# Inspires effective alternatives to backpropagation: predictive coding help understanding and building learning

Zhenghua Xu[*, #], Miao Yu[#], Yuhang Song

Artificial neural networks are capable of machine learning by simulating the hierarchical structure of the human brain. To enable learning by brain and machine, it is essential to accurately identify and correct the prediction errors, referred to as credit assignment (Lillicrap et al., 2020). It is critical to develop artificial intelligence by understanding how the brain deals with credit assignment in neuroscience.

As a theory of efficient credit assignment, backpropagation enables roaring in artificial neural networks (Rumelhart et al., 1986), where the patterns of activity as observed in the cortex are reproduced (Lillicrap et al., 2013). Currently, the models based on backpropagation have been widely applied to describe brain learning in large-scale cognitive tasks. In the process of backpropagation, the change in synaptic weights is a complex function of weights and activities of neurons that are not directly associated with the modified synapse. In contrast, the biological synaptic changes are determined by only the activity of pre-synaptic and post-synaptic neurons. Backpropagation does not necessarily follow the rules of brain learning to update weights and disseminate information (Lillicrap et al., 2016), which affects the biological plausibility of backpropagation in neuroscience. This prompts the attempt to construct a biologically credible model that can inspire a potentially better solution to credit assignment than backpropagation.

**Energy-based predictive coding:** Energy-based models (EBMs) are the models that can capture the correlation between variables by imposing range-limited energies to each configuration of the variables, which can be viewed as the measures of similarity or fit between data samples and models (Scellier et al., 2023). Relying on local learning, EBMs are highly flexible, versatile, and expressive (i.e., weight updates are local). According to the conclusions of statistical mechanics, any probability distribution can be transformed into EBMs, which are an important concept in relation to machine learning.

Studies show that EBMs are applicable to process model information in different brain networks at the neuronal level (Auksztulewicz et al., 2016). The application of EBMs to replace backpropagation lays a new foundation for learning, which highlights their potential to replace backpropagation. Distinct from the backpropagation-based artificial neural networks in which weights change directly to minimize the errors on output neurons from the target pattern, EBMs first clamp output neurons to the target, then modify neural activity to a configuration that minimizes a predefined energy function and finally update the weights to further decrease the same energy function.

Predictive coding networks (PCNs) are widely used EBMs in which one or more signals are used to predict the next signal, before coding of the difference between actual value and predicted error. Compared with other EBMs, PCNs are advantageous in the robustness to interference, flexibility, and simplicity of implementation. Research shows that predictive coding is closely related to backpropagation in layered networks and the same prediction is made with the same weights and input pattern (Salvatori et al., 2022). Predictive coding enables the description of a network architecture in which a particularly simple neural implementation is sufficient for such learning (Whittington et al., 2017). Therefore, PCN is a widely used framework for the description of information processing in the cortex (Rao et al., 1999), where each cortical area estimates both latent sensory states and actions, with the consequence of actions predicted by the cortex as a whole at multiple hierarchical levels (Rao et al., 2024). By minimizing predictive errors, PCN accounts for neural responses in the brain from a functional perspective. When describing the processing of information by the brain, PCNs follow the general principles including: (1) the brain is organized into a hierarchy of areas; (2) each area predicts the activity at a lower level; (3) prediction errors drive relaxation and learning.

**Limitations of predictive coding:** Nevertheless, some features of PCNs are arguably inconsistent with the known properties of the corresponding biological networks. One is that there is a one-to-one connection between values and error nodes. The other two relate to the symmetric forward-backward weights and nonlinear functions that affect only some outputs of the neurons.

Desirably, some recent studies show that the variants of predictive coding can be developed without the aforementioned implausible elements. Therefore, more and more scholarly attention has been drawn to exploring the variants of predictive coding. In a study (Salvatori et al., 2021), it was proposed that generative PCNs can implement associative memories like the brain and that the supervised restored task of low-level vision can be accomplished. In another study (Salvatori et al., 2022), predictive coding was extended to multilayer perceptrons for reverse differentiation, approximating to backpropagation. In a subsequent study (Rao et al., 2024), it was proposed that the neocortex implements active predictive coding and that it can be used to explain cortex activity of the brain.

**Prospective configuration in predictive coding:** There are two main stages in brain learning to external stimuli: "learn" and "predict";

the difference between them lies in whether supervisory signals exist. In brain learning, error signals exist only in the last layer of neural network and they are transmitted to the whole system through energy-diffusion. While, error signals exist independently in backpropagation, which causes its transmission from the last layer to all preceding layers. Thus, the weight update of backpropagation is affected by all the neurons in the system.

Physics knowledge is exercised to understand PCNs, where the energy-based system tends to maintain a low level of energy until equilibrium. In this system, "predict" is to fix one side of the energy system while allowing it to relax by moving nodes (**Figure 1A**). Since error signals have been transmitted into the system, the weight update in PCNs requires only the reduction in local potential energy for convergence of the system. "Learn" is to fix both sides of the energy system and relax it first by moving nodes and then tuning rods (**Figure 1B**). At this moment, PSNs first reduce the potential energy of the system by changing the activation, and then synaptic weight changes until the system converges. **Figure 1** shows the essence of energy-based networks. The relaxation before weight modification during learning allows the network to settle to a new configuration of neural activities, corresponding to those that would have come about after the error was corrected by weight modification, namely prospective configuration.

**Differences between prospective configuration in predictive coding and backpropagation:** To learn from feedback efficiently often requires changes in synaptic weights in many cortical areas (Whittington et al., 2017). Assume that there is a bear fishing aside a river. In this case, an important indicator of successful fishing is the combination of water sound and fish smell in the bear's brain. One day, the bear got its ears injured, which means the synaptic weights need to be modified not only in the auditory region but also in the associative (smell) and visual regions.

Backpropagation requires a global control signal to trigger computation. As for gradients, they must be the backworks sequentially computed via the computational graph (Salvatori et al., 2024). In the above example, backpropagation would proceed by backpropagating the negative errors to reduce the weights on the path between visual and auditory neurons. Consequently, the bear's expectation of smell the next time is affected.

In contrast, it is assumed by the PCNs with a prospective configuration that learning begins from re-configuring the neurons into a new configuration. Then, the weights are modified to produce and consolidate them to such a stable configuration. This configuration corresponds to the activity supposed to arise after learning, namely "prospective." Due to this prospective configuration, predictive coding can be performed to "anticipate" the side effects of potential weight modifications and compensate for them dynamically while maintaining the correct output. Therefore, the side effects of learning an association within only one learning iteration can be corrected through prospective configuration, while backpropagation requires multiple rounds of iteration (Song et al., 2024).
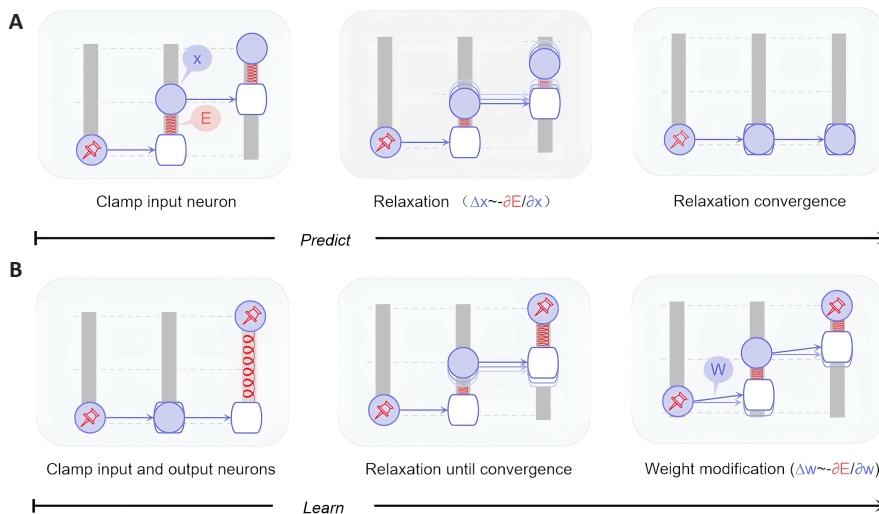
**Figure 1 | Energy system helps understand prospective configuration mechanism in the stages of "predict" (A) and "learn" (B).**
The neuron "x" is indicated by a solid circle, the activity of a neuron reflects the height of a node sliding on the post, and the input to the neuron is represented by a hollow node on the same post. A synaptic connection corresponds to a rod pointing from a solid to a hollow node. The arrow indicates the direction in which weight is transmitted at the synapse, and weight determines how the input to a postsynaptic neuron is determined by the activity of a pre-synaptic neuron. The pin indicates that neural activity is fixed to the input or target pattern. Neural activity maps to the vertical position of a node, synaptic connections map to rods that point from one node to another, and synaptic weight determines the relationship between the terminal and initial positions of rods. In the predictive coding networks with prospective configuration, relaxation, and weight modification are both driven by the minimization of system energy. Created with Microsoft PowerPoint.

**\*Correspondence to:** Zhenghua Xu, PhD,
xuzhenghua1987@gmail.com.
https://orcid.org/0000-0002-6719-7333
(Zhenghua Xu)
#These authors contributed equally to this work.

# References

Auksztulewicz R, Karl F (2016) Repetition suppression and its contextual determinants in predictive coding. Cortex 80:125-140.

Lillicrap TP, Stephen HS (2013) Preference distributions of primary motor cortex neurons reflect control solutions optimized for limb biomechanics. Neuron 77:168-179.

Lillicrap TP, Cownden D, Tweed DB, Akerman CJ (2016) Random synaptic feedback weights support error backpropagation for deep learning. Nat Commun 7:13276.

Lillicrap TP, Santoro A, Marris L, Akerman CJ, Hinton G (2020) Backpropagation and the brain. Nat Rev Neurosci 21:335-346.

Rao R, Ballard D (1999) Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. Nat Neurosci 2:79-87.

Rao R (2024) A sensory-motor theory of the neocortex. Nat Neurosci 27: 1221-1235.

Rumelhart DE, Hinton GE, Williams RJ (1986) Leaming internal representations by error propagation. Nature 323:533-536.

Salvatori T, Song Y, Hong Y, Sha L, Frieder S, Xu Z, Bogacz R, Lukasiewicz T (2021) Associative memories via predictive coding. Adv Neural Inf Process Syst 34:3874-3886.

Salvatori T, Song Y, Xu Z, Lukasiewicz T, Bogacz R (2022) Reverse differentiation via predictive coding. Proc AAAI Conf Artif Intell 36:8150-8158.

Salvatori T, Song Y, Yordanov Y, Millidge B, Xu Z, Sha L, Emde C, Bogacz R, Lukasiewicz T (2024) A stable, fast, and fully automatic learning algorithm for predictive coding networks. arXiv doi:10.48550/arXiv.2212.00720.

Scellier B, Ernoult M, Kendall J, Kumar S (2023) Energy-based learning algorithms for analog computing: a comparative study. arXiv [preprint] doi:10.48550/arXiv.2312.15103.

Song Y, Lukasiewicz T, Xu Z, Bogacz R (2020) Can the brain do backpropagation?-Exact implementation of backpropagation in predictive coding networks. Adv Neural Inf Process Syst 33:22566-22579.

Song Y, Millidge B, Salvatori T, Lukasiewicz T, Xu Z, Bogacz R (2024) Inferring neural activity before plasticity as a foundation for learning beyond backpropagation. Nat Neurosci 27:348-358.

Whittington, JCR, Bogacz R (2017) An approximation of the error backpropagation algorithm in a predictive coding network with local hebbian synaptic plasticity. Neural Comput 29:1229-1262.

*C-Editors: Zhao M, Sun Y, Qiu Y; T-Editor: Jia Y*

**Evidence and effectiveness of prospective configuration in predictive coding:** Bear's example shows that learning about one stimulus affects memory for another stimulus, while the brain can reduce this impact by inferring a latent state of the environment from the feedback. Considering the superior performance and biological plausibility of the prospective configuration, it is expected that its signatures can be observed through the learning behavior of animals. In a study by Song et al. (2024), the modeling analysis of human motion capture, human reinforcement learning, and mice-conditioned reflex was conducted to reveal that this inference can be reproduced through prospective configuration instead of BP in neural circuits.

In another study (Song et al., 2024), comparative experiments were also conducted on various learning situations in biological systems, including online learning, continual learning of multiple tasks, learning in changing environments, and small amount of data learning. In all of the aforementioned scenarios of learning, the experimental results show that prospective configuration is suitable to deal with various learning problems encountered by biological systems. The PCNs with prospective configuration demonstrate a notable advantage over backpropagation by reducing interference, with only local computation and biological plasticity required.

**Conclusions:** To find out the learning principle of biological neural systems and reverse engineer them as algorithms or even specialized hardware, such lines of research play a positive role in improving our understanding of the brain and true artificial intelligence. Given the advantages of prospective configuration, it may be applicable in machine learning to improve the efficiency and performance of deep neural networks, which enhances the connection between machine learning and the neuroscience community.

The research in predictive coding with prospective configuration can drive the development of brain science, as reflected in the following three aspects: (1) Brain cognition, which supports humans to better understand the mechanism of the brain in the process of prediction and learning; (2) Brain-inspired artificial intelligence, which provides a new guiding principle for the design of more efficient and self-adaptive neural network structure; (3) Brain disease diagnosis, which provides not only a new perspective on and strategy of exploring brain disease, but also new ideas for the diagnosis and treatment of various brain diseases such as neurodegeneration and cognitive impairment.

Future studies aim to close the gaps between abstract models and real brains for a better understanding of how prospective configuration is implemented in anatomically identified cortical networks. Due to the essential difference in the way to run machine learning and biological brains, it is time-consuming to simulate the PCNs with prospective configuration on the existing computers. A new type of computer, or specialized for brain-inspired hardware is thus needed to achieve prospective configuration rapidly with low power consumption.

**Zhenghua Xu\*,#, Miao Yu#, Yuhang Song**
School of Health Sciences and Biomedical Engineering, Hebei University of Technology, Tianjin, China (Xu Z, Yu M)
Department of Computer Science, University of Oxford, Oxford, UK  (Xu Z, Song Y)