive Science*.

³I once asked Hopfield how he decided to change fields from condensed matter physics to biophysics. “I didn’t change,” he said, “I kept doing the same things and the fields moved past me.”

John Jumper, leaders of the team that built AlphaFold [48], an AI system that predicts protein structures from amino acid sequences. This is the clearest example thus far of AI solving a scientific problem, raising questions about what comes next. In the extreme, one might worry that these developments will progress to the point where science ceases to be a human activity.

A more modest and concrete point is that many steps needed to handle the flood of data in modern, large-scale experiments can now be done by training a neural network rather than through algorithms that are hand tuned by expert humans. In particle physics the integration of new methods is so complete that the regular community review of results—once a compact listing of masses, lifetimes, and branching ratios—now includes a chapter on machine learning [49]. AI tools also are becoming integral to large-scale astrophysical surveys, and a recent textbook provides a path into these developments [50].

While not at the scale of particle physics, we have seen an explosion of data on animal behavior. Throughout the 20th century psychophysicists quantified behavioral decisions in response to sensory inputs, while ethologists made qualitative observations on unconstrained behaviors in the wild. The physics of behavior combines these approaches, using modern imaging tools to provide more complete high-resolution views of movement in increasingly natural conditions and over increasingly long times [51,52]. The first problem is to reduce these large datasets to lower dimensional time series of meaningful variables. Before 2015 methods were hand tuned to each example, but machine learning makes it possible to convert, for example, megapixel movies into the trajectories of body parts or postures with minimal human input [53–55]. These methods are general, which means we can look forward to high-quality measurements of naturalistic movements in a wider range of organisms and contexts. This will provide an unprecedented opportunity for the physicists’ drive toward simplicity and generality to confront the complexity and diversity of animal behavior.

Data reduction is just a first step. While AI now is integrated into the particle physics analysis pipeline, nobody is proposing that these tools can fully replace humans in uncovering what the data mean. A controlled experiment would be to ask a deep network to produce the standard model of particle physics just from the data available up to ~ 1973 , and this seems unlikely to work (although we could be surprised). In the physics of behavior one goal is to discover dynamical equations that describe trajectories through the space of postures, and there has been considerable progress in the simple case of crawling motions in the worm *Caenorhabditis elegans* [56]. There are proposals to use machine learning to automate this discovery [57], and some features of *C. elegans* dynamics have been captured in this way [58].

The physics of behavior provides just one set of biological physics problems in which new science is being enabled by AI; an exhaustive review would be out of date almost as soon as it was finished. In data reduction there are open problems about interpretability, as in many uses of AI. The automated discovery of underlying dynamical equations is more ambitious and in many ways strikes at the core of what we mean by understanding.

B. Physics problems in AI

Qualitatively striking phenomena often have deep theoretical explanations. By any measure, the behavior of current AI systems is a striking phenomenon and captures our imagination not just as humans but as physicists. At some basic level, we do not understand why it works. In fact, it did not work until networks had enough degrees of freedom and datasets for training became large enough. Human level performance of AI systems thus is an emergent phenomenon in the same sense that we use the word in physics. Indeed, the discovery of scaling laws for system performance, similar to the behavior of materials near second-order phase transitions [4,5], provided part of the motivation for the push that led to ChatGPT-3 [59,60]. It thus seems plausible that there is a statistical physics description of modern AI systems, and of their learning dynamics, and this is being pursued in several directions. Pedagogical reviews characterize the interactions between physics and AI circa 2019, just before the emergence of large language models [61–63].

Practical learning algorithms are stochastic because examples are chosen at random, and there is progress on understanding the structure of this noise [64–66]. Early efforts to study learning using methods from the statistical mechanics of disordered systems are being dramatically extended to encompass new network architectures and highly structured data [67–72]. A very different direction grows from the realization that the dynamics of learning can be implemented in simpler systems, from molecular self-assembly to flow networks [73].

Central issues on which we can expect progress in the coming years include mechanisms for emergence of long timescales in large language models and the nature of the (astronomically) high-dimensional landscape for learning. As AI systems make the transition to more dramatic forms of generalization, such as in context learning [74], it becomes more urgent to give crisp definitions: can we find order parameters that quantify our impressions of increasingly humanlike performance? In the background is the need for tractable models that capture the statistical structure of real world signals.

C. Lessons for the physics of life

It matters that modern AI has roots in physics and especially in physicists’ efforts to capture the principles at work in living systems. Perhaps the most important idea is that there *are* physical principles to be found. Certainly this was Hopfield’s vision, not just with neural networks but with protein dynamics, electron transfer, and especially kinetic proofreading.

It is natural to connect artificial networks with the dynamics of real brains, and this program began almost immediately after Hopfield’s work [75]. The development of models with continuous attractors made it possible to see the control of eye movements [22–24] as being related to the sense of space encoded by the place and grid cells in rodents [76,77]. This conceptual unity became even clearer with the discovery in the fly’s brain of a ring attractor whose states represent direction, with neurons arranged on a literal ring [78,79]. There are efforts to identify the responses of neurons at different layers of the primate visual processing pathway with the activity of units at different layers in a deep network [80], and

deep learning methods are being used to construct semirealistic models for the circuitry of the retina as it responds to rich naturalistic stimuli [81]. Ising models equivalent to Boltzmann machines have had success in capturing the patterns of correlated activity observed in large populations of neurons [82]. These developments began when physics-style theories were seen as metaphors for the obviously more complicated mechanisms of real brains. Today these theories can describe data on real brains in quantitative detail. Theories may be swept away by new data, but we should not lose our drive for this precise connection between theory and experiment, so characteristic of physics' greatest successes.

A particular neural network—or a particular network of real neurons—is described by one set of parameters. But the earliest work on learning emphasized that we should think about an ensemble of networks with parameters chosen from a probability distribution, and it is this distribution that evolves as we learn [32]. In the simplest case, this distribution is a Boltzmann distribution in which the performance of the network plays the role of the (negative) energy, and this is the maximum entropy distribution consistent with a certain mean level of performance. One can also think about maximum entropy distributions of internal biochemical states that are consistent with a certain mean growth rate for bacterial cells [83–85], or an ensemble of genomes consistent with a certain mean fitness [86,87]. We can also view these distributions as priors on the space of parameters as we try to learn a model for the system's underlying mechanisms [88]. This view combines “top-down” and “bottom-up” approaches and provides a path to make quantitative inferences from data about the approach to optimality.

Training a modern network solves a high-dimensional, nonconvex optimization problem, and in many cases the end point of training is demonstrably a true optimum. With many near-degenerate optima, different networks that reach similar performance will be different, and training history will leave an imprint on the system. But historical constraints do not impede optimization: the relevant landscape of performance vs parameters is much more navigable than many people had guessed. Perhaps the landscape of fitness as a function of genotype also is more navigable than usually assumed, and there are hints of this from the direct experimental exploration of sequence space [89,90]. Better understanding of machine

learning will lead to a deeper view of the interplay among history, variability, and optimality in living systems.

V. CODA

We are nervous when our explanation of natural phenomena depends on finely tuned parameters. Neural networks provide a concrete example of how fixed parameters become dynamical variables, and then these dynamics drive the system to special points in parameter space.⁴ The emergent, functional behavior of the network rests entirely on this tuning. A similar tuning happens within each neuron as it adjusts the mixture of ion channels that it inserts into the cell membrane, shaping its intrinsic dynamics as an element of the network [96,97]. These phenomena in neurons and networks remind us of the protein folding problem—just as random connections or random channel mixtures typically are not functional for neural computation, random amino acid sequences typically do not fold into unique compact structures, and so the sequences of globular proteins that occur in nature must be highly nongeneric [98,99]. For both proteins and neural networks, function also does not require all parameters to be set precisely. None of this is what we expected in physics problems from before Hopfield and Hinton, and I do not think we have seen the full implications yet [100]. Exploring systems that learn, adapt, and evolve will lead us to new ideas about what is “natural” for living systems and for physical systems more generally.

ACKNOWLEDGMENTS

It is a special pleasure to thank J. Hopfield for teaching me so much about so many things, now over several decades. Thanks also to colleagues who gave advice on the manuscript. Remaining errors are mine alone.

DATA AVAILABILITY

No data were created or analyzed in this study.

⁴It is perhaps surprising that the same idea of dynamical tuning arises in particle physics and cosmology, to explain the near-zero value of (dimensionless) time-reversal violating terms in the strong interaction [91–93] and the very small value of the cosmological constant when it is measured in natural units [94,95].

-
- [1] The Nobel Committee for Physics, Royal Swedish Academy of Sciences, Scientific background to the Nobel Prize in Physics (2024), <https://www.nobelprize.org/prizes/physics/2024/advanced-information/>.
 - [2] J. J. Hopfield, Neural networks and physical systems with emergent collective computational abilities, *Proc. Natl. Acad. Sci. USA* **79**, 2554 (1982).
 - [3] P. W. Anderson, *Basic Notions of Condensed Matter Physics* (Benjamin/Cummings, Menlo Park, CA, 1984).
 - [4] J. P. Sethna, *Statistical Mechanics: Entropy, Order Parameters, and Complexity* (Oxford University, Oxford, 2021).
 - [5] S. A. Kivelson, J. M. Jiang, and J. Chang, *Statistical Mechanics of Phases and Phase Transitions* (Princeton University, Princeton, NJ, 2024).
 - [6] M. Mézard, G. Parisi, and M. A. Virasoro, *Spin Glass Theory and Beyond* (World Scientific, Singapore, 1987).
 - [7] B. G. Cragg and H. N. V. Temperley, The organisation of neurones: A co-operative analogy, *Electroencephalogr. Clin. Neurophysiol.* **6**, 85 (1954).
 - [8] L. N. Cooper, A possible organization of animal memory and learning, in *Collective Properties of Physical Systems: Proceedings of Nobel Symposium*, edited by B. Lundqvist

- and S. Lundqvist (Academic Press, New York, 1973), pp. 252–264.
- [9] W. A. Little, Existence of persistent states in the brain, *Math. Biosci.* **19**, 101 (1974).
 - [10] W. A. Little and G. L. Shaw, Analytic study of the memory storage capacity of a neural network, *Math. Biosci.* **39**, 281 (1978).
 - [11] T. Kohonen, Correlation matrix memories, *IEEE Trans. Comput.* **C-21**, 353 (1972).
 - [12] S. I. Amari, Learning patterns and pattern sequences by self-organizing nets of threshold elements, *IEEE Trans. Comput.* **C-21**, 1197 (1972).
 - [13] S. I. Amari, Neural theory of association and concept-formation, *Biol. Cybern.* **26**, 175 (1977).
 - [14] W. Bialek, Moving boundaries: An appreciation of John Hopfield, [arXiv:2412.18030](https://arxiv.org/abs/2412.18030).
 - [15] R. G. Shulman, S. Ogawa, and J. J. Hopfield, An allosteric model of hemoglobin: II. The assumption of independent binding, *Arch. Biochem. Biophys.* **151**, 68 (1972).
 - [16] J. J. Hopfield, Relation between structure, co-operativity and spectra in a model of hemoglobin action, *J. Mol. Biol.* **77**, 207 (1973).
 - [17] J. J. Hopfield, Electron transfer between biological molecules by thermally activated tunneling, *Proc. Natl. Acad. Sci. USA* **71**, 3640 (1974).
 - [18] J. J. Hopfield, Kinetic proofreading: A new mechanism for reducing errors in biosynthetic processes requiring high specificity, *Proc. Natl. Acad. Sci. USA* **71**, 4135 (1974).
 - [19] J. Ninio, Kinetic amplification of enzyme discrimination, *Biochimie* **57**, 587 (1975).
 - [20] C. H. Bennett, The thermodynamics of computation—A review, *Int. J. Theor. Phys.* **21**, 905 (1982).
 - [21] L. Peliti and S. Pigolotti, *Stochastic Thermodynamics: An Introduction* (Princeton University, Princeton, NJ, 2021).
 - [22] H. S. Seung, How the brain keeps the eyes still, *Proc. Natl. Acad. Sci. USA* **93**, 13339 (1996).
 - [23] G. Major, R. Baker, E. Aksay, B. Mensh, H. S. Seung, and D. W. Tank, Plasticity and tuning by visual feedback of the stability of a neural integrator, *Proc. Natl. Acad. Sci. USA* **101**, 7739 (2004).
 - [24] G. Major, R. Baker, E. Aksay, H. S. Seung, and D. W. Tank, Plasticity and tuning of the time course of analog persistent firing in a neural integrator, *Proc. Natl. Acad. Sci. USA* **101**, 7745 (2004).
 - [25] J. J. Hopfield and D. W. Tank, “Neural” computation of decisions in optimization problems, *Biol. Cybern.* **52**, 141 (1985).
 - [26] J. J. Hopfield and D. W. Tank, Computing with neural circuits: A model, *Science* **233**, 625 (1986).
 - [27] D. W. Tank and J. J. Hopfield, Collective computation in neuron-like circuits, *Sci. Am.* **257**, 104 (1987).
 - [28] S. Kirkpatrick, C. D. Gelatt, Jr., and M. Vecchi, Optimization by simulated annealing, *Science* **220**, 671 (1983).
 - [29] Y. Fu and P. W. Anderson, Application of statistical mechanics to NP-complete problems in combinatorial optimisation, *J. Phys. A: Math. Gen.* **19**, 1605 (1986).
 - [30] R. Monasson, R. Zecchina, S. Kirkpatrick, B. Selman, and L. Troyansky, Determining computational complexity from characteristic ‘phase transitions’, *Nature (London)* **400**, 133 (1999).
 - [31] E. Gardner, The space of interactions in neural network models, *J. Phys. A: Math. Gen.* **21**, 257 (1988).
 - [32] E. Levin, N. Tishby, and S. A. Solla, A statistical approach to learning and generalization in layered neural networks, *Proc. IEEE* **78**, 1568 (1990).
 - [33] E. Gardner, Multiconnected neural network models, *J. Phys. A: Math. Gen.* **20**, 3453 (1987).
 - [34] D. H. Ackley, G. E. Hinton, and T. J. Sejnowski, A learning algorithm for Boltzmann machines, *Cognit. Sci.* **9**, 147 (1985).
 - [35] P. Smolensky, Information processing in dynamical systems: Foundations of harmony theory, in *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Vol. 1: Foundations*, edited by D. E. Rumelhart and J. L. McClelland (MIT, Cambridge MA, 1986), pp. 194–281.
 - [36] G. Hinton, Boltzmann machines, Nobel Prize Lecture, 2024, <https://www.nobelprize.org/prizes/physics/2024/hinton/lecture/>.
 - [37] H. D. Block, The perceptron: A model for brain functioning. I, *Rev. Mod. Phys.* **34**, 123 (1962).
 - [38] H. D. Block, B. W. Knight, Jr., and F. Rosenblatt, Analysis of a four-layer series-coupled perceptron. II, *Rev. Mod. Phys.* **34**, 135 (1962).
 - [39] Y. LeCun, Y. Bengio, and G. Hinton, Deep learning, *Nature (London)* **521**, 436 (2015).
 - [40] G. E. Hinton, S. Osindero, and Y. W. Teh, A fast learning algorithm for deep belief nets, *Neural Comput.* **18**, 1527 (2006).
 - [41] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, I. Polosukhin, Attention is all you need, *Adv. Neural Inf. Process. Syst.* **30**, 5999 (2017).
 - [42] D. Krotov and J. J. Hopfield, Dense associative memory for pattern recognition, *Adv. Neural Inf. Process. Syst.* **29**, 1180 (2016).
 - [43] H. Ramsauer, B. Schäfl, J. Lehner, P. Seidl, M. Widrich, T. Adler, L. Gruber, M. Holzleitner, M. Pavlović, G. K. Sandve, V. Greiff, D. Kreil, M. Kopp, G. Klambauer, J. Brandstetter, and S. Hochreiter, Hopfield networks is all you need, [arXiv:2008.02217](https://arxiv.org/abs/2008.02217).
 - [44] W. Bialek, *Biophysics: Searching for Principles* (Princeton University, Princeton, NJ, 2012).
 - [45] W. Bialek, B. Carragher, I. Cissé, M. M. Desai, O. K. Dudko, D. I. Goldman, J. Kondev, P. B. Littlewood, A. J. Liu, M. E. Moxon, J. N. Onuchic, M. J. Schnitzer, and C. M. Waterman, *Physics of Life: A Consensus Study* (National Academies, Washington, DC, 2022).
 - [46] *From Statistical Physics to Statistical Inference and Back*, edited by P. Grassberger and J. P. Nadal (Kluwer Academic, Dordrecht, 1994).
 - [47] J. J. Hopfield, Physics is a point of view. Nobel Prize Lecture, 2024, <https://www.nobelprize.org/prizes/physics/2024/hopfield/lecture/>.
 - [48] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Židek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back *et al.*, Highly accurate protein structure prediction with AlphaFold, *Nature (London)* **596**, 583 (2021).
 - [49] R. L. Workman *et al.* (Particle Data Group), Review of particle physics, *Prog. Theor. Exp. Phys.* **2022**, 083C01 (2022).
 - [50] V. Acquaviva, *Machine Learning for Physics and Astronomy* (Princeton University, Princeton, 2023).

- [51] S. R. Datta, D. J. Anderson, K. Branson, P. Perona, and A. Leifer, Computational neuroethology: A call to action, *Neuron* **104**, 11 (2019).
- [52] C. A. Riiska, C. Nguyen, O. Peleg, and J. M. Rieser, The physics of animal behavior: Form, function, and interactions, *Ann. Rev. Condens. Matter Phys.* **15**, 325 (2024).
- [53] A. Mathis, P. Mamidanna, K. M. Cury, T. Abe, V. N. Murthy, M. W. Mathis, and M. Bethge, DeepLabCut: Markerless pose estimation of user-defined body parts with deep learning, *Nat. Neurosci.* **21**, 1281 (2018).
- [54] T. D. Pereira, D. E. Aldarondo, L. Willmore, M. Kislin, S. S.-H. Wang, M. Murthy, and J. W. Shaevitz, Fast animal pose estimation using deep neural networks, *Nat. Methods* **16**, 117 (2019).
- [55] T. D. Pereira, N. Tabris, A. Matsliah, D. M. Turner, J. Li, S. Ravindranath, E. S. Papadoyannis, E. Normand, D. S. Deutsch, Z. Y. Wang, G. C. McKenzie-Smith, C. C. Mitelut, M. D. Castro, J. D’Uva, M. Kislin, D. H. Sanes, S. D. Kocher, S. S.-H. Wang, A. L. Falkner, J. W. Shaevitz *et al.*, SLEAP: A deep learning system for multi-animal pose tracking, *Nat. Methods* **19**, 486 (2022).
- [56] T. Ahamed, A. C. Costa, and G. J. Stephens, Capturing the continuous complexity of behaviour in *Caenorhabditis elegans*, *Nat. Phys.* **17**, 275 (2021).
- [57] B. C. Daniels and I. Nemenman, Automated adaptive inference of phenomenological dynamical models, *Nat. Commun.* **6**, 8133 (2015).
- [58] B. C. Daniels, W. S. Ryu, and I. Nemenman, Automated, predictive, and interpretable inference of *Caenorhabditis elegans* escape dynamics, *Proc. Natl. Acad. Sci. USA* **116**, 7226 (2019).
- [59] J. Kaplan, S. McCandlish, T. Henighan, T. B. Brown, B. Chess, R. Child, S. Gray, A. Radford, J. Wu, and D. Amodei, Scaling laws for neural language models, [arXiv:2001.08361](https://arxiv.org/abs/2001.08361).
- [60] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. Ziegler, J. Wu, C. Winter, C. Hesse *et al.*, Language models are few-shot learners, *Adv. Neural Inf. Process. Syst.* **33**, 1877 (2020).
- [61] P. Mehta, M. Bukov, C. H. Wang, A. G. R. Day, C. Richardson, C. K. Fisher, and D. J. Schwab, A high-bias, low-variance introduction to machine learning for physicists, *Phys. Rep.* **810**, 1 (2019).
- [62] G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby, L. Vogt-Maranto, and L. Zdeborová, Machine learning and the physical sciences, *Rev. Mod. Phys.* **91**, 045002 (2019).
- [63] Y. Bahri, J. Kadmon, J. Pennington, S. S. Schoenholz, J. Sohl-Dickstein, and S. Ganguli, Statistical mechanics of deep learning, *Annu. Rev. Condens. Matter Phys.* **11**, 501 (2020).
- [64] Y. Feng and Y. Tu, The inverse variance–flatness relation in stochastic gradient descent is critical for finding flat minima, *Proc. Natl. Acad. Sci. USA* **118**, e2015617118 (2021).
- [65] F. Mignacco and P. Urbani, The effective noise of stochastic gradient descent, *J. Stat. Mech.* **2022**, 083405 (2022).
- [66] A. Sclocchi and M. Wyart, On the different regimes of stochastic gradient descent, *Proc. Natl. Acad. Sci. USA* **121**, e2316301121 (2024).
- [67] L. Tiberi, F. Mignacco, K. Irie, and H. Sompolinsky, Dissecting the interplay of attention paths in a statistical mechanics theory of transformers, [arXiv:2405.15926](https://arxiv.org/abs/2405.15926).
- [68] S. Goldt, M. Mézard, F. Krzakala, and L. Zdeborová, Modeling the influence of data structure on learning in neural networks: The hidden manifold model, *Phys. Rev. X* **10**, 041044 (2020).
- [69] J. Garnier-Brun, M. Mézard, E. Moscato, and L. Saglietti, How transformers learn structured data: Insights from hierarchical filtering, [arXiv:2408.15138](https://arxiv.org/abs/2408.15138).
- [70] A. Nguyen, D. J. Schwab, and V. Ngampruetikorn, Generalization vs. specialization under concept shift, [arXiv:2409.15582](https://arxiv.org/abs/2409.15582).
- [71] M. Kamb and S. Ganguli, An analytic theory of creativity in convolutional diffusion models, [arXiv:2412.20292](https://arxiv.org/abs/2412.20292).
- [72] A. Sclocchi, A. Favero, and M. Wyart, A phase transition in diffusion models reveals the hierarchical nature of data, *Proc. Natl. Acad. Sci. USA* **122**, e2408799121 (2025).
- [73] M. Stern and A. Murugan, Learning without neurons in physical systems, *Ann. Rev. Condens. Matter Phys.* **14**, 417 (2023).
- [74] C. Goddard, L. M. Smith, V. Ngampruetikorn, and D. J. Schwab, When can in-context learning generalize out of task distribution? [arXiv:2506.05574](https://arxiv.org/abs/2506.05574).
- [75] D. J. Amit, *Modeling Brain Function: The World of Attractor Neural Networks* (Cambridge University, Cambridge, England, 1989).
- [76] A. Treves, O. Miglino, and D. Parisi, Rats, nets, maps, and the emergence of place cells, *Psychobiology* **20**, 1 (1992).
- [77] K. J. Yoon, M. A. Buice, C. Barry, R. Hayman, N. Burgess, and I. R. Fiete, Specific evidence of low-dimensional continuous attractor dynamics in grid cells, *Nat. Neurosci.* **16**, 1077 (2013).
- [78] J. D. Seelig and V. Jayaraman, Neural dynamics for landmark orientation and angular path integration, *Nature (London)* **521**, 186 (2015).
- [79] S. S. Kim, H. Rouault, S. Druckmann, and V. Jayaraman, Ring attractor dynamics in the *Drosophila* central brain, *Science* **356**, 849 (2017).
- [80] D. L. K. Yamins and J. J. DiCarlo, *Nat. Neurosci.* **19**, 356 (2016).
- [81] N. Maheswaranathan, L. T. McIntosh, H. Tanaka, S. Grant, D. B. Kastner, J. B. Melander, A. Nayeibi, L. E. Brezovec, J. H. Wang, S. Ganguli, and S. A. Baccus, Interpreting the retinal neural code for natural scenes: From computations to neurons, *Neuron* **111**, 2742.E4 (2023).
- [82] L. Meshulam and W. Bialek, Statistical mechanics for networks of real neurons, [arXiv:2409.00412](https://arxiv.org/abs/2409.00412) [Rev. Mod. Phys. (to be published)].
- [83] S. F. Taylor, N. Tishby, and W. Bialek, Information and fitness, [arXiv:0712.4382](https://arxiv.org/abs/0712.4382).
- [84] D. D. Martino, F. Capuani, and A. D. Martino, Growth against entropy in bacterial metabolism: The phenotypic trade-off behind empirical growth rate distributions in *E. coli*, *Phys. Biol.* **13**, 036005 (2016).
- [85] D. De Martino, A. M. Andersson, T. Bergmiller, C. C. Guet, and G. Tkačik, Statistical mechanics for metabolic networks during steady state growth, *Nat. Commun.* **9**, 2988 (2018).
- [86] N. H. Barton and H. P. de Vladar, Statistical mechanics and the evolution of polygenic quantitative traits, *Genetics* **181**, 997 (2009).

- [87] K. Bod'ová, G. Tkačik, and N. H. Barton, A general approximation for the dynamics of quantitative traits, *Genetics* **202**, 1523 (2016).
- [88] W. Młynarski, M. Hledák, T. R. Sokolowski, and G. Tkačik, Statistical analysis and optimality of neural systems, *Neuron* **109**, P1227.E5 (2021).
- [89] M. Carneiro and D. L. Hartl, Adaptive landscapes and protein evolution, *Proc. Natl. Acad. Sci. USA* **107**, 1747 (2010).
- [90] A. Papkou, L. Garcia-Pastor, J. Escudero, and A. Wagner, A rugged yet easily navigable fitness landscape, *Science* **382**, eadh3860 (2023).
- [91] R. D. Peccei and H. R. Quinn, CP conservation in the presence of pseudoparticles, *Phys. Rev. Lett.* **38**, 1440 (1977).
- [92] S. Weinberg, A new light boson? *Phys. Rev. Lett.* **40**, 223 (1978).
- [93] F. Wilczek, Problem of strong P and T invariance in the presence of instantons, *Phys. Rev. Lett.* **40**, 279 (1978).
- [94] L. F. Abbott, A mechanism for reducing the value of the cosmological constant, *Phys. Lett. B* **150**, 427 (1985).
- [95] P. J. Steinhardt and N. Turok, Why the cosmological constant is small and positive, *Science* **312**, 1180 (2006).
- [96] G. LeMasson, E. Marder, and L. F. Abbott, Activity-dependent regulation of conductances in model neurons, *Science* **259**, 1915 (1993).
- [97] G. Turrigiano, L. F. Abbott, and E. Marder, Activity-dependent changes in the intrinsic properties of cultured neurons, *Science* **264**, 974 (1994).
- [98] J. N. Onuchic, Z. Luthey-Schulten, and P. G. Wolynes, Theory of protein folding: The energy landscape perspective, *Annu. Rev. Phys. Chem.* **48**, 545 (1997).
- [99] H. Li, C. Tang, and N. S. Wingreen, Are protein folds atypical? *Proc. Natl. Acad. Sci. USA* **95**, 4987 (1998).
- [100] M. Bauer, W. Bialek, C. Goddard, C. M. Holmes, K. Krishnamurthy, S. E. Palmer, R. Pang, D. J. Schwab, and L. Susman, Optimization and variability can coexist, [arXiv:2505.23398](https://arxiv.org/abs/2505.23398).