Evolutionary learning in the brain by heterosynaptic plasticity

Zedong Bi[1,2,*], Guozhang Chen[3], Dongping Yang[4,5], Yu Zhou[6]

[1]Institute for Future, Qingdao University, Qingdao 266071, China

[2]School of Automation, Shandong Key Laboratory of Industrial Control Technology, Qingdao University, Qingdao 266071, China

[3]Institute for Theoretical Computer Science, Graz University of Technology, 8010 Graz, Austria

[4]Quanzhou Institute of Equipment Manufacturing, Haixi Institute, Chinese Academy of Sciences, Quanzhou 362200, China

[5]Fujian Institute of Research on the Structure of Matter, Chinese Academy of Sciences, Fuzhou 350002, China

[6]Department of Physiology and Pathophysiology, School of Basic Medical Sciences, Qingdao University, Qingdao 266071, China

*zedong.bi@outlook.com

Abstract

The way in which the brain modifies synapses to improve the performance of complicated networks remains one of the biggest mysteries in neuroscience. Existing proposals lack sufficient experimental support, and neglect inter-cellular signaling pathways ubiquitous in the brain. Here we show that the heterosynaptic plasticity between hippocampal or cortical pyramidal cells mediated by diffusive nitric oxide and astrocyte calcium wave, together with flexible dendritic gating of somatostatin interneurons, implies an evolutionary algorithm (EA). In simulation, this EA is able to train deep networks with biologically plausible binary weights in MNIST classification and Atari-game playing tasks up to performance comparable with continuous-weight networks trained by gradient-based methods. Our work leads paradigmatically fresh understanding of the brain learning mechanism.

Introduction

Understanding the cellular-level learning mechanisms underlying brain intelligence is a grand task of neuroscience. Inspired by the great success of gradient-based algorithms in training artificial neural networks, a large body of literature has dedicated to biologically plausible mechanisms to fulfill gradient-based learning in the brain (Lillicrap et al., 2020). For example, error signals may be back-propagated by action potentials in dendrites (Schiess et al., 2016) or through feedback synaptic connections (Lillicrap et al., 2016; Payeur et al., 2021). Gradient-based algorithms can also be effectively realized by exploiting the correlation between the perturbed fluctuation of neuronal activities or synaptic weights with a broadcasted scalar reward (Fiete and Seung, 2006; Miconi, 2017; Legenstein et al., 2010), or by back-propagating the training target (instead of errors) toward lower layer through feedback connections (LeCun, 1986; Lee et al., 2015). Recently, it is also suggested that evolution may endow each synapse a specific set of parameters of Hebbian learning, so that the brain can learn in a manner that different synapses change in different rules according to its pre- and post-synaptic activities (Miconi et al., 2020; Najarro and Risi, 2020).

However, as mentioned in Lillicrap et al. (2020), none of the proposed mechanisms above has been experimentally confirmed. Besides, these mechanisms are proposed under the guidance of the intuition of locality, which means that the change of a synapse should only depend on its pre- and post-synaptic activities, at most up to a globally available reward signal. Such intuition, however, neglects the ubiquitous inter-cellular signaling pathways in the brain, making the proposed cellular-level learning theory unlikely to be complete. For example, it is

well known that the diffusive nitric oxide (NO) associated with a potentiated synapse is able to potentiate a neighboring synapse (Schuman and Madison, 1994; Jacoby et al., 2001); the pre-synaptic release of glutamate is able to stimulate long-range propagating calcium waves in astrocytes, depressing synapses in a broader spatial range (Chen et al., 2013; Deitmer et al., 1998). The computational roles of such heterosynaptic plasticity remain unclear. It has been proposed that heterosynaptic plasticity may stabilize the run-away dynamics introduced by Hebbian-type rules and enhance synaptic competition (Chistiakova et al., 2014; Chistiakova and Volgushev, 2009), which seems, though important, auxiliary for learning. It is unknown whether heterosynaptic plasticity implies a standalone learning rule, and if it does, how its performances are in complex tasks.

Here, we collect various cellular-level experimental evidence (including heterosynaptic potentiation mediated by NO (Schuman and Madison, 1994; Jacoby et al., 2001), heterosynaptic depression mediated by astrocyte calcium wave (Chen et al., 2013; Deitmer et al., 1998), dendritic gating (Kepecs and Fishell, 2014), binary synapses (O'Connor et al., 2005), meta-plasticity (Duszkiewicz et al., 2019; Hulme et al., 2013), dopamine gating of synaptic plasticity (Roelfsema and Holtmaat, 2018), etc.) into a coherent picture, and show that the heterosynaptic plasticity of the synapses between hippocampal or cortical pyramidal cells implies an evolutionary algorithm (EA) (Mirjalili, 2019; Gao et al., 2019). In simulations, this EA is able to train neural networks with biologically plausible binary synaptic weights in MNIST classification and Atari-game playing tasks up to performance comparable with continuous-weight networks trained by gradient-based algorithms. Therefore, EA induced by heterosynaptic plasticity is a biologically-supported and powerful brain learning paradigm.

Results

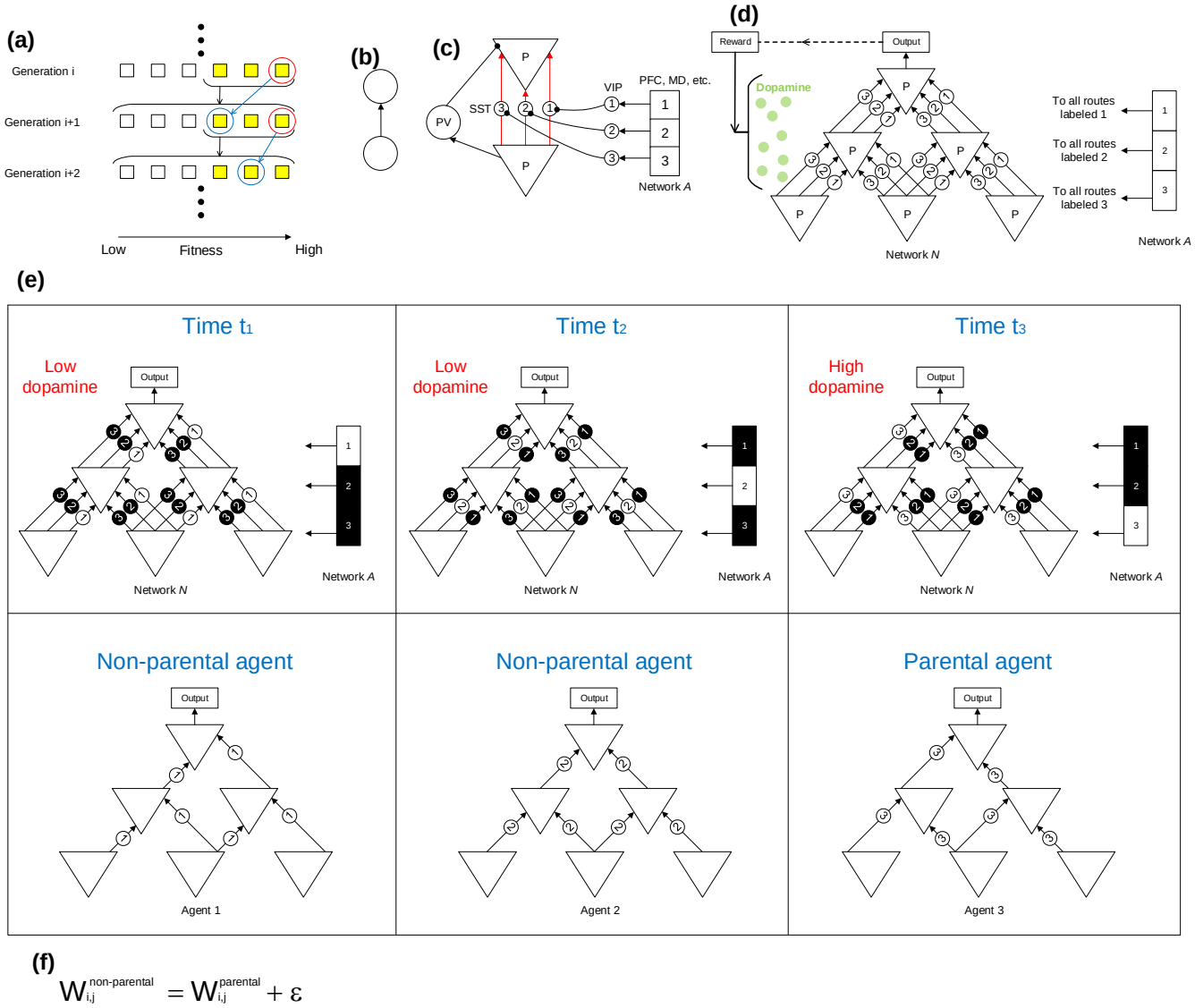
Network structure

In traditional computational models of neuroscience, two neurons are connected by a single connection (**Figure 1b**). This is in contrast with the reality that in the brain two pyramidal cells are connected by synapses on multiple dendrites (**Figure 1c**) (Hiratani and Fukai, 2018). These dendritic routes can be independently gated by dendrite-inhibiting somatostatin (SST) interneurons (Kepecs and Fishell, 2014), which are controlled by a hub area, e.g., prefrontal cortex (PFC) or mediodorsal thalamus (MD) (Oldham and Fornito, 2019), through long-distance corticocortical or thalamocortical connections to vasoactive intestinal peptide (VIP) interneurons that inhibit SST interneurons (Wall et al., 2016; Pfeffer et al., 2013).

This neural substance provides a natural basis for implementing EA. EA is an optimization algorithm based on the evolution of a population of agents (Mirjalili, 2019; Gao et al., 2019) (**Figure 1a**). Each generation of agents are procreated by the parental agents (i.e., the agents with high fitness, or in other words, better task performance) in the previous generation through mutation or cross-over; elite agents (i.e., the agents with the highest fitness) become agents in the next generation without mutation, a technique called elitism.

The implementation of EA in brain learning is illustrated in **Figure 1d, e**, where we suppose that the dendrites of pyramidal cells in a network $\mathcal{N}$ representing hippocampus or neocortex are gated by a network $\mathcal{A}$ representing PFC or MD. If neuron k in network $\mathcal{A}$ is active, all the dendrites labeled k in network $\mathcal{N}$ are gated-on, with others gated-off. Neurons in network $\mathcal{A}$ are alternatively activated one at a time, such that dendrites with different labels are gated-on alternatively (**Figure 1e, upper subplots**). With different dendrites gated-on, network $\mathcal{N}$ gives different outputs in response to the same stimulus, leading to different rewards, either feedback from the real world or estimated by the brain (Stott and Redish, 2015). These reward signals stimulate different levels of neuromodulators, such as dopamine and serotonin, informed dispersedly to the whole network $\mathcal{N}$ to guide synaptic plasticity (Roelfsema and Holtmaat, 2018; Fischer and Ullsperger, 2017; Crispino et al., 2020) (**Figure 1d**). The sub-network of network $\mathcal{N}$ connected by the dendrites labeled k is regarded as the kth agent in EA. Dopamine level stimulated by an agent indicates the fitness of that agents, such that high (or low) dopamine level indicates parental (or non-parental) agents (**Figure 1e, lower subplots**).

In the next subsection, we will argue that the heterosynaptic plasticity between pyramidal cells is equivalent to a learning rule in which the synapse $w_{ij}^{\text{non-parental}}$ between the ith and jth pyramidal cells on a non-parental dendrite becomes approximately the synapse w_{ij}^{parental} on a parental dendrite (**Figure 1f**). In this way, non-parental agents are updated to be descendants of the parental agent with some mutations (represented by ϵ in **Figure 1f**) caused by the stochastic nature of the nervous system (Mongillo et al., 2017; Faisal et al., 2008). There is no cross-over in this EA. All the synapses to and from inhibitory interneurons are kept fixed in our model (**Figure 1c**).



(f)

$$W_{i,j}^{\text{non-parental}} = W_{i,j}^{\text{parental}} + \varepsilon$$

Figure 1: The network architecture. (a) Schematic of the evolutionary algorithm. Each generation of agents (squares) are procreated by multiple parental agents (yellow squares) in the previous generation with high fitness (black arrows and brackets). The one with the highest fitness (red circle) is the elite and becomes an agent of the next generation without any mutation (blue arrows and circles), a technique called elitism. (**b**) Traditional models suppose that two neurons are connected by a single connection. (**c**) Our model supposes that two pyramidal cells are connected through multiple dendritic routes, which are gated by MD or PFC (network $\mathcal{A}$) through SST and VIP interneurons. Only synapses (red) between pyramidal cells are plastic in our model, all the other synapses to and from PV, SST and VIP interneurons remain fixed. (**d**) Dendritic routes labeled k ($k = 1, 2, 3$) in network $\mathcal{N}$ are gated by neurons labeled k in network $\mathcal{A}$. Dopamine, which represents the evaluation of the output of network $\mathcal{N}$, are released dispersedly to network $\mathcal{N}$. (**e**) Neurons in network $\mathcal{A}$ are activated alternately one at a time. At time t_k ($k = 1, 2, 3$), only neuron k in network $\mathcal{A}$ is activated (white square) with others slient (black squares), so that the dendritic routes labeled k in network $\mathcal{N}$ are gated on (white circles) with others gated off (black circles). The sub-network connected by all the kth dendritic routes in network $\mathcal{N}$ is regarded as the kth agent in evolutionary algorithm (lower subplots). The 3rd agent stimulates a high level of dopamine (red text in the upper-right subplot), is therefore the parental agent; the other two agents stimulate a low level of dopamine, are therefore non-parental agents. (**f**) The plasticity rule. The synaptic efficacy $w_{i,j}^{\text{non-parental}}$ between the ith and jth pyramidal cells in a non-parental agent becomes the synaptic efficacy between the same pair of pyramidal cells in a parental agent plus a small value ϵ.

Basic learning process

We consider a delayed reward process widely in nature or in the laboratory: neuronal activities result in eligibility traces in synapses, according to which sparse and delayed rewards then guide synaptic plasticity. The above plasticity rule (**Figure 1f**) can be realized by heterosynaptic plasticity through the following cellular-level mechanisms:

(1) If the synapse on a gated-on dendrite has large efficacy (e.g., synapse c_1 from neuron 1 in **Figure 2b**), then a pre-synaptic spike has high probability to stimulate a post-synaptic spike. This pairing of pre- and post-synaptic spikes produces diffusive messengers such as nitric oxide (NO) (Schuman and Madison, 1991, 1994), which produces potentation eligibility traces in neighboring synapses (synapses a_1 and b_1 in **Figure 2b**) on gated-off dendrites in the condition of the simultaneous arriving spikes from neuron 1 (Bonhoeffer et al., 1989; Schuman and Madison, 1994; Kossel et al., 1990). This heterosynaptic potentiation has spatial scope comparable with the dendritic arbor of a pyramidal cell ($\sim 150\mu$m in hippocampal CA1 (Bonhoeffer et al., 1989; Schuman and Madison, 1994)). Synapses from neuron 2 in **Figure 2b** (i.e., synapses a_2, b_2 and c_2) cannot be potentiated by NO without simultaneous presynaptic spikes.

(2) If the synapse on a gated-on dendrite has low efficacy (e.g., synapse c_2 in **Figure 2c**), then a pre-synaptic spike has low probability to stimulate a post-synaptic spike, such that the potentiation eligibility traces are low in the neighboring synapses sharing the same pre-synaptic neurons (i.e., synapses a_2 and b_2).

(3) Pre-synaptic release of glutamate stimulates astrocytes to spread calcium wave (Cornell-Bell et al., 1990; Deitmer et al., 1998), which induces depression eligibility traces in synapses spatially widespread (Scanziani et al., 1996; Min and Nevian, 2012; Chen et al., 2013) (at least 300$\sim$500μm in hippocampal CA1 (Andersson et al., 2007; Chen et al., 2013)).

Terms 1 and 3 above are best established in the hippocampus (Chistiakova et al., 2014) (also see the references above), and have also been observed in the neocortex (Chistiakova and Volgushev, 2009; Kossel et al., 1990; Min and Nevian, 2012; Lee et al., 2012; Tsumoto and Suda, 1979).

Collectively, if a synapse on a gated-on dendrite has high (or low) efficacy, then it induces high (or low) potentiation eligibility traces in the neighboring synapses from the same pre-synaptic neuron on gated-off dendrites; the depression eligibility traces on different synapses are similar (**Figure 2d**). If the delayed reward is small so that the dopamine level is low (i.e., the gated-on dendrite is non-parental in EA, see **Figure 1e**), no synaptic change happens; if the delayed reward is large so that the dopamine level is high (i.e., the gated-on dendrite is parental in **Figure 1e**), the synapses with large (or small) potentiation eligibility traces will be potentiated (or depressed) (Roelfsema and Holtmaat, 2018) (**Figure 2e**). As biological synapses take binary efficacies (O'Connor et al., 2005), potentiation (or depression) only works in synapses with low (or high) efficacies, whereas those already with high (or low) efficacies remain unchanged (**Figure 2e**). As a result, synapses on non-parental dendrites are updated towards those on parental dendrites connecting the same pair of pre- and post-synaptic neurons (i.e., synapses a_1 and b_1 are updated towards c_1, whereas a_2 and b_2 towards c_2), whereas those on parental dendrites (i.e., synapses c_1 and c_2) remain unchanged (**Figure 2e**), fulfilling the requirement of EA (**Figure 1f**).

Generally speaking, the above argument still holds if we regard each neuron in the model as a population of closely neighboring pyramidal cells. First, closely neighboring pyramidal cells have high firing synchrony (Takahashi and Sakurai, 2009), so the synchronous pre-synaptic firing required for potentiation still holds if neuron 1 in **Figure 2b** represents a population of closely neighboring neurons. Second, both the heterosynaptic potentiation and depression mechanisms (**Figure 2b, c**) are mediated by inter-cellular pathways, therefore still hold for synapses on different post-synaptic neurons spatially nearby (Kossel et al., 1990; Bonhoeffer et al., 1989; Chen et al., 2013). Third, the synapses from a population of neighboring pyramidal cells with high synchrony spatially cluster on the same compartment of a post-synaptic dendrite (Takahashi et al., 2012) with correlated synaptic efficacies (Lu and Zuo, 2017; Hayama et al., 2013; Engert and Bonhoeffer, 1997), therefore can be considered as a big synapse taking binary values.

Multiple parents and elitism

In EA of computer science, each generation usually has multiple parents and at least one elite (Mirjalili, 2019; Gao et al., 2019) (**Figure 1a**). The plasticity process illustrated in **Figure 2**, however, corresponds to the case that a generation has only one parent, because the agents at any time step are the descendants of a single parent (**Figure 3a**). Here we show that both multiple parents and elitism can be implemented

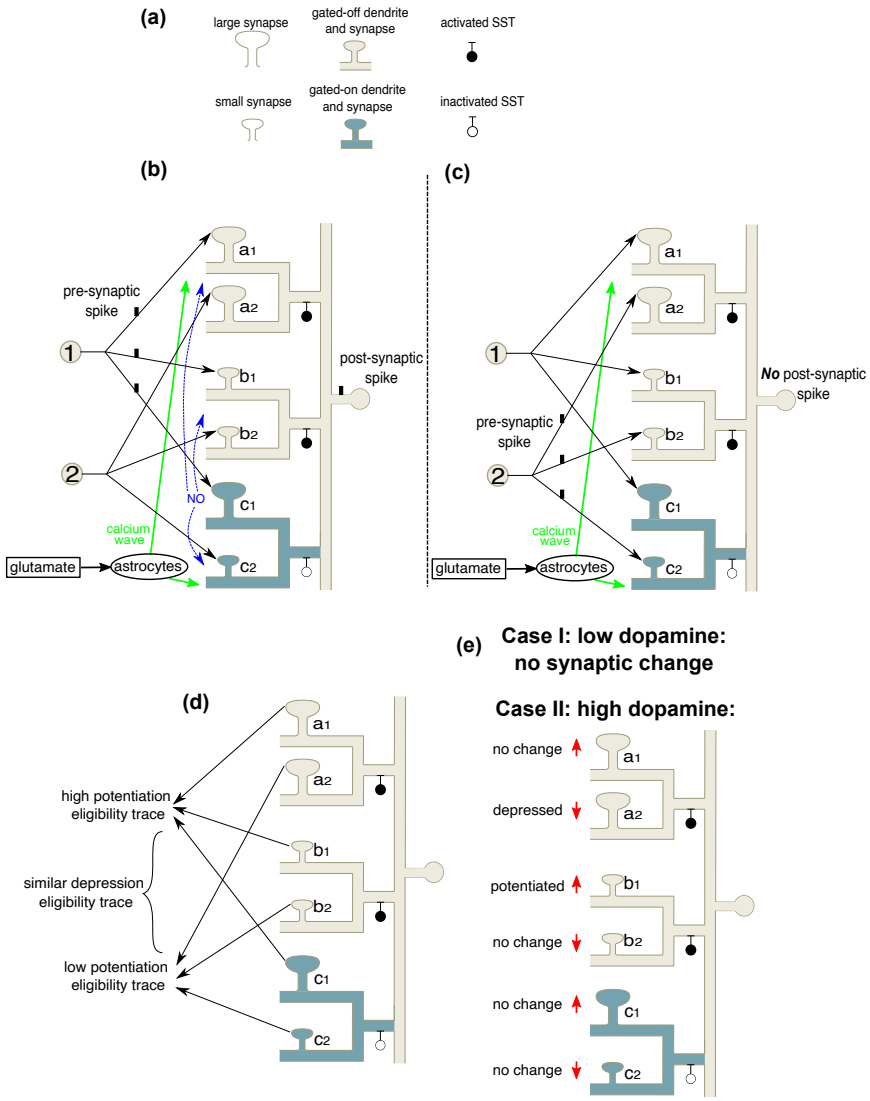


Figure 2: Heterosynaptic plasticity induces EA. (**a**) Elements of the schematics below. (**b**) When neuron 1 emits a spike, the large synapse c_1 on the gated-on dendrite stimulates a post-synaptic spike; this pairing of pre- and post-synaptic spikes produces diffusive NO, which, together with the spikes simultaneously invading synapses a_1 and b_1, induces potentiation eligibility traces in these synapses. (**c**) When neuron 2 emits a spike, the small synapse c_2 on the gated-on dendrite does not stimulate a post-synaptic spike, therefore no potentiation eligibility traces in synapses a_2 and b_2. In both panels **b** and **c**, pre-synaptic release of glutamate stimulates astrocytes to produce widely spreading calcium waves, which induce depression eligibility traces in all synapses. (**d**) The synapses targeted by neuron 1 have higher potentiation eligibility traces than those targeted by neuron 2, but the depression eligibility traces in these synapses are similar. (**e**) With a low dopamine level, no synaptic change happens. With a high dopamine level, the synapses a_1, b_2, c_1 (or a_2, b_2, c_2) with high (or low) potentiation eligibility traces tend to be potentiated (or depressed), see upper (or lower) red arrows. Synapses a_1, c_1 (or b_2, c_2) remain unchanged because they already have large (or small) efficacy.

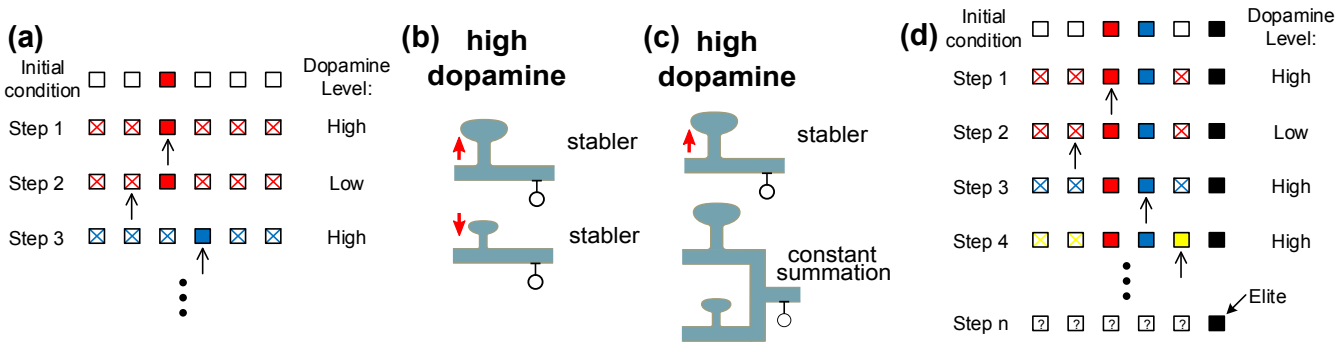


Figure 3: Multiple parents and elitism. (**a**) The learning process under the plasticity rule illustrated in **Figure 2**. The initial condition contains one parental agent (red square) that can induce high dopamine level and five non-parental agents (white squares). In step 1, the red parental agents are activated (black arrow), so that all non-parental agents are updated towards the red agent, becoming its descendants (red crosses). In step 2, the activated agent cannot induce high dopamine level, so this agent remains non-parental. In step 3, the activated agent induces high dopamine level, so all the other agents (including the red parental agent) are updated toward the activated agent, becoming its descendants (blue crosses). (**b**) The synaptic stabilization mechanism. If a large (upper panel) or small (lower panel) synapse on a gated-on dendrite (white circle) receives potentiation or depression signal (upward or downward arrow) at a high dopamine level, then the synapse is stabilized at large or small efficacy. (**c**) The stabilization of small synapses can also be realized by the stabilization of large synapses (upper panel) together with dendrite-level synaptic homeostasis (lower panel). (**d**) How multiple parents and elitism are fulfilled using the stabilization mechanism in panels **b** and **c**. Notice that the parental agents (red, blue, yellow and black squares) are stabilized, so that in steps 1, 3 and 4, these parental agents do no update toward the activated agent that stimulates high dopamine level. After many steps (last row), parental agents may also lose their synaptic configurations (the question marks mean we do not know the states of those agents); but the elite agent (black square), which can induce the highest dopamine level, is the stablest one, and can last for the longest time.

in our model by supposing a meta-plasticity mechanism. In this mechanism, if a synapse with large (or small) efficacy on a gate-on dendrite receives potentiation (or depression) signals under high dopamine level, then the efficacy of this synapse gets stabler at large (or small) level, harder to be switched in future learning process (**Figure 3b**). With this mechanism, if agent k is parental, then the synapses on the dendrites with label k get stabilized, so that if another agent m induces a high dopamine level afterwards (i.e., agent m is also parental), the synapses with label k remain unchanged, instead of updated towards the synapses with label m. Therefore, if agent k is activated again later, high dopamine level may still be induced, so that non-parental agents (whose synapses are not stabilized) can still be updated toward agent k, becoming descendants of agent k (**Figure 3d**). In this way, the agents at any time step are the descendants of multiple parents. Despite of the stability, agent k may also gradually lose its synaptic configuration after sufficiently long time. The parental agent that can induce the highest dopamine level (i.e., the elite) also has the highest stability, and therefore can be maintained for the longest time, fulfilling the elitism technique in EA (**Figure 3d**).

In a phenomenological model (Fusi et al., 2005; Benna and Fusi, 2016), a binary synapse takes small or large efficacy if a hidden state in that synapse is below or above a threshold value. According to that understanding, the hidden state of a synapse of an activated parental agent gets increased (or decreased) if that synapse already has high (or low) efficacy. Physiologically, it is well known that memory formation, which implies stabilized synaptic changes, requires inhibiting SST or parvalbumin (PV) interneurons to increase the excitability of pyramidal cells (Holtmaat and Caroni, 2016; Josselyn and Tonegawa, 2020), and that dopamine stabilizes potentiated synapses (Duszkiewicz et al., 2019), so the potentiation part of the stabilization mechanism (upper subplot of **Figure 3b**) is well established. The depression part (lower subplot of **Figure 3b**) may be related to a heterosynaptic meta-plasticity process: the stimulation of a pre-synaptic pathway to a hippocampal region makes long-term potentiation harder to be induced in another pathway in subsequent experiment (Abraham et al., 2001; Hulme et al., 2012), probably due to an extra-cellular mechanism related to astrocytes (Hulme et al., 2013) similar to heterosynaptic depression (**Figure 2b, c**). However, the modulation of this meta-plasticity by SST interneurons and dopamine level remains to be experimentally tested.

A better established process to fulfill the depression stabilization (lower subplot of **Figure 3b**) is dendrite-level synaptic homeostasis (Sutton et al., 2006; Barnes et al., 2017), which constrains the total synaptic efficacy on a dendrite (**Figure 3c**). With this constraint, small-efficacy synapses are hard to be potentiated if other large-efficacy synapses on the same dendrite are stabilized, fulfilling the stabilization of small-efficacy synapses. If we regard a single neuron model in **Figures 1 and 2** as a population of closely neighboring biological pyramidal cells (as we discussed in the previous subsection), and information route is gated at the neuron level (instead of the dendrite level), then neuron-level synaptic homeostasis (Royer and Paré, 2003; White et al., 1990; Turrigiano, 2011) also fulfills this task.

EA is compatible with off-line replay

According to the no-free-lunch theorem (Wolpert and Macready, 1997), no single learning algorithm is advantageous in every problem. Therefore, it is necessary for the brain to combine a variety of learning strategies for the best living of the animal; and it is necessary for us to understand the scenarios where EA is suitable, thereby getting insight into the functional roles which EA plays in.

Consider the following on-line learning scenario which possibly happens in a waking animal (Scenario 1, **Figure 4a**): in a learning session, the performance of a gated-on dendritic route in response to an input stimulus is scored by the reward feedback from the real world, inducing corresponding dopamine level; learning sessions may not be immediately successive, but are separated by intervals of variable durations. This scenario is not compatible with EA. To see this, suppose a situation where dendritic route 1 brings 10 scores of reward in response to stimulus 1, and brings 0 scores in response to stimulus 2; dendritic route 2 brings 6 scores of reward in response to both stimuli 1 and 2 (**Figure 4c**). If we suppose that high dopamine level is induced only in sessions corresponding to the highest reward, then only the sessions in which dendritic route 1 is gated on and stimulus 1 is inputted induce high dopamine level, which makes the synapses on dendritic route 2 updated towards those on dendritic route 1 (**Figures 1f and 2**). In other words, dendritic route 1 is the parental route in the on-line learning scenario of **Figure 4a**. However, this parental selection is incorrect because it is dendritic route 2 that brings more rewards on average over the two input stimuli (**Figure 4c**).

This problem can be solved if we suppose that the evaluation of dendritic routes happens during off-line replay (Scenario 2, **Figure 4b**): a gated-on dendritic route is quickly evaluated over a number of replayed inputs, and the dopamine level is released according to the average reward (estimated by the brain) associated with a number of input stimuli just replayed (**Figure 4b**). In this case, the dopamine level indicates the average performance of the gated-on dendritic route over a batch of input samples just replayed. If we choose the batch size to be 2, then dendritic route 2 in **Figure 4c** can be selected to be parental during the replay. This off-line replay scenario (**Figure 4b**) is a prediction of our model to be experimentally tested.

The incompatibility of EA with Scenario 1 lies in the very nature of EA. In gradient-descent algorithms, the updating direction of a synaptic weight is computable and equal to the average over the directions computed from individual input stimuli. Therefore, the neural network can be optimized by accumulating small updating steps in response to individual stimuli during the on-line learning in Scenario 1. However, in EA, such updating direction is uncomputable, and synaptic updating relies on the comparison of different dendritic routes. This comparison should depend on the performance of dendritic routes over a number of stimuli samples (as an estimation of the performance over the whole stimuli set) instead of a single stimulus, which makes Scenario 1 inappropriate. The compatibility of EA with off-line replay suggests that EA plays an important role in transforming episodic memory into unconscious dexterous skills at rest or in sleep (Buch et al., 2021; Rasch and Born, 2013). In our following simulations, EA is performed in an off-line manner.

Numerical simulation

To illustrate the EA above, we trained a spiking neural network to perform XOR task. In this network (**Figure 5a**), a population of pyramidal cells receive through dendrites gated by SST interneurons from two neuronal groups emitting Poisson spike trains, and give output to a single neuron. The two input groups also impose global inhibition to the pyramidal cells through PV interneurons. The network was trained so that the output neuron has high firing rate when one input group has high firing rate while the other is silent, and the output neuron has low firing rate when both input groups have high firing rates (**Figure 5c**). During training, the synapses from the two input groups to the pyramidal cells were changed according to the heterosynaptic rules, and the synapses to and from PV interneurons as well as the output synapses of the pyramidal cells were kept fixed.

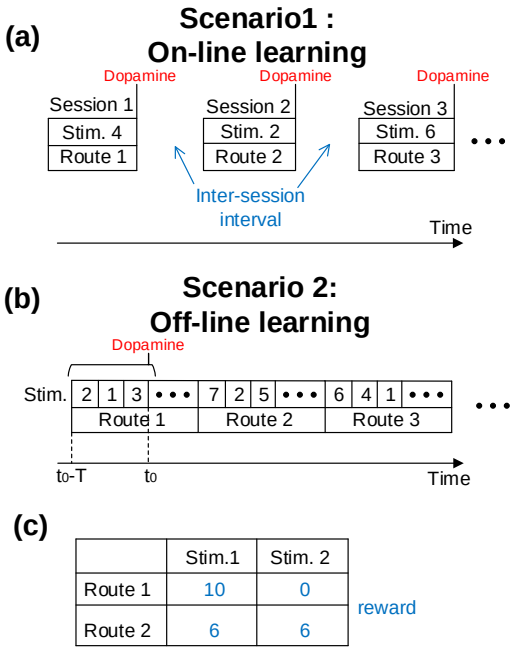


Figure 4: EA is compatible with off-line replay. (**a**) During an on-line learning scenario, the performance of a gated-on dendritic route in response to an input stimulus is represented by the level of dopamine given at the end of each session. There may be intervals of variable durations between sessions (blue text). (**b**) During an off-line learning scenario, the brain can quickly sweep over samples of input stimuli when a dendritic route k ($k = 1, 2, 3, \cdots$) is gated on. The dopamine level at time t_0 represents the average reward associated with the replayed stimuli from $t_0 - T$ to t_0, with T being a period in which several (3, in this schematic) stimuli are replayed. (**c**) An example situation considered in the text. Route 1 brings the highest reward (blue text) in response to a single stimulus, but route 2 brings the highest average reward over both stimuli.

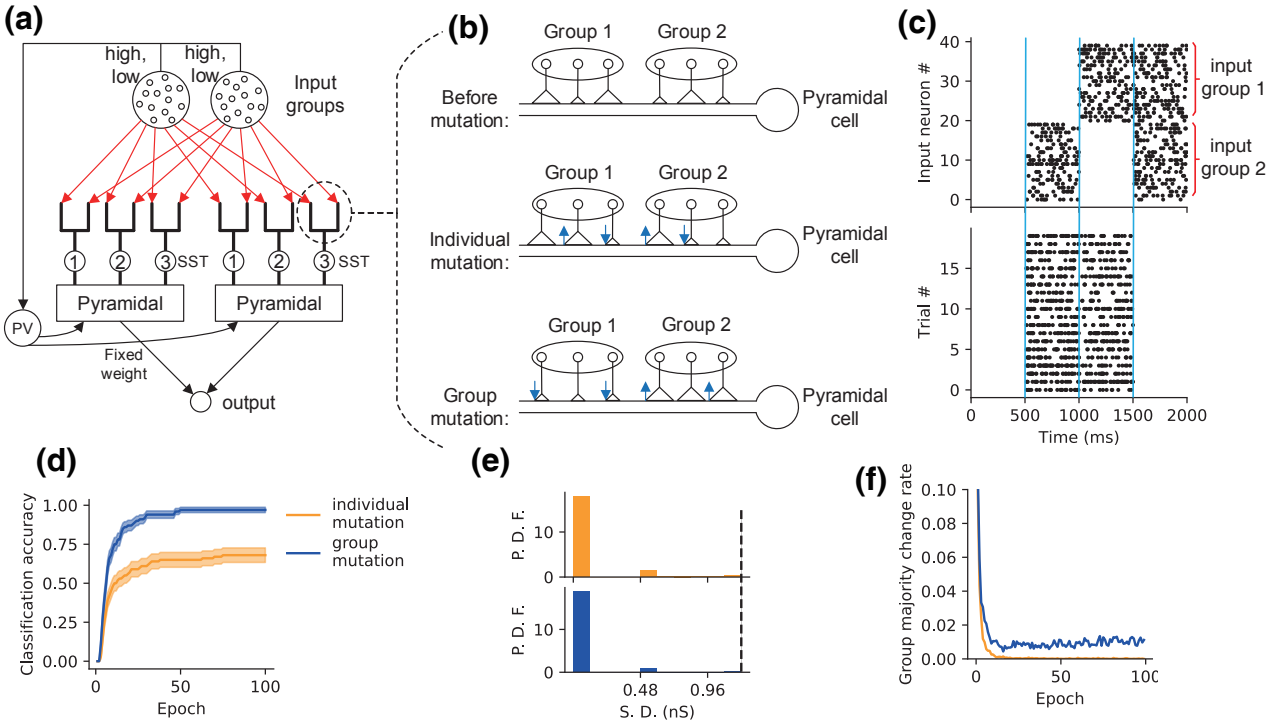


Figure 5: Training neural networks on XOR task. (**a**) The structure of the network to perform XOR task. Pyramidal cells receive from two input groups (whose activity levels can be high or low) through synapses (red arrows) on dendrites gated by SST interneurons. Red synapses are to be changed during training, and the other synapses are fixed. (**b**) The mutation on a single dendrite (dashed circle in panel **a**). Individual mutation: synapses are mutated independently. Group mutation: synapses from the same input group are potentiated or depressed simultaneously. Large (small) triangles represent large (small) synapses. Upward (downward) blue arrows indicate potentiation (depression). (**c**) Raster plot of the input neurons in one simulation trial (upper) and the single output neuron in multiple trials (lower) after training under group mutation. Vertical blue lines delimit four intervals in which the two input groups have different high or low activities. (**d**) Classification accuracy during the training process for individual (orange) and group (blue) mutation. Error belt indicates standard error of mean (s.e.m.) over 100 training trials. (**e**) Probability density function (P. D. F.) of the standard deviation (S. D.) of $\{w_{d,i}\}_{i\in G_1} \cup \{w_{d,i}\}_{i\in G_2}$. Dashed vertical line indicate the S. D. when synaptic efficacies in $\{w_{d,i}\}_{i\in G_1}$ and $\{w_{d,i}\}_{i\in G_2}$ have large or small values with equal probability. (**f**) The probability that the majority value of $\{w_{d,i}\}_{i\in G_1}$ or $\{w_{d,i}\}_{i\in G_2}$changes during the training process. Average over 100 training trials.

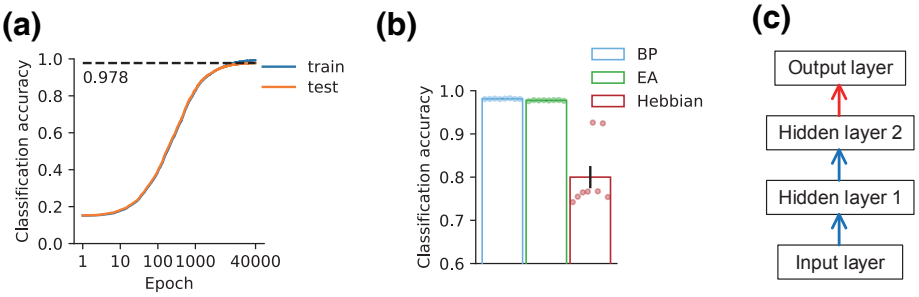


Figure 6: Training neural networks on MNIST classification task. (**a**) Classification accuracy during the training process of EA. Horizontal dashed line with number indicates the final accuracy on the test dataset. (**b**) Classification accuracy on the test dataset when training the deep network using BP, EA and Hebbian rule. (**c**) The network structure. When studying Hebbian algorithm, the lower layers were trained using Hebbian algorithm (blue arrows), and the top layer was trained by gradient-descent algorithm supervisedly (red arrows). In panel **b,** error bars represent s.e.m. over 8 training trials. Error bars in panel **a** (not shown) are comparable with the width of the lines.

We studied two types of mutation: individual mutation (IM), in which each synapse was randomly selected to flip its efficacy; and group mutation (GM), in which the synapses from the same input group were simultaneously potentiated or depressed (**Figure 5b**). GM is a better mutation strategy than IM, because GM enables the network to quickly achieve high classification accuracy whereas IM saturates the network at a low performance (**Figure 5d**).

To understand the reason for the good performance of GM, we first noticed that in both mutation strategies, $\{w_{d,i}\}_{i\in G_1}$ or $\{w_{d,i}\}_{i\in G_2}$ (i.e., the set of synapses from all the neurons i in group G_1 or G_2 through the same dendrite d) either mostly have large efficacy or mostly have small efficacy (**Figure 5e**). This majority efficacy in $\{w_{d,i}\}_{i\in G_1}$ or $\{w_{d,i}\}_{i\in G_2}$ may change occasionally under GM, but hardly under IM (**Figure 5f**). To understand this point, suppose all synapses in $\{w_{d,i}\}_{i\in G_1}$ except $w_{d,1}$ have large efficacy. Due to the high synchrony between the neurons in the same group G_1 (**Figure 5c, upper panel**), a pre-synaptic spike invading $w_{d,1}$ has high chance to pair with a post-synaptic spike stimulated by the activation of the other large-efficacy synapses in $\{w_{d,i}\}_{i\in G_1}$, thereby potentiating $w_{d,1}$. Similarly, if all synapses in $\{w_{d,i}\}_{i\in G_1}$ except $w_{d,1}$ have low efficacy, then without the help of the other synapses in the same group, it is hard for $w_{d,1}$ alone to stimulate post-synaptic spikes, so $w_{d,1}$ is very likely to be depressed. This makes $\{w_{d,i}\}_{i\in G_1}$ or $\{w_{d,i}\}_{i\in G_2}$ stabilized at the majority efficacy. GM, however, mutates $\{w_{d,i}\}_{i\in G_1}$ at the same time, breaking the above shackle on the exploration of weight configurations (**Figure 5f, blue curve**). Consistent with GM, it has been experimentally observed that functionally similar presynaptic neurons tend to form synapses in the same dendritic compartment, and these spatially clustering synapses tend to be simultaneously potentiated or depressed (Larkum and Nevian, 2008; Ujfalussy and Makara, 2020; Lu and Zuo, 2017; Engert and Bonhoeffer, 1997; Hayama et al., 2013).

To manifest the effectiveness of the biologically supported EA in training neural networks in more complicated tasks, we then trained a feedforward neural network with two hidden layers (**Figure 6c**) to classify MNIST images. In this network, each neuron represents the mean firing rate of a neuronal group, each synaptic weight takes binary values, one positive, one negative, modeling the effective synaptic weight when the excitatory synapse has large or small efficacy under the global inhibition of PV interneurons (**Figure 5a**). EA was performed so that descendants were mutated from parental agents with a small probability of synaptic flip (**Figure 1f,** see Methods). The final classification performance is comparable with continuous-weight networks with the same structure trained by back propagation (BP), see **Figure 6a, b**. When the summation of synaptic weights on the same dendrite is fixed during training (**Figure 3c, lower panel**), the performance is hardly influenced (**Supplementary Figure 1b**). We also trained continuous-weight networks layer-by-layer using a competitive Hebbian algorithm (Krotov and Hopfield, 2019), except that the last layer was trained by gradient-descent algorithm supervisedly (**Figure 6c**), and found worse final performance (**Figure 6b, third bar**). Interestingly, if we supervisedly trained the last layer on top of the first hidden layer, then the performance was better than the two-hidden-layer case (**Supplementary Figure 1c, d**), similar to the finding in Amato et al. (2019). This phenomenon implies that Hebbian algorithm lacks the ability to coordinate weights in different layers to fulfill better performance.

The capability of the biologically supported EA in training deep neural networks is further demonstrated in **Table 1**, where EA is shown to train networks with binary weights to play Atari games up to performance comparable to continuous-weight networks trained by gradient-based methods such as DQN and A3C (Mnih et al., 2015, 2016). Such good performance of EA has also been observed in training

Game name	DQN, continuous weight, (Such et al., 2017)	A3C, continuous weight, (Such et al., 2017)	EA, continuous weight, (Such et al., 2017)	EA, binary weight, ours	Hebbian-Q, continuous weight, ours
amidar	978	264	263	251	10
assault	4,280	5,475	714	873	130
asterix	4,359	22,140	1,850	1,923	92
asteroids	1,365	4,475	1,661	**1,596**	146
atlantis	279,987	911,091	76,273	74,563	1,400
enduro	729	-82	60	**32**	5
frostbite	797	191	4,536	**4,520**	30
gravitar	473	304	476	**589**	26
kangaroo	7,259	94	3,790	**3,997**	151
seaquest	5,861	2,355	798	800	85
skiing	-13,062	-10,911	-6,502	**-6,503**	-28,383
venture	163	23	969	**1,070**	22
zaxxon	5,363	24,622	6,180	**6,195**	24

Table 1: Binary-weight networks trained by EA is competitive with continuous-weight networks trained by DQN and A3C in Atari games. In the fifth column, bold black numbers indicate the scores of EA-trained binary-weight network that are higher than both the scores of DQN and A3C. Red (or blue) numbers indicate the scores that are higher than those of DQN (or A3C), but lower than A3C (or DQN).

continuous-weight networks (Such et al., 2017). We also tried a Hebbian-Q algorithm, where low-level layers were trained by a competitive Hebbian algorithm (Krotov and Hopfield, 2019), whereas the top layer was trained by Q learning (Mnih et al., 2015), and found that the resulted performance was significantly worse than DQN, A3C and EA (**Table 1**). These results demonstrate that the biological supported EA is a powerful algorithm capable to train deep neural networks on complicated tasks.

Discussion

Overall, we collect a broad spectrum of biologically experimental evidence into a coherent picture, and strongly suggest that the heterosynaptic plasticity between hippocampal or cortical pyramidal cells mediated by diffusive NO and astrocyte calcium signaling implies a powerful general-purpose EA in the brain. Comparing to the homosynaptic mechanisms reviewed in Lillicrap et al. (2020), our mechanism also has the following computational advantages:

(1) EA is model-free, it does not depend on knowing anything about the task objective or neural network structure, thereby broadly applicable without the need of much gene-encoded knowledge. For example, a recent proposal states that each synapse has an evolution-endowed specific Hebbian learning rule coded in genes (Miconi et al., 2020; Najarro and Risi, 2020). Considering the 100 trillion synapses in the brain, this proposal requires infeasibly huge information coded in DNA. As a second example, the node perturbation approach estimates the gradient of reward respecting to synaptic weights by observing the correlation between the injected noises to neurons and the rewards (Fiete and Seung, 2006; Miconi, 2017; Legenstein et al., 2010). This approach requires the linear dependence of membrane conductance on the synaptic weights and injected noises (Fiete and Seung, 2006), which may be problematic in biological neurons with nonlinear dendrites; besides, this approach has never successfully trained large deep networks for difficult problems (Lillicrap et al., 2020), possibly due to its sensitivity to the inaccuracy of mean subtractions of the covariant reward and neuronal activities (Loewenstein, 2008). As a third example, the target propagation approach requires to reconstruct the activities of lower layers from those of higher layers (LeCun, 1986; Lee et al., 2015). This top-down reconstruction is consistent with predictive-coding theory of perception (Rao and Ballard, 1999), which receives support from visual and auditory systems (Friston, 2018; Heilbron and Chait, 2018). However, when training recurrent neural networks on temporal tasks, this approach requires a neuron to remember its whole temporal trajectory to guide the updating of its synapses, which is biologically problematic (Manchev and Spratling, 2020). On the contrary, our EA does not need heavy coding in DNA, and requires no specific network structure to be applicable.

(2) EA is particularly good at creatively dealing with multi-objective optimization problems (Zhou et al., 2019). Real-world problems usually have to address several conflicting objectives. For example, because a cheap house in the downtown is usually unavailable, a house buyer may consider both an expensive house in the downtown and a cheap house in the suburb, due to their own merits to be either convenient or cheap. Gradient-based methods can solve this problem using a single objective as the linear combination of these multiple objectives with a pre-assigned preferential weghts: e.g., a house buyer may emphasize more on price or convenience before making the decision. In contrast, a population of agents trained by EA are able to simultaneously enumerate a number of possible solutions each with its unique merit (more precisely, the Pareto front (Zhou et al., 2019)), thereby significantly improving the creativity in front of conflicting objectives. In neuro-imaging studies, it has been found that the functional connectivity of the brain is dynamically reconfigured, and this reconfiguration is closely related to the creativity of human subjects (Braun et al., 2015; Li et al., 2017). From our model, this dynamic reconfiguration is caused by the alternation of activated agents (**Figure 1e**), reflecting the solution enumeration process mentioned above.

Comparing with gradient-based methods, the drawback of EA is its slow learning speed. Classification of MNIST images requires over 10,000 learning epochs using EA (**Figure 6a**), instead of dozens or hundreds of epochs using gradient-based methods. However, learning speed may not be a critical aspect, because EA mainly mediates the off-line learning during resting or sleeping (**Figure 4**), whose neuronal firing sequences are temporally compressed by about 20-fold relative to those in real experience (Buch et al., 2021). Besides, the brain may use several strategies to speed up learning. First is the combination of homosynaptic mechanisms compatible with gradient-based methods (Lillicrap et al., 2020). To manifest the sole power of heterosynaptic plasticity, we do not add any Hebbian-type homosynaptic rules in our model. Adding Hebbian-type homosynaptic rules can guide the mutation direction using hints of gradients, potentially speeding up learning. Second, a recent theoretical study (Qian et al., 2020) suggests that comparing with EA with only mutation, EA with crossover operation is faster by polynomial order of the population size. Our model only contains mutation, but can easily embed in crossover. Suppose network $\mathcal{N}$ (**Figure 1d**) has two modules $\mathcal{N}_A$ and $\mathcal{N}_B$, whose dendritic routes are separately controlled by two networks $\mathcal{A}$ and $\mathcal{B}$, such that when a neuron $a_i \in A$ (or $b_j \in B$) is activated, the dendritic routes labeled by i (or j) in $\mathcal{N}_A$ (or $\mathcal{N}_B$) are gated on, with other routes gated off. If at the early stage of training, a_1 and b_1 (or a_2 and b_2) are activated simultaneously, such that dendritic routes with the same label 1 (or 2) in both $\mathcal{N}_A$ and $\mathcal{N}_B$ are gated on simultaneously; but at a later stage, a_1 and b_2 (or a_2 and b_1) are activated simultaneously, such that dendritic routes with different labels in $\mathcal{N}_A$ and $\mathcal{N}_B$ are gated on simultaneously: then this later stage fulfills crossover operation.

In our model, only excitatory synapses between pyramidal cells are plastic, whereas the synapses to and from inhibitory interneurons are kept fixed. Physiologically, inhibitory plasticity surely plays an important role in normal functioning of the brain (Hennequin et al., 2017; Ostroumov and Dani, 2018). Inhibitory plasticity has more diverse forms than excitatory plasticity, so understanding its functional roles is challenging (Hennequin et al., 2017). An emerging principle is that inhibitory plasticity restores the balance of excitation and inhibition up to dendrite-specific level, in response to the imbalance caused by excitatory plasticity or environmental change (Herstel and Wierenga, 2021; Hennequin et al., 2017). For example, the activation of NMDA receptors potentiates the dendrite-targeting synapses from SST interneurons (Chiu et al., 2018), which may help SST interneurons better gate off the dendrites frequently or strongly activated by excitatory inputs.

Overall, our work provides new computational aspects to understand heterosynaptic plasticity in the brain, and suggests neuroscientists further investigate the biological backend of EA.

Methods

XOR task

In the network of **Figure 5a**, there are 20 pyramidal cells that receive from two groups of excitatory neurons (each group have 20 neurons) through 20 dendrites gated by different SST interneurons. These pyramidal cells are inhibited by PV interneurons, which receive input from both groups of excitatory input neurons and another group of inhibitory neurons (not depicted in **Figure 5a**). Pyramidal cells also give output to a single neuron. Pyramidal cells, PV interneurons and the output neuron are modeled with integrate-and-fire equations:

$$C\frac{\mathrm{d}V_i}{\mathrm{d}t} = -g_L(V_i - V_L) - \sum_j g_{ij}(V_i - V_{ij}) + I_{background}, \tag{1}$$

where $V_L = -74$mV, $g_L = 25$nS, $C = 500$pF, and $I_{background} = 425$pA. When the membrane potential V_i reaches the firing threshold $V_\theta = 54$mV, a spike is recorded and V_i is reset to $V_{reset} = -60$mV . The reversal potential V_{ij} is 0mV if the synapse from j to i is excitatory, or -70mV if this synapse is inhibitory. The dynamics of the synaptic conductance g_{ij} is

$$\tau_{ij}\frac{\mathrm{d}g_{ij}}{\mathrm{d}t} = -g_{ij} + w_{ij}\sum_l \delta(t - t_{l;j}), \tag{2}$$

with $t_{l;j}$ being the time of the lth spike of the jth neuron. The time constant τ_{ij} is 10ms for the synapses from the PV interneurons to the pyramidal cells and 5ms for the other synapses. The synaptic efficacy w_{ij} is 4.8nS from the excitatory input neurons to the PV interneurons, 2.4nS from the inhibitory input neurons to the PV interneurons, 2.4nS from the PV interneurons to the pyramidal cells, and 4.8nS from the pyramidal cells to the single output neuron: all the above synapses are connected with probability 0.4, and are fixed during training. The synapses from the excitatory input neurons to the pyramidal cells are all-to-all connected in every dendrite of the pyramidal cells, and have either large efficacy of 2.4nS or small efficacy of 0nS. Both the excitatory and inhibitory input neurons produce Poisson spike trains. The firing rate of the inhibitory input neurons is fixed at 40Hz. The firing rates of the two groups of excitatory input neurons are respectively (0Hz, 0Hz), (20Hz, 0Hz), (0Hz, 20Hz), and (20Hz, 20Hz) in the four successive intervals with duration 500ms in each training epoch (**Figure 5c, upper panel**). Our simulation was performed using exponential Euler method with time step of 0.1ms using Brian simulator (Stimberg et al., 2019).

The potentiation eligibility trace of the synapse from a pre-synaptic excitatory input neuron to a post-synaptic pyramidal cell through the dth dendrite is

$$e^{pot}_{pre,post,d} = A(d)\sum_{l,m}\Theta(t_{m;post} - t_{l;pre})e^{-(t_{m;post}-t_{l;pre})/\tau_{pot}}, \tag{3}$$

where $t_{m;post}$ (or $t_{l;pre}$) is the time of the mth (or lth) spike of the post- (or pre-) synaptic neuron, $\tau_{pos} = 20$ms is the time scale of the time window indicating the simultaneity of the pre- and post-synaptic spikes that induces potentiation eligibility trace. $A(d) = 1$ for a synapse on the gated-on dendrite; $A(d) = 0.5$ for a synapse on a gated-off dendrite, modeling the effect of diffusive NO emitted from the gated-on dendrite. Here we suppose that the potentiation effect of the NO emitted from a gated-on dendrite of a pyramidal cell can only influence the dendrites of the same pyramidal cell, and has no influence on the dendrites of other pyramidal cells, based on the experimental observation that the spatial scope of this potentiation effect is comparable with the dendritic arbor of a pyramidal cell (Bonhoeffer et al., 1989; Schuman and Madison, 1994).

Synaptic efficacy is updated according to the eligibility trace and the dopamine level, which tries to satisfy the following rules or constraints, explained below:

$$\begin{cases} w_{ij,d}(T) = 0 \text{ or } w_{max} \\ w_{ij,d}(T) = w_{ij,d}(T-1) + \Theta(e^{pot}_{ij,d}(T-1) - e^{pot}_\theta)B(r(T), \{r(t)\}_{t<T}) \\ \sum_j w_{ij,d}(T) = \sum_j w_{ij,d}(T-1) \\ w_{ij,d}(T) = w_{ij,d}(T-1), \text{ if } h_{ij,d}(T-1) > 0 \text{ and } w_{ij,d}(T-1) = w_{max} \\ h_{ij,d}(T) = h_{ij,d}(T-1) + \Theta(e^{pot}_{ij,d}(T-1) - e^{pot}_\theta)C(r(T), \{r(t)\}_{t<T}), \text{ if } w_{ij,d}(T) = w_{max} \\ h_{ij,d}(T) = h_{ij,d}(T-1) - 1 \\ 0 \le h_{ij,d}(T) \le h_{max} \end{cases} \tag{4}$$

The first equation indicates that the synapse $w_{ij,d}(T)$ from the jth excitatory input neuron to the ith pyramidal cell through the dth dendrite at the Tth epoch has binary efficacies 0 or w_{max}. The second equation indicates that $w_{ij,d}$ is potentiated if $e^{pot}_{ij,d}$ is larger than a threshold value $e^{pot}_\theta = 0.45$ and the factor B controlled by the reward $r(T)$ at the Tth epoch is larger than zero. In our simulation, $B = 1$ if $r(T)$ is not smaller than the second largest reward obtained in the last 20 training epochs, and $B = 0$ otherwise. The reward at an epoch is defined as

$$r = -[o(\text{high}, \text{high})]_+ - [-o(\text{high}, \text{low}) + 10]_+ - [-o(\text{low}, \text{high}) + 10]_+, \tag{5}$$

where $o(\text{high}, \text{low})$ means the number of spikes of the output neuron during the interval of 500ms when Group 1 of the excitatory input

neurons have high firing rate and Group 2 have low firing rate; $o(\text{high}, \text{high})$ and $o(\text{low}, \text{high})$ have similar definitions. The third equation in eqs. 4 indicates dendrite-level synaptic homeostasis (**Figure 3c, lower panel**), which keeps constant the total synaptic efficacy on a dendrite. The fourth equation indicates that if a hidden state $h_{ij,d}$ in the synapse is larger than zero, then a synapse with large efficacy remains unchanged. In other words, a large synapse can only be depressed when its hidden state is zero. This hidden state remains at zero if the synapse has low efficacy. The fifth equation indicates that if the eligibility trace $e^{pot}_{ij,d}$ is larger than e^{pot}_{θ} and the factor C controlled by the reward is larger than zero, then the hidden state in the large synapse is increased. In our simulation, C is 20 if $r(T)$ is no smaller than the largest reward obtained in the last 20 training epochs, C is 10 if $r(T)$ is smaller than the first largest but no smaller than the second largest reward obtained in the last 20 training epochs, $C = 0$ otherwise. The sixth equation indicates that the hidden state in every large synapse is decreased by 1 in each training epoch, so that the synapse gradually becomes less stable. The seventh equation indicates that the hidden state is non-negative and is no larger than a maximum value $h_{max} = 20$.

The second and third equations in eqs. 4 may conflict with each other. In our simulation, when too many synapses had large potentiation eligibility traces and were to be potentiated, we kept synaptic homeostasis (i.e., the third equation) non-violated, and set the synapses with largest potentiation eligibility traces at large efficacy. Specifically, at the Tth epoch when $B > 0$, we collected the synapses either with $w_{ij,d}(T-1) > 0$ and $h_{ij,d}(T-1) = 0$ or with $w_{ij,d}(T-1) = 0$ and $e^{pot}_{ij,d}(T-1) > e^{pot}_{\theta}$ into a set $\mathcal{S}$, sorted the potentiation eligibility traces in the synapses in $\mathcal{S}$ in descending order, and let the first $N_{\text{input}}/2 - N_{\text{stable large}}$ synapses have large efficacy, and let the rest synapses have small efficacy. Here, $N_{\text{input}} = 40$ is the number of synapses on a single dendrite (which is the total number of excitatory neurons in the two input groups), and we constrained the number of synapses with large efficacy on a dendrite fixed at $N_{\text{input}}/2$, modeling dendrite-level synaptic homeostasis (the third equation of eqs. 4). $N_{\text{stable large}}$ is the number of synapses with $w_{ij,d}(T-1) > 0$ and $h_{ij,d}(T-1) > 0$ on the dth dendrite.

At each training epoch, the dendrites with a randomly selected label (**Figure 1e**) were gated-on, with other dendrites gated off. Synaptic mutation was performed at the gated-on dendrite at the begining of each epoch. In individual mutation (**Figure 5b, upper panel**), synapses with large efficacy and zero hidden state were set to small efficacy with probability 0.25 at the gated-on dendrite; at the same time, randomly selected synapses with small efficacy were also set to have large efficacy, to maintain the total number of large synapses on the gated-on dendrite. In group mutation (**Figure 5b, lower panel**), all the synapses (except for those synapses with non-zero hidden states) from a randomly selected input group to the gated-on dendrite of a randomly selected pyramidal cell were set to have small efficacy, at the same time a number of synapses on the gated-on dendrite from another input group to the selected pyramidal cell were set to large efficacy to maintain the total number of large synapses on the gated-on dendrite. In both mutation strategies, a single large synapse with zero hidden state has almost equal probability 0.025 to be mutated (**Supplementary Figure 1a**), so that the performance difference under these two stategies (**Figure 5d-f**) is caused by these strategies themselves, instead of the difference of mutation rate.

MNIST task

In **Figure 6**, we trained a multi-layer perceptron with 2 hidden layers with sizes 256 and 128 respectively to classify MNIST images (with image size $28 \times 28 = 784$) into 10 classes. Adjacent layers in the feedforward network are all-to-all connected. The non-linear activation function is relu.

EA was performed on networks whose synaptic weights took binary values $\{-\sqrt{1/N_{\text{in}}}, \sqrt{1/N_{\text{in}}}\}$, where N_{in} represents input size, which is 784 for weights between the input layer and the first hidden layer, 256 for weights between the first and second hidden layers, and 128 between the second hidden layer and the output layer. We initialized 100 network configurations. At every training epoch, negative log-likelihood loss averaging over all 60000 images in the training dataset was evaluated for every configuration. The elite configuration (which is the one with the smallest loss) remained unchanged, whereas each weight of the other configurations was updated to be (almost always) the corresponding weight of the elite configuration, but with probability 0.00005 to be different from the weight of the elite configuration. Therefore, in this algorithm, there is only one parent in each generation of agents. We also studied the case of dendrite-level synaptic homeostasis (**Figure 3c, lower panel**), when the number of synapses that have large efficacy on a dendrite were kept fixed after the random initialization of synaptic weights. We found adding this homeostasis mechanism hardly influences the classification performance (**Supplementary Figure 1b**).

Back-propagation was performed on networks whose synaptic weights took continuous values. We used Adam optimizer of Pytorch to perform the training. Batch size was 100, learning rate was 0.0001, the number of training epochs was 400.

Hebbian learning was also performed on continuous-weight networks. All the low-level layers (except the top layer) were trained by a competitive Hebbian learning (Krotov and Hopfield, 2019). In every step, the weights were updated by

$$\Delta w_{ij} \propto g_i(2u_j - (\sum_k w_{ik}u_k)w_{ij}), \tag{6}$$

where w_{ij} is the synaptic weights from the jth pre-synaptic neuron to the ith post-synaptic neuron, u_j is the input from the jth pre-synaptic neuron, g_i is 1 for the highest activated post-synaptic neuron, -0.4 for the second highest activated post-synaptic neuron, and 0 for others. Batch normalization was performed on the pre-synaptic input u_j. Layers were trained one after another, such that when a higher-level layer was trained, all the lower-level layers were fixed. Adam optimizer was used to supervisedly train the top layer. Following Krotov and Hopfield (2019), the learning rate of Hebbian learning linearly decreased from the maximal value 0.04 at the first epoch to 0 at the last epoch, and the learning rate of supervised learning was kept at 0.0001. Both Hebbian and supervised training were performed using minibatches of size 100 for 1000 epochs.

Atari task

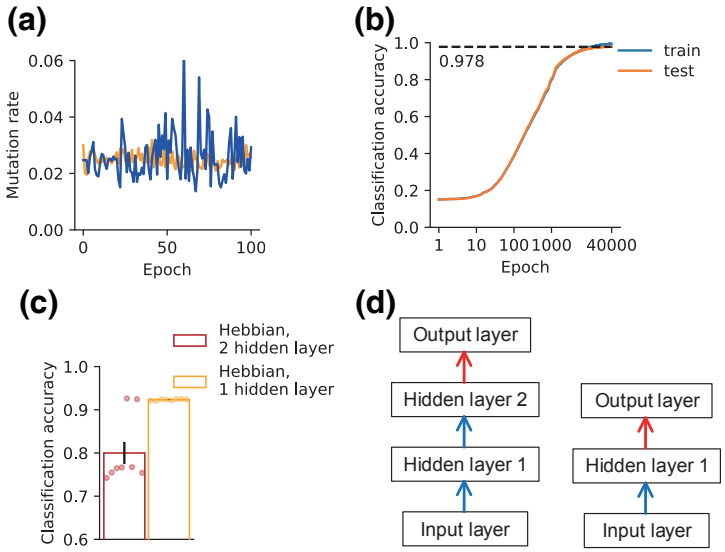
The structure of the deep neural network to play Atari games closely follows that of Mnih et al. (2015); Such et al. (2017). Specifically, the neural network maps 4 recent frames of size 84×84 to actions through 3 convolutional layers and 2 fully-connected layers. Of the three convolutional layers, the kernel sizes are 8×8, 4×4 and 3×3 respectively, the strides are 4, 2 and 1 respectively, and the number of features are 32, 64 and 64 respectively. The first fully-connected layers have 512 neurons. Relu activation function is used.

EA was used to train neural network whose synaptic weights took binary values $\{-\sqrt{2/N_{\text{in}}}, \sqrt{2/N_{\text{in}}}\}$, with N_{in} being the input size. The training protocol closely followed that of Such et al. (2017). Specifically, each generation had 1000 agents, each was evaluated by one episode (i.e., from the start to the end of a game, when the player succeeded or was killed in the game). The top 20 agents were selected to be parents and procreated the next generation by flipping their binary weights with probability 0.002. Then the top 10 agents were further evaluated by 30 episodes, so that the elite agent was selected to be the agent that achieved the highest score. The elite agent became a member of the next generation without any mutation. The performance of EA at any training epoch was the average score of the elite agent in further 200 episodes. We trained the network in 2.5×10^8 epochs. Each number of EA in **Table 1** represents the median value of the scores in the final epochs of 5 training trials.

To perform Hebbian-Q learning, batch normalization was inserted between each layer of the network. The low-level layers were trained using eq. 6, and the last layer was trained supervisedly using Adam optimizer. When training a higher-level layer, all lower-level layers were kept fixed.

When unsupervisedly training a low-level layer, at each time step, a random action was taken of a game (Anand et al., 2019), then minibatches of size 32 were randomly extracted from a buffer of 10^6 recent frames to train the network using eq. 6. To train convolutional layers, we averaged over the weight updating at different spatial locations to update the sharing weights. Similar to Krotov and Hopfield (2019), learning rate linearly decreased from the maximal value 0.02 at the first epoch to 0 at the last epoch. The three convolutional layers were trained in 1.6×10^5 epochs, and the first fully-connected layer was trained in 4.8×10^5 epochs.

When supervisedly training the top layer, minibatches of size 32 were randomly extracted from a buffer of 10^6 recent frames at each epoch in a similar protocol to Mnih et al. (2015). The action policy during training was ϵ-greedy, with ϵ linearly annealed from 1 to 0.1 in the first 10^6 epochs. The target network was cloned from the Q-network every 10000 epochs. Every 2×10^5 epochs, the performance of the agent was evaluated by the average score over 200 episodes of games. We trained the top layer for 2×10^6 epochs, and found no sign of continuing increase of performance. Each number of Hebbian-Q in **Table 1** represents the maximal value of the evaluation scores in 2 training trials.



Supplementary Figure 1: More simulation results. (**a**) The mutation rate of large synapses with zero hidden states under individual mutation (orange) and group mutation (blue) as a function of training epoch, in XOR task. Notice that they are almost the same at our parameter values. (**b**) Classification accuracy during the training process under dendrite-level homeostasis, in MNIST classification task. Notice that dendrite-level homeostasis hardly influences performance, comparing with **Figure 6a**. (**c**) Classification accuracy on the test dataset when training deep networks with 2 hidden layers or 1 hidden layer using Hebbian rule, in MNIST classification task. (**d**) Schematic of the structures of the networks examined in panel **c**, one with two hidden layers and the other with one hidden layer. Lower layers were trained using Hebbian algorithm (blue arrows), and the top layer was trained by gradient-descent algorithm supervisedly (red arrows). In panel **c,** error bars represent s.e.m. over 8 training trials. Error bars in panel **b** (not shown) are comparable with the width of the lines.

Acknowledgement

This work was supported by the National Natural Science Foundation of China (Grant no. 32000694 to Z. B., Grant no. 12175242 to D. Y. and Grant no. 32071141 to Y. Z.), the Natural Science Foundation of Shandong province (Grant no. ZR2019ZD34 to Y. Z.) and the start-up fund of the Institute for Future of Qingdao University (to Z. B.).

References

Abraham, W. C., Mason-Parker, S. E., Bear, M. F., Webb, S. and Tate, W. P. (2001), 'Heterosynaptic metaplasticity in the hippocampus in vivo: a BCM-like modifiable threshold for LTP', *Proc. Natl. Acad. Sci. USA* **98**, 10924–10929.

Amato, G., Carrara, F., Falchi, F., Gennaro, C. and Lagani, G. (2019), Hebbian learning meets deep convolutional neural networks, *in* 'International Conference on Image Analysis and Processing', pp. 324–334.

Anand, A., Racah, E., Ozair, S., Bengio, Y., Côté, M.-A. and Hjelm, R. D. (2019), Unsupervised state representation learning in Atari, *in* 'Conference on Neural Information Processing Systems'.

Andersson, M., Blomstrand, F. and Hanse, E. (2007), 'Astrocytes play a critical role in transient heterosynaptic depression in the rat hippocampal CA1 region', *J. Physiol.* **585**, 843–852.

Barnes, S. J., Franzoni, E., Jacobsen, R. I., Erdelyi, F., Szabo, G., Clopath, C., Keller, G. B. and Keck, T. (2017), 'Deprivation-induced homeostatic spine scaling in vivo is localized to dendritic branches that have undergone recent spine loss', *Neuron* **96**, 871–882.

Benna, M. K. and Fusi, S. (2016), 'Computational principles of synaptic memory consolidation', *Nat. Neurosc.* **19**, 1697–1706.

Bonhoeffer, T., Staiger, V. and Aertsen, A. (1989), 'Synaptic plasticity in rat hippocampal slice cultures: local "Hebbian" conjunction of pre- and postsynaptic stimulation leads to distributed synaptic enhancement', *Proc. Nat Acad. Sci. USA* **86**, 8113–8117.

Braun, U., Schäfer, A., Walter, H., Erk, S., Romanczuk-Seiferth, N., Haddad, L., Schweiger, J. I., Grimm, O., Heinz, A., Tost, H., Meyer-Lindenberg, A. and Bassett, D. S. (2015), 'Dynamic reconfiguration of frontal brain networks during executive cognition in humans', *Proc. Natl Acad. Sci. USA* **112**, 11678–11683.

Buch, E. R., Claudino, L., Quentin, R., Bönstrup, M. and Cohen, L. G. (2021), 'Consolidation of human skill linked to waking hippocampo-neocortical replay', *Cell Rep.* **35**, 109193.

Chen, J., Tan, Z., Zeng, L., Zhang, X., He, Y., Gao, W., Wu, X., Li, Y., Bu, B., Wang, W. and Duan, S. (2013), 'Heterosynaptic long-term depression mediated by ATP released from astrocytes', *Glia* **61**, 178–191.

Chistiakova, M., Bannon, N. M., Bazhenov, M. and Volgushev, M. (2014), 'Heterosynaptic plasticity: Multiple mechanisms and multiple roles', *Neuroscientist* **263**, 532–536.

Chistiakova, M. and Volgushev, M. (2009), 'Heterosynaptic plasticity in the neocortex', *Exp Brain Res.* **199**, 377–390.

Chiu, C. Q., Martenson, J. S., Yamazaki, M., Natsume, R., Sakimura, K., Tomita, S., Tavalin, S. J. and Higley, M. J. (2018), 'Input-specific NMDAR-dependent potentiation of dendritic GABAergic inhibition', *Neuron* **97**, 368–377.

Cornell-Bell, A. H., Finkbeiner, S. M., Cooper, M. S. and Smith, S. J. (1990), 'Glutamate induces calcium waves in cultured astrocytes: Long-range glial signaling', *Science* **247**, 470–473.

Crispino, M., Volpicelli, F. and Perrone-Capano, C. (2020), 'Role of the serotonin receptor 7 in brain plasticity: From development to disease', *Int J Mol Sci.* **21**, 505.

Deitmer, J. W., Verkhratsky, A. J. and Lohr, C. (1998), 'Calcium signalling in glial cells', *Cell Calcium* **24**, 405–416.

Duszkiewicz, A. J., McNamara, C. G., Takeuchi, T. and Genzel, L. (2019), 'Novelty and dopaminergic modulation of memory persistence: A tale of two systems', *Trends Neurosci.* **42**, 102–114.

Engert, F. and Bonhoeffer, T. (1997), 'Synapse specificity of long-term potentiation breaks down at short distances', *Nature* **388**, 279–284.

Faisal, A. A., Selen, L. P. J. and Wolpert, D. M. (2008), 'Noise in the nervous system', *Nat. Rev. Neurosci.* **9**, 292–303.

Fiete, I. R. and Seung, H. S. (2006), 'Gradient learning in spiking neural networks by dynamic perturbation of conductances', *Phys. Rev. Lett.* **97**, 048104.

Fischer, A. G. and Ullsperger, M. (2017), 'An update on the role of serotonin and its interplay with dopamine for reward', *Front. Hum. Neurosci.* **11**, e1004638.

Friston, K. (2018), 'Does predictive coding have a future?', *Nat. Neurosci.* **21**, 1019–1021.

Fusi, S., Drew, P. J. and Abbott, L. F. (2005), 'Cascade models of synaptically stored memories', *Neuron* **45**, 599–611.

Gao, S., Zhou, M., Wang, Y., Cheng, J., Yachi, H. and Wang, J. (2019), 'Dendritic neuron model with effective learning algorithms for classification, approximation, and prediction', *IEEE Trans. Neural Netw. Learn. Syst.* **30**, 601–614.

Hayama, T., Noguchi, J., Watanabe, S., Takahashi, N., Hayashi-Takagi, A., Ellis-Davies, G. C. R., Matsuzaki, M. and Kasai, H. (2013), 'GABA promotes the competitive selection of dendritic spines by controlling local Ca2+ signaling', *Nat. Neurosci.* **16**, 1409–1416.

Heilbron, M. and Chait, M. (2018), 'Great expectations: Is there evidence for predictive coding in auditory cortex?', *Neuroscience* **389**, 54–73.

Hennequin, G., Agnes, E. J. and Vogels, T. P. (2017), 'Inhibitory plasticity: Balance, control, and codependence', *Annu. Rev. Neurosci.* **40**, 557–579.

Herstel, L. J. and Wierenga, C. J. (2021), 'Network control through coordinated inhibition', *Curr. Opin. Neurobiol.* **67**, 34–41.

Hiratani, N. and Fukai, T. (2018), 'Redundancy in synaptic connections enables neurons to learn optimally', *Proc. Nat Acad. Sci. USA* **115**, E6871–E6879.

Holtmaat, A. and Caroni, P. (2016), 'Functional and structural underpinnings of neuronal assembly formation in learning', *Nat. Neurosci.* **19**, 1–10.

Hulme, S. R., Jones, O. D., Ireland, D. R. and Abraham, W. C. (2012), 'Calcium-dependent but action potential independent bcm-like metaplasticity in the hippocampus', *J. Neurosci.* **32**, 6785–6794.

Hulme, S. R., Jones, O. D., Raymond, C. R., Sah, P. and Abraham, W. C. (2013), 'Mechanisms of heterosynaptic metaplasticity', *Phil. Trans. R. Soc. B* **369**, 20130148.

Jacoby, S., Sims, R. E. and Hartell, N. A. (2001), 'Nitric oxide is required for the induction and heterosynaptic spread of long-term potentiation in rat cerebellar slices', *J. Physiol.* **535**, 825–839.

Josselyn, S. A. and Tonegawa, S. (2020), 'Memory engrams: Recalling the past and imagining the future', *Science* **367**, eaaw4325.

Kepecs, A. and Fishell, G. (2014), 'Interneuron cell types are fit to function', *Nature* **505**, 318–326.

Kossel, A., Bonhoeffer, T. and Bolz, J. (1990), 'Non-hebbian synapses in rat visual cortex', *Neuroreport* **1**, 115–118.

Krotov, D. and Hopfield, J. J. (2019), 'Unsupervised learning by competing hidden units', *Proc. Natl Acad. Sci. USA* **116**, 7723–7731.

Larkum, M. E. and Nevian, T. (2008), 'Synaptic clustering by dendritic signalling mechanisms', *Curr. Opin. Neurobiol.* **18**, 321–331.

LeCun, Y. (1986), Learning processes in an asymmetric threshold network, *in* F. Fogelman-Soulié, E. Bienenstock and G. Weisbuch, eds, 'Disordered Systems and Biological Organization', Springer-Verlag, Les Houches, pp. 233–240.

Lee, C. M., Stoelzel, C., Chistiakova, M. and Volgushev, M. (2012), 'Heterosynaptic plasticity induced by intracellular tetanization in layer 2/3 pyramidal neurons in rat auditory cortex', *J. Physiol.* **590**, 2253–2271.

Lee, D.-H., Zhang, S., Fischer, A. and Bengio, Y. (2015), Difference target propagation, *in* 'Joint Eur. Conf. Machine Learning Knowl. Discov. Databases'.

Legenstein, R., Chase, S. M., Schwartz, A. B. and Maass, W. (2010), 'A reward-modulated hebbian learning rule can explain experimentally observed network reorganization in a brain control task', *J. Neurosci.* **30**, 8400–8410.

Li, J., Zhang, D., Liang, A., Liang, B., Wang, Z., Cai, Y., Gao, M., Gao, Z., Chang, S., Jiao, B., Huang, R. and Liu, M. (2017), 'High transition frequencies of dynamic functional connectivity states in the creative brain', *Sci. Rep.* **7**, 46072.

Lillicrap, T. P., Cownden, D., Tweed, D. B. and Akerman, C. J. (2016), 'Random synaptic feedback weights support error backpropagation for deep learning', *Nat. Commun.* **7**, 13276.

Lillicrap, T. P., Santoro, A., Marris, L., Akerman, C. J. and Hinton, G. (2020), 'Backpropagation and the brain', *Nat. Rev. Neurosci.* **21**, 335–346.

Loewenstein, Y. (2008), 'Robustness of learning that is based on covariancedriven synaptic plasticity', *PLoS Comput. Biol.* **4**, e1000007.

Lu, J. and Zuo, Y. (2017), 'Clustered structural and functional plasticity of dendritic spines', *Brain Res. Bull.* **129**, 18–22.

Manchev, N. and Spratling, M. (2020), 'Target propagation in recurrent neural networks', *J. Mach. Learn Res.* **7**, 46072.

Miconi, T. (2017), 'Biologically plausible learning in recurrent neural networks reproduces neural dynamics observed during cognitive tasks', *eLife* **97**, e20899.

Miconi, T., Rawal, A., Clune, J. and Stanley, K. O. (2020), Backpropamine: training self-modifying neural networks with differentiable neuromodulated plasticity, *in* 'International Conference on Learning Representations'.

Min, R. and Nevian, T. (2012), 'Astrocyte signaling controls spike timing-dependent depression at neocortical synapses', *Nat. Neurosci.* **15**, 746–753.

Mirjalili, S. (2019), *Evolutionary Algorithms and Neural Networks: Theory and Applications*, Springer, Switzerland.

Mnih, V., Badia, A. P., Mirza, M., Graves, A., Harley, T., Lillicrap, T. P., Silver, D. and Kavukcuoglu, K. (2016), 'Asynchronous methods for deep reinforcement learning', *arXiv:1602.01783* .

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Fidjeland, M. R. A. K., Ostrovski, G. et al. (2015), 'Human-level control through deep reinforcement learning', *Nature* **518**, 529–533.

Mongillo, G., Rumpel, S. and Loewenstein, Y. (2017), 'Intrinsic volatility of synaptic connections - a challenge to the synaptic trace theory of memory', *Curr. Opin. Neurobiol.* **46**, 7–13.

Najarro, E. and Risi, S. (2020), Meta-learning through hebbian plasticity in random networks, *in* 'Advances in Neural Information Processing Systems'.

O'Connor, D. H., Wittenberg, G. M. and Wang, S. S.-H. (2005), 'Graded bidirectional synaptic plasticity is composed of switch-like unitary events', *Proc. Natl Acad. Sci. USA* **102**, 9679–9684.

Oldham, S. and Fornito, A. (2019), 'The development of brain network hubs', *Dev. Cogn. Neurosci.* **36**, 100607.

Ostroumov, A. and Dani, J. A. (2018), 'Inhibitory plasticity of mesocorticolimbic circuits in addiction and mental illness', *Trends Neurosci.* **41**, 898–910.

Payeur, A., Guerguiev, J., Zenke, F., Richards, B. A. and Naud, R. (2021), 'Burst-dependent synaptic plasticity can coordinate learning in hierarchical circuits', *Nat. Neurosci.* **24**, 1010–1019.

Pfeffer, C. K., Xue, M., He, M., Huang, Z. J. and Scanziani, M. (2013), 'Inhibition of inhibition in visual cortex: the logic of connections between molecularly distinct interneurons', *Nat. Neurosci.* **16**, 1068–1076.

Qian, C., Bian, C. and Feng, C. (2020), Subset selection by pareto optimization with recombination, *in* 'Proceedings of the AAAI Conference on Artificial Intelligence'.

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

Rasch, B. and Born, J. (2013), 'About sleep's role in memory', *Physiol Rev.* **93**, 681–766.

Roelfsema, P. R. and Holtmaat, A. (2018), 'Control of synaptic plasticity in deep cortical networks', *Nat. Rev. Neurosci.* **19**, 166–180.

Royer, S. and Paré, D. (2003), 'Conservation of total synaptic weight through balanced synaptic depression and potentiation', *Nature* **422**, 518–522.

Scanziani, M., Malenka, R. C. and Nicoll, R. A. (1996), 'Role of intercellular interactions in heterosynaptic long-term depression', *Nature* **380**, 446–450.

Schiess, M., Urbanczik, R. and Senn, W. (2016), 'Somato-dendritic synaptic plasticity and error-backpropagation in active dendrites', *PLoS Comput. Biol.* **12**, e1004638.

Schuman, E. M. and Madison, D. V. (1991), 'A requirement for the intercellular messenger nitric oxide in long-term potentiation', *Science* **254**, 1503–1506.

Schuman, E. M. and Madison, D. V. (1994), 'Locally distributed synaptic potentiation in the hippocampus', *Science* **263**, 532–536.

Stimberg, M., Brette, R. and Goodman, D. F. M. (2019), 'Brian 2, an intuitive and efficient neural simulator', *eLife* **8**, e47314.

Stott, J. J. and Redish, A. D. (2015), 'Representations of value in the brain: An embarrassment of riches?', *PLoS Biol.* **13**, e1002174.

Such, F. P., Madhavan, V., Conti, E., Lehman, J., Stanley, K. O. and Clune, J. (2017), 'Deep neuroevolution: Genetic algorithms are a competitive alternative for training deep neural networks for reinforcement learning', *arXiv:1712.06567* .

Sutton, M. A., Ito, H. T., Cressy, P., Kempf, C., C.Woo, J. and Schuman, E. M. (2006), 'Miniature neurotransmission stabilizes synaptic function via tonic suppression of local dendritic protein synthesis', *Cell* **125**, 785–799.

Takahashi, N., Kitamura, K., Matsuo, N., Mayford, M., Kano, M., Matsuki, N. and Ikegaya, Y. (2012), 'Locally synchronized synaptic inputs', *Science* **335**, 353–356.

Takahashi, S. and Sakurai, Y. (2009), 'Sub-millisecond firing synchrony of closely neighboring pyramidal neurons in hippocampal ca1 of rats during delayed non-matching to sample task', *Front. Neural Circuits* **3**, 9.

Tsumoto, T. and Suda, K. (1979), 'Cross-depression: an electrophysiological manifestation of binocular competition in the developing visual cortex', *Brain Res.* **168**, 190–194.

Turrigiano, G. (2011), 'Too many cooks? Intrinsic and synaptic homeostatic mechanisms in cortical circuit refinement', *Annu. Rev. Neurosci.* **34**, 89–103.

Ujfalussy, B. B. and Makara, J. K. (2020), 'Impact of functional synapse clusters on neuronal response selectivity', *Nat. Commun.* **11**, 1413.

Wall, N. R., Parra, M. D. L., Sorokin, J. M., Taniguchi, H., Huang, Z. J. and Callaway, E. M. (2016), 'Brain-wide maps of synaptic input to cortical interneurons', *J. Neurosci.* **36**, 4000–4009.

White, G., Levy, W. B. and Steward, O. (1990), 'Spatial overlap between populations of synapses determines the extent of their associative interaction during the induction of long-term potentiation and depression', *J. Neurophysiol.* **64**, 1186–1198.

Wolpert, D. and Macready, W. (1997), 'No free lunch theorems for optimization', *IEEE Trans. Evol. Comput.* **1**, 67–82.

Zhou, Z.-H., Yu, Y. and Qian, C. (2019), *Evolutionary Learning: Advances in Theories and Algorithms*, Springer, Switzerland.