PLOS | COMPUTATIONAL BIOLOGY

RESEARCH ARTICLE

Multi-layer network utilizing rewarded spike time dependent plasticity to learn a foraging task

Pavel Sanda[1☯], **Steven Skorheim**[2☯], **Maxim Bazhenov**[1*]

1 Department of Medicine, University of California, San Diego, La Jolla, California, United States of America, **2** Information and Systems Sciences Lab, HRL Laboratories, LLC, Malibu, California, United States of America

☯ These authors contributed equally to this work.
* bazhenov@salk.edu

OPEN ACCESS

Citation: Sanda P, Skorheim S, Bazhenov M (2017) Multi-layer network utilizing rewarded spike time dependent plasticity to learn a foraging task. PLoS Comput Biol 13(9): e1005705. https://doi.org/10.1371/journal.pcbi.1005705

Editor: Abigail Morrison, Research Center Jülich, GERMANY

Received: January 18, 2017

Accepted: July 26, 2017

Published: September 29, 2017

Data Availability Statement: Additional data related to this project and its further developments together with the model demonstrations can be found online at http://www.bazhlab.ucsd.edu/brg/decision-making/.

Funding: This work was supported by Office of Naval Research grant (Multi University Research Initiative: N000141612829, N000141612415) [MB] and by the National Institute of Health (R01 MH087631 and R01 DC012943) [MB]. The funders had no role in study design, data collection

Abstract

Neural networks with a single plastic layer employing reward modulated spike time dependent plasticity (STDP) are capable of learning simple foraging tasks. Here we demonstrate advanced pattern discrimination and continuous learning in a network of spiking neurons with multiple plastic layers. The network utilized both reward modulated and non-reward modulated STDP and implemented multiple mechanisms for homeostatic regulation of synaptic efficacy, including heterosynaptic plasticity, gain control, output balancing, activity normalization of rewarded STDP and hard limits on synaptic strength. We found that addition of a hidden layer of neurons employing non-rewarded STDP created neurons that responded to the specific combinations of inputs and thus performed basic classification of the input patterns. When combined with a following layer of neurons implementing rewarded STDP, the network was able to learn, despite the absence of labeled training data, discrimination between rewarding patterns and the patterns designated as punishing. Synaptic noise allowed for trial-and-error learning that helped to identify the goal-oriented strategies which were effective in task solving. The study predicts a critical set of properties of the spiking neuronal network with STDP that was sufficient to solve a complex foraging task involving pattern classification and decision making.

Author summary

This study explores how intelligent behavior emerges from the basic principles known at the cellular level of biological neuronal network dynamics. Compared to the approaches used in the artificial intelligence community, we applied biologically realistic modeling of neuronal dynamics and plasticity. The building blocks of the model are spiking neurons, spike-time dependent plasticity (STDP) and homeostatic rules, known experimentally, which are shown to play a fundamental role in both keeping the network stable and capable of continous learning. Our study predicts that a combination of these principles makes possible a foraging behavior in a previously unknown environment, including pattern

and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

classification to distinct between environment shapes which are rewarded and those which are punished and decision making to select the optimal strategy to acquire the maximal number of the rewarded elements. To solve this complex task we used multi-layer neuronal processing that implemented pattern generalization by unsupervised STDP at the earlier processing step, as commonly observed in the animal and human sensory processing, followed by reinforcement learning at the later steps. In the model, the intelligent behavior emerged spontaneously due to the network organization implementing both local unsupervised plasticity and reward feedback resulting from a successful behavior in the environment.

Introduction

Biologically inspired neural networks should be capable of performing sophisticated information processing that the human and animal brains can perform. Information processing by the brain is deeply multilayered and involves many sequential steps before sensory information can be interpreted and translated into behavior. What makes this cascade powerful and unique is its capability to learn and respond to an ever changing environment. In the most studied sensory pathway—the visual one—the sensory input gets progressively more general, though the stages at which visual learning occurs are still a matter of controversy and different plasticity mechanisms might be operating at different processing steps [1]. Eventually sensory information reaches decision centers (such as lateral intraparietal cortex) which govern behavior and those centers are under the influence of reward signals [2, 3]. It has been shown that reinforcement based on the reward can be vital to visual learning [4]. To what extent reward influences learning in the sensory cortices seems to be task dependent and different studies showed both learning with [5, 6] and without [7, 8] the presence of reward.

While a great deal of research has gone into understanding the mechanisms of learning, it is still not fully known how learning at the cellular level gives rise to the learning at the level of animal behavior. Most learning models successfully employ hebbian type learning principles formulated as an abstract rule [9] without being concerned with the cellular level mechanisms. The most promising model of learning at the synaptic level is spike time dependent plasticity (STDP) [10]. In STDP, when a presynaptic cell fires shortly before a post synaptic cell, the synapse between them increases in strength. If the postsynaptic cell fires before the presynaptic cell, the synapse decreases in strength. Such basic form of STDP is only capable of unsupervised learning. By storing STDP events in the form of synaptic tags [11], a delayed reward signal can enable a network to perform reinforcement learning [12]. This mechanism relies on the dopamine modulation of synaptic tags created by STDP [13, 14, 15]. A small number of studies have attempted to perform task learning using rewarded STDP including studies that attempted to maintain the stability of a virtual robotic arm [16, 17].

In theoretical studies, STDP commonly leads to the net increase of synaptic strength across the network [18], causing large numbers of synapses to climb toward their maximum values. This would result in a highly dysfunctional brain and can eventually promote an epileptic state [19]. To prevent runaway synaptic dynamics, various regulatory synaptic mechanisms maintain the distribution of synaptic weights in biological networks. This includes homeostatic mechanisms when a group of cells that experiences high firing frequencies over a prolonged period becomes less responsive to excitatory input by effectively reducing incoming excitatory synaptic weights [20]. Another regulatory mechanism, heterosynaptic plasticity, responds to the increase of a synaptic weight by reducing all other incoming synaptic strengths of the same

post synaptic neuron [21, 22, 23, 24] preventing neurons from becoming overactive [25, 26, 27, 28].

In this new work, we built a model containing two plastic layers inspired by the pathways known in mammalian brain from sensory to higher order sensory-motor areas responsible for decision making [29, 4]. The first layer of the model uses unsupervised (unrewarded) learning to classify the input while the second layer (based on rewarded STDP) is responsible for decision making. As a whole the network simulated the brain of an agent moving through an unknown environment, continuously learning distinct input patterns of food and adjusting synaptic weights controlling its movement according to the reward and punishment signals based on the shape of the different configurations of food particles that it acquires.

We demonstrate that such multilayered network combining rewarded and nonrewarded STDP and synaptic regulatory mechanisms is capable of solving a more complex foraging task, involving discrimination between elementary patterns of food, than it was previously reported in the networks with only one plastic layer based on rewarded STDP [30].

Results

The organization of the foraging neural network

In this study we have expanded upon a network with a single plastic layer which was designed to learn and perform a foraging task in a virtual environment [30]. The new model incorporated two layers of plastic synaptic connections and implemented both rewarded and non-rewarded spike time dependent plasticity (STDP), see Fig 1. It included one-to-many connections from the input layer to the large middle layer and all to all connections from the middle layer to the output layer. Each middle layer cell received inputs from a few randomly selected input neurons. Reducing the number of inputs to each middle layer cell greatly reduced the computational power required for the simulation and was also a realistic approximation of the brain anatomy [31]. Connections between all three layers were set up initially at random and as such any correlation between the input pattern and the network response was random. As the simulation progressed, the learning in the form of changes in synaptic weights caused the network to become more proficient in obtaining food. Connections from the input to the middle layer employed non-rewarded STDP. This form of unsupervised learning allowed the network to respond more readily to the common features (such as spatial patterns), and common combinations of features, within the environment. Connections from the middle to output layers were controlled by the rewarded STDP. Thus, STDP traces were recorded as synaptic tags [13] and only applied if a reward signal was received later. The strength of traces declined, with time causing older traces to have less impact if a reward was eventually applied. A global reward signal was received whenever the network directed the virtual agent to a goal (”food” location). Synapses that observed pre-before-post combination of the action potentials within a few movement cycles before the reward became stronger; therefore, it increased the probability for the same postsynaptic cell to spike when the same input occurred in the future. The supervised learning allowed the network to develop approach or avoidance behavior based on the reward predicting value of the stimuli.

To evaluate the training performance, we compared our STDP based learning algorithm to a set of simple heuristic algorithms. In the best performing heuristic algorithm, when no food was present in the visual field, the agent either continued in the same direction it moved on the last step with a 98% probability or turned at 45 degrees left or right with a 2% probability. When the food was present, the heuristic algorithm searched through all possible combinations of 5 moves within the visual field. It then chose the set of moves that would result in the most food being obtained and made the first move from that set. If multiple sets of moves

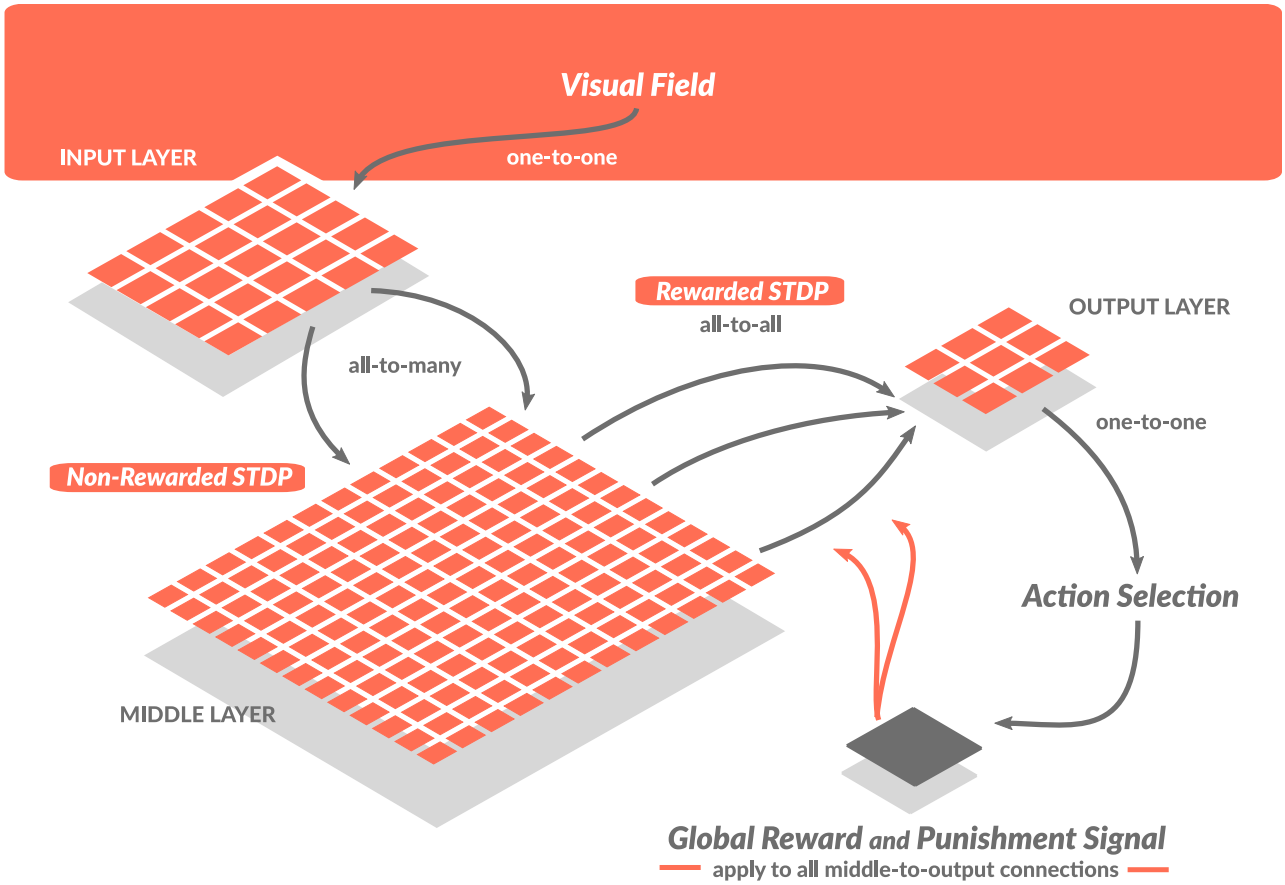

Fig 1. Network organization. The network organization is a simplification of the information processing flow known in the visual pathway, involving mapping of the sensory input into the higher level representations and then using them for decision making in the prefrontal cortex [4]. Input indicating the position of food particles relative to the virtual agent (positioned in the center of the field) was simulated as a set of excitatory inputs to the input layer neurons. In the model, each input layer cell sends one excitatory and one inhibitory connection to each of the cells in the middle layer where object representation is built. Each middle layer cell sends one excitatory and one inhibitory connection to 9 cells in the output layer. The most active cell in the output layer (size 3x3) decides the direction of subsequent movement. Excitatory connections from the input to the middle layer are subject to non-rewarded STDP. Excitatory connections from the middle layer to the output layer are subject to rewarded STDP where reward depends on whenever a move results in food acquisition. Inhibitory connections from a given cell always match the average strength of the excitatory outputs of the same presynaptic cell.

https://doi.org/10.1371/journal.pcbi.1005705.g001

obtained the same number of food particles, the set which obtained food sooner was selected. If multiple sets had the same sequence of food being obtained, one of those sets was chosen randomly. Under standard conditions (such as given density of food particles in the environment) this strategy had an average food acquisition rate of 56%. Our multilayered network combining rewarded and nonrewarded STDP was able to achieve very close acquisition rate of 52%. For comparison, in the previous study we were able to achieve average food acquisition rates of about 48% [30].

This paper is divided into two main sections. In the first part we considered a simple task to obtain food at any location regardless of the spatial pattern of the food particles. In the second part, we increased the complexity of the task, so the agent was trained to pick up food organized within a specific spatial pattern (horizontal bar) and avoid any food particles organized within another pattern (vertical bar). The last task could not be achieved by the simpler network with only one plastic layer [30]. Thus, the main achievement of the current model is that

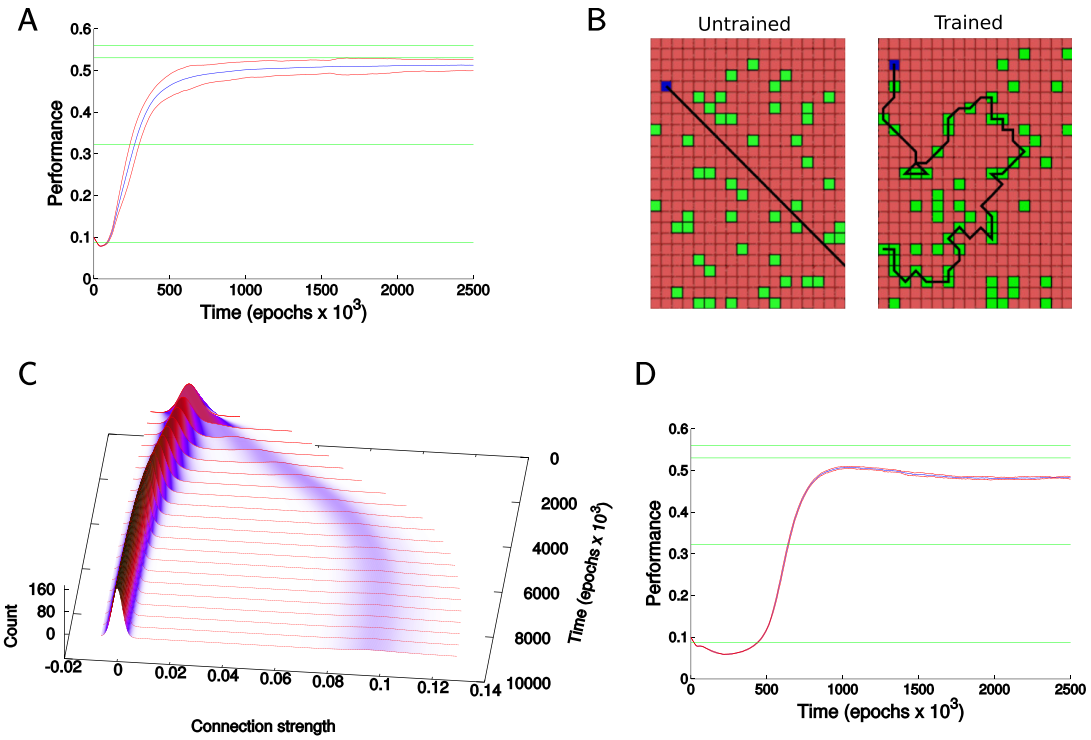

Fig 2. Basic two plastic layers model. A) Dynamics of the network performance over time. The Y axis is the rate of food acquisition per move as an exponential moving average. The X axis is time in epochs (each epoch consists of 600 time steps). Final food acquisition rates often exceeded 52% which is higher than in the previous models [30] and performance was very reliable across trials. Blue lines represent the mean performance and red lines represent the mean the standard deviation. Green lines represent performance of heuristic algorithms—lowest: random movement (98% go straight, 2% turn); second to lowest: as the lowest, but move to any food if directly adjacent; third to lowest: move toward closest food among all particles within the visual field; highest: try all possible combinations within the visual field, then take the first move of most successful set. B) Examples of the agent movement through the environment. Left before training, right after $\sim 120 \times 10^3$ training steps. C) Time evolution of the strength of excitatory connections from the input to the middle layer during the first 10000 epochs (the profile at any point in time corresponds to the histogram of strengths). Note a group of synapses that significantly increased their strength. D) Performance of the model without normalization of synaptic potentiation based on spike rate. Note that the model showed declines in performance occurring after initial peak in all of the trials and over a wide range of testing parameters. Green lines as in A.

https://doi.org/10.1371/journal.pcbi.1005705.g002

it sets the stage for further development into the learning of more complex tasks, however, it also shows a modest improvement compared to the simpler models performing with food acquisition rates above 51%.

Simple foraging behavior in the model with two plastic layers

In Fig 2 we show the dynamics of a complete model for a simple foraging task of food acquisition regardless of the spatial arrangement of the food particles. Various homeostatic mechanisms were applied in order to obtain the best performance for a simple foraging task of acquiring food; the contribution of individual mechanisms will be discussed in detail in the following sections. Over time performance increased (Fig 2A) as the agent learned to move into the direction of food (Fig 2B). The Fig 2C shows the dynamics of the strengths of excitatory connections from the input to the middle layer over time. As the network learned a task, a small fraction of synapses increased and moved apart from the majority of other synapses

which were slightly reduced. This small subset of synapses played the decisive role in learning, and maintaining their size and stable distance from the bulk of synapses kept the network behavior stable.

Normalization of synaptic potentiation. In the early versions of the model we observed a small but consistent decline in performance over time after performance reached its peak (see Fig 2D). This dynamic seems to be an inherent property of the networks that experience a positive net reward value. Indeed, synapses that are regularly active would have an advantage in gaining synaptic strength regardless of whether their activation improves the probability of reward or not. We implemented a new normalization mechanism to counteract this effect. The average value of STDP traces created by the synapse over an extended period (thousands of cycles) was recorded for each synapse between the middle and the output layer. The magnitude of each STDP trace at a synapse was then normalized by this average value before being applied as a change in synaptic strength. Thus a synapse that regularly experienced strong traces from STDP events, had progressively reduced potentiation from new STDP events. As a result, the slow decline in network performance was completely eliminated and significant improvements were observed in the learning speed (compare Fig 2A and 2D). Learning times are likely improved due to much more rapid development of rarely used and weak connections early in the learning process. Indeed, in this new model synaptic changes due to rewarded events may be significantly larger for the connections experiencing only rare positive STDP events. Thus implementation of synaptic trace normalization promoted competition between synapses that have very few strong STDP traces and those with many strong STDP events and eliminated bias toward synapses with higher activity. Importantly this normalization mechanism is consistent with experimental data [32, 33, 34].

Balancing of synaptic input: Effect of heterosynaptic plasticity. Next we tested the impact of the different synaptic balancing mechanisms on the model performance (Fig 3). Heterosynaptic plasticity was the most important single balancing rule [27, 28]. Its role was to ensure that the total synaptic strength of all incoming synapses to an individual neuron remains constant. Whenever a single input changed in strength, all other incoming synaptic connections of that neuron were adjusted to compensate (see Methods). As a result of maintaining a constant synaptic input, individual synapses must compete to be effective at activating the postsynaptic neuron and only those synapses which are regularly rewarded can reach sufficient strength to control the network. Without input side balancing (heterosynaptic plasticity), performance fell to near chance levels as synaptic strengths diverged to the ceiling and floor values (see Fig 3, orange line). Even when synaptic weight totals were maintained by keeping the sum of outgoing synaptic strength constant (output counter part of heterosynaptic plasticity, which would be a stronger version of the output balancing rule we discuss in the next section), performance remained at the near chance level (Fig 3, purple line) and distribution of synaptic weights became bimodal with a small number of connections reaching ceiling values.

Output balancing. Our previous study revealed the importance of balancing the strength of the output synapses [30]. In this new study, implementation of the output part of synaptic balancing was to reduce the rate of synaptic growth in the neurons that already had high total synaptic output. This effectively prevented a very small number of neurons from controlling the entire network. Thus, for each middle-layer cell, increments of the strength of outgoing synapses resulting from rewarded STDP events were divided by the ratio of the current sum of synaptic outputs to the initial sum of synaptic outputs of the same cells (see Methods). The result was that synapses originating from the neurons with many strong outputs were not able to increase their synaptic strength as quickly as synapses from the neurons with a weak output. This gave a competitive advantage to the later neurons. It helped to control synaptic output,

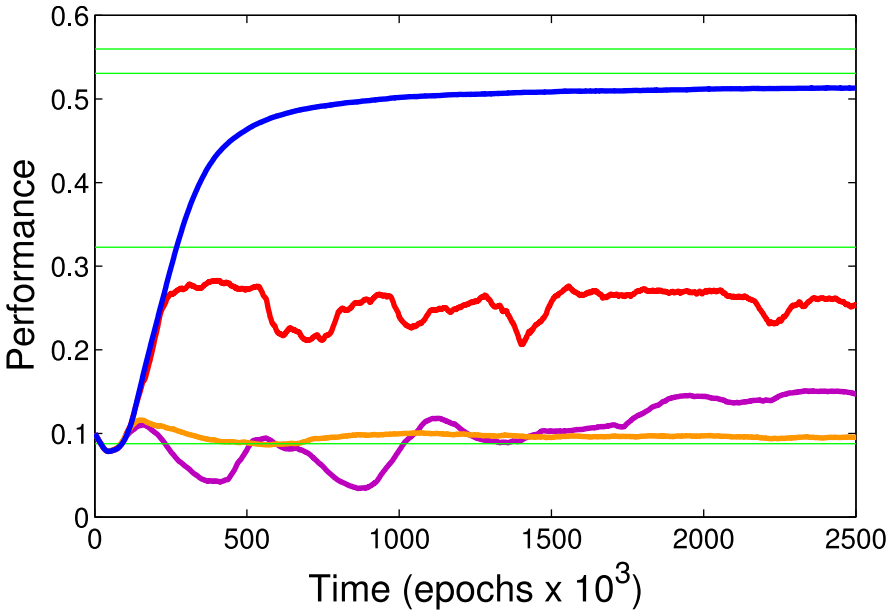

Fig 3. Effect of specific synaptic balancing rules on the model performance. Blue—full model with the input side heterosynaptic plasticity and output balancing. Orange—no input side heterosynaptic plasticity but output balancing is still implemented. Purple—both input side heterosynaptic plasticity and output balancing are removed. To partially compensate for the loss of the output balancing we applied another rule—the sum of all the outgoing synaptic strengths was held constant. Nevertheless, performance was further reduced to almost chance level. Red—input side heterosynaptic plasticity is implemented but no output balancing. Note the greatest impact of heterosynaptic plasticity on the model performance. Output balancing made lesser impact but still was crucial to maintaining high performance.

https://doi.org/10.1371/journal.pcbi.1005705.g003

thus preventing over-representation by the cells whose activities were most often correlated with the rewards (see S1 Fig). The performance of the full model simulated without this rule is shown by the red line in Fig 3.

Homeostatic gain control. To maintain the average desired firing rate of neurons (over long time), we implemented homeostatic scaling [20] independently for the middle and the output layer neurons. When the actual firing rate was below (above) the target rate, the efficacy of all excitatory synaptic inputs was scaled up (down) in small steps, which promoted the firing rate move towards the target. Over time this caused the firing rate of the network to gravitate around the target value. By changing the target firing rate of neurons within each layer, it was possible to control the sparseness of activity within that layer. Thus we next tested the impact of the target (homeostatic) firing rate on the model performance.

For each simulation experiment, the target firing rate of neurons was set to a different level (Fig 4). We then calculated the performance for each combination of the target firing rates in two layers. When firing rates were too low, it would often prevent information to propagate down the network. Target firing rates that were too high potentially led to high peak performance but also unstable dynamics and to some form of overlearning causing reduction in performance over time. Lower firing rates and sparse activity in the middle layer were important in achieving high performance. A higher firing rate of spiking (>0.9 Hz) in the output layer revealed the best performance. This was observed because increasing mean firing rates of the output layer neurons reduced the likelihood of the spike count ties, which improved decision making. It also reduced the fraction of the movement cycles where no output spikes occurred at all (see S2 Fig in supporting information).

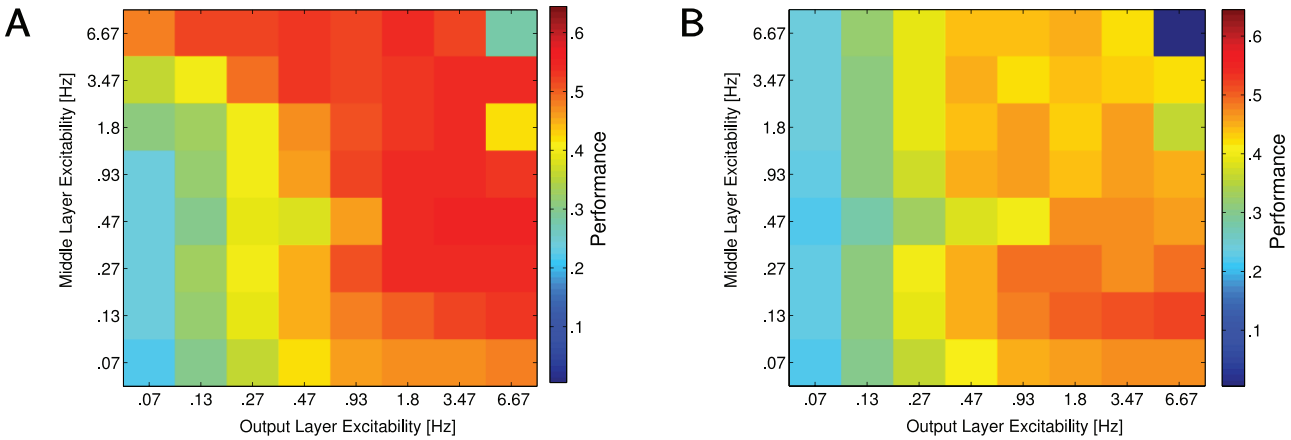

Fig 4. Performance plot over excitability in the middle and output network layers. A) A heat chart of peak performance under a range of target firing rates in the middle layer (y-axis) and the output layer (x-axis). B) A chart showing final performance over the same range of conditions. Points where large differences exist between the peak and the final performance charts generally indicate conditions where the network performed much better early in the simulation but suffered declines in performance over time.

https://doi.org/10.1371/journal.pcbi.1005705.g004

Effect of performance decline after reaching a maximum was likely related to an inability of the network to maintain synaptic efficacy in a moderate range. When excitability of the post synaptic cells was too high, spike events that normally only caused post synaptic spikes when occurring in coincidence with other events could eventually result in connections strong enough that even a single spike event may cause spike responses in postsynaptic neurons.

Role of inhibition. Feedforward synaptic inhibition was implemented in the model (see Methods) and was necessary for optimal behavior of the network. Thus each layer projecting excitatory connections to the following layer was also projecting inhibition and the total strength of inhibition was equal to the total strength of excitation. Removing the inhibition, especially from the middle to the output layer, caused drastic reduction in performance. Fig 5A shows results of simulations in three networks with different inhibitory connectivity. The green line shows the baseline model. Removing the inhibition from the input to the middle layer revealed only a moderate decline in performance (Fig 5A, black line). When the inhibition was removed from the middle to the output layer (blue line) or there was no inhibition at all (red line), the network performed near chance level.

We found that layers not receiving inhibitory input had far greater variance in their activity per epoch (Fig 5B and 5C). Large variance in the output layer activity was associated with very poor performance. Indeed, large variance when middle->output inhibition was turned off indicates a great number of epochs with either no output activity or simultaneous spiking activity in all of the output layer neurons (see S3 Fig). Synaptic inhibition was necessary to maintain moderate levels of activity despite changes in the input strength. Thus, e.g., when many food particles were found within the network "visual field", the activity in the input layer was high. Without inhibition at all high activity of the input layer propagated all the way to the output layer leading to increased variability between trials and great performance reduction. Eliminating inhibition only from the input to middle layers greatly increased the variance of activity in the middle layer but had far less impact on the final performance because activity in the output layer could still be constrained by feedforward inhibition to that layer.

Role of synaptic noise. The model included some level of randomness in all of its synaptic connections. In particular, each neurotransmitter release, following presynaptic spike, was modulated by the variable implementing a noisy component. Based on our previous work

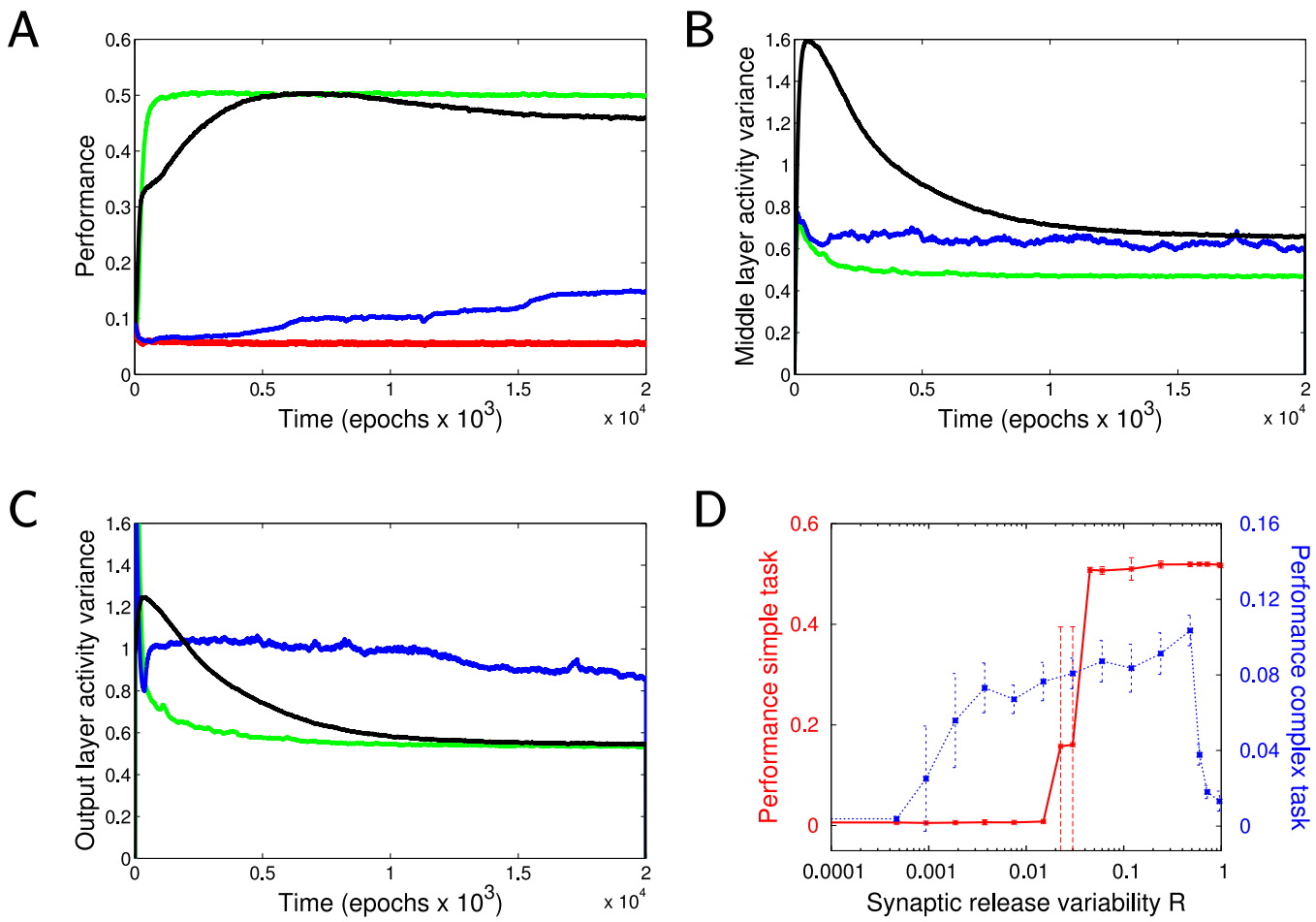

Fig 5. Effect of synaptic inhibition and noise. A) Model performance reduced when no inhibition was present in the network. Green—baseline (inhibition is present), black—no inhibition from input to middle layer, blue—no inhibition from middle to output layer, red—no inhibition in both layers. B) Coefficient of variance in the activity per epoch of the middle layer. C) Coefficient of variance in the output layer for the same sets of trials. Colors have the same meaning for all three panels. D) Model performance for different level of variability in synaptic release. Red—performance for the simple foraging task, blue—performance for the complex foraging task. Each dot is the average (10 independent trials) of performance measurements at the time of 2.10^4 epochs.

https://doi.org/10.1371/journal.pcbi.1005705.g005

using the simplified model [30], we expected that such variability would be necessary in order for the agent to learn. Indeed, reward based learning is a trial and error process and noise allows opportunities for the correct response to occur by chance and then be rewarded and learned. Fig 5D (red line) shows that the low noise amplitude prevented the model from learning successfully, increasing the level of noise beyond a certain level led to an abrupt performance increase. When similar analysis was applied to the complex foraging task, as described in the next section, we found a smoother transition (Fig 5D, blue line). Nevertheless, a certain level of noise was still required for our model to learn successfully.

Complex foraging behavior: Pattern classification and learning

Our previous model with only one plastic layer was able to solve a simple foraging task of food acquisition at any location regardless of the food pattern shape [30]. It failed, however, to distinguish between patterns of different shape and to acquire only food particles arranged in a particular way. In contrast, the new model with two plastic layers, was able to accomplish this

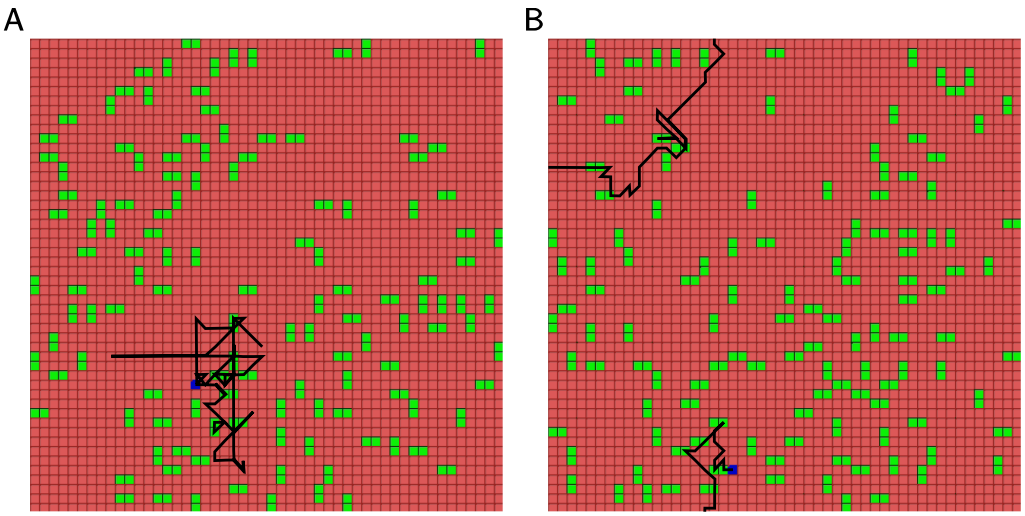

Fig 6. Environment arrangement and agent behavior. Green particles represent food; horizontal patterns are rewarded, vertical patterns are punished. Blue dot is a starting point of the agent. Black line represents its movement in sequential steps. A) Moving behavior before learning. B) Moving behavior after learning. Note the avoidance of vertical patterns.

https://doi.org/10.1371/journal.pcbi.1005705.g006

more complex task. In the following experiments the food in the environment was arranged into vertical and horizontal pairs of locations. When the virtual agent arrived at the location of either particle from a pair, both particles were removed and another pair of food was generated at a random location on the map. Obtaining horizontally arranged food was rewarded, while obtaining vertically arranged food was punished. Thus, vertically arranged food was considered to be "bad" or "toxic". When food particles were placed in the environment, it was arranged in such a way that one pattern was never adjacent to another to avoid ambiguity of the input. The two layer model was able to quickly learn avoiding vertical food arrangements and only acquire food particles arranged horizontally (see Fig 6).

Complementary roles of rewarded and non-rewarded STDP. Simple pattern classification and approach/avoidance behavior was achieved in our model through a combination of rewarded and non-rewarded STDP. Inputs to the middle layer were affected by non-reward modulated STDP (see Methods). The upper limit on the strength of synaptic connections from the input to the middle layer neurons was such that no single synaptic input was sufficient to evoke a spike in a middle layer cell. Inhibitory connections from the input to the middle layer were shown to have a minimal impact on the base model performance (Fig 5A, black line) and we did not include inhibition in the pattern-classifying network because it reduced the reliability for a pair of inputs to evoke a response in the middle layer.

Over time, middle layer neurons learned to respond only when a specific pair of input layer neurons fired together. More commonly co-occurring pairs of events developed significantly more representation in the middle layer. As food particles were presented as horizontal or vertical pairs in the environment, most middle layer neurons became responsive randomly to either a vertical or a horizontal pair of cells in the input layer (see Fig 7), thus giving an example of a basic generalization behavior while still maintaining location based specificity [35]. The connectivity pattern from the input layer to the middle layer impacted both the model performance and its computational complexity. We found that the optimal compromise between the learning and the computational performance was reached with fan-in about 9

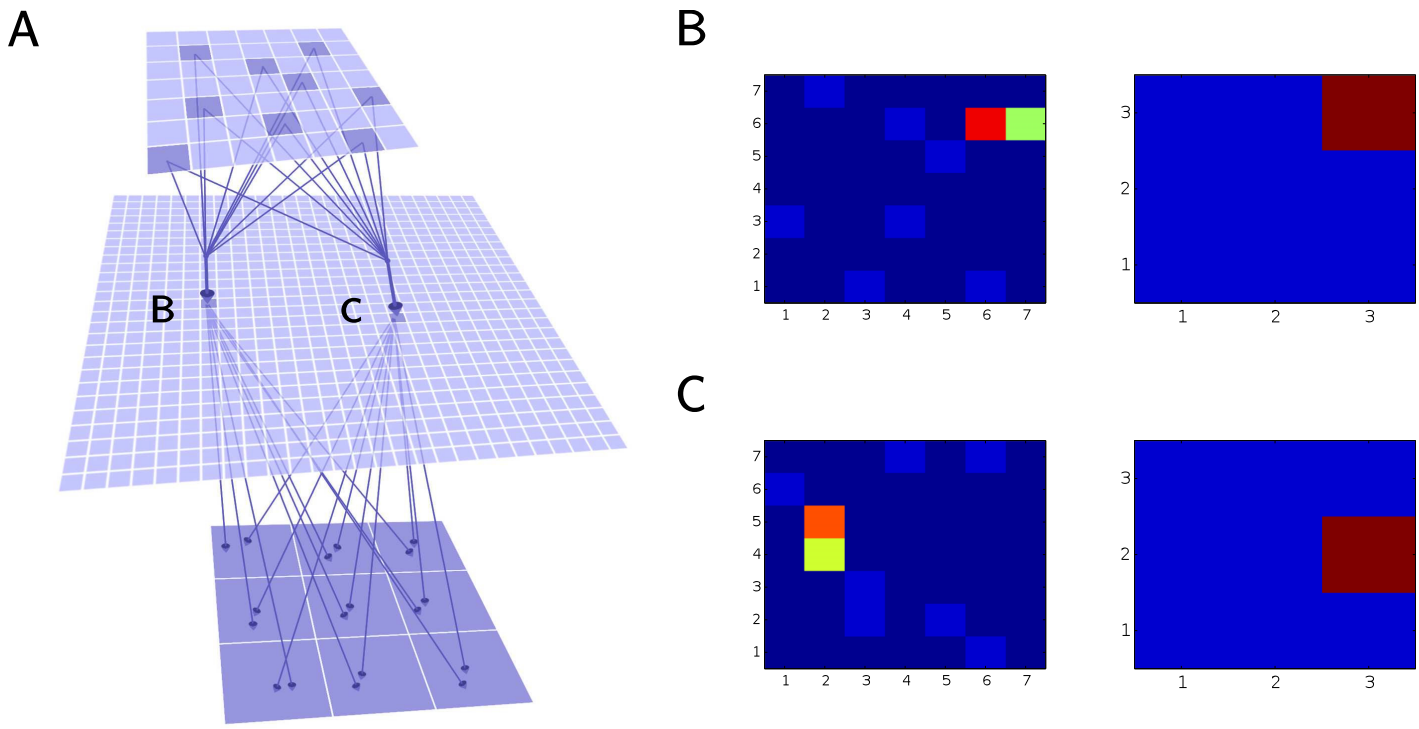

Fig 7. Structure of synaptic connectivity after learning. A) Network connectivity diagram for two typical middle layer cells. On top is input layer (visual field), below is the middle layer network and bottom is the output layer (direction of next movement). B&C) Strengths of the synaptic inputs (left) and outputs (right) of two typical middle layer neurons (labeled in panel A as 'B','C'), after successful training. The network has been trained to move toward horizontal food pairs and away from vertical food pairs. Red represents the highest synaptic strength while blue represents the lowest. B) A characteristic middle layer cell that has become responsive to a pair of horizontally arranged inputs on the top right of the visual field and excites the top-right output cell that directs movement of the virtual agent toward the food pair. C) Another characteristic middle layer cell that has become responsive to a pair of vertically arranged inputs on the left of the visual field. It has learned to excite the right-direction cell that directs movement away from food pair.

https://doi.org/10.1371/journal.pcbi.1005705.g007

input neurons (see S4 Fig) and this was used in the rest of the study. Interestingly, for complex tasks increase of connectivity beyond this limit led to some decrease of performance.

The output layer received all to all connections from the middle layer. Recent STDP traces were rewarded when the virtual agent acquired horizontally arranged food pairs and were punished when the agent acquired vertically arranged food pairs. Thus,the network learned to move toward horizontal (rewarded) pairs and to avoid the vertical (punished) pairs over the course of training.

Characteristic structure of synaptic connectivity after training was formed for the simpler version of a foraging task, where for a given middle layer cell the input from specific input layer cell would become prominent and its output increased to the corresponding quadrant of the "decision" layer, pursuing approaching behavior (see S5 Fig in supplementary information).

Performance. Fig 8A shows performance of the model for a complex discrimination task. In order to let enough "freedom of movement" to the virtual agent, the total density of rewarded food particles in the new environment was 25% compared to the base model where no compound shapes were used (see Fig 2). This implies that the value of the optimal performance of an ideal agent for complex discrimination task should be approximately one fourth of that in the base model (performance is defined as a rate of obtained food). Accuracy in discriminating between the two food arrangements was often above 80% with high rate of food acquisition (Fig 8B). When high performance levels were obtained, the network developed in

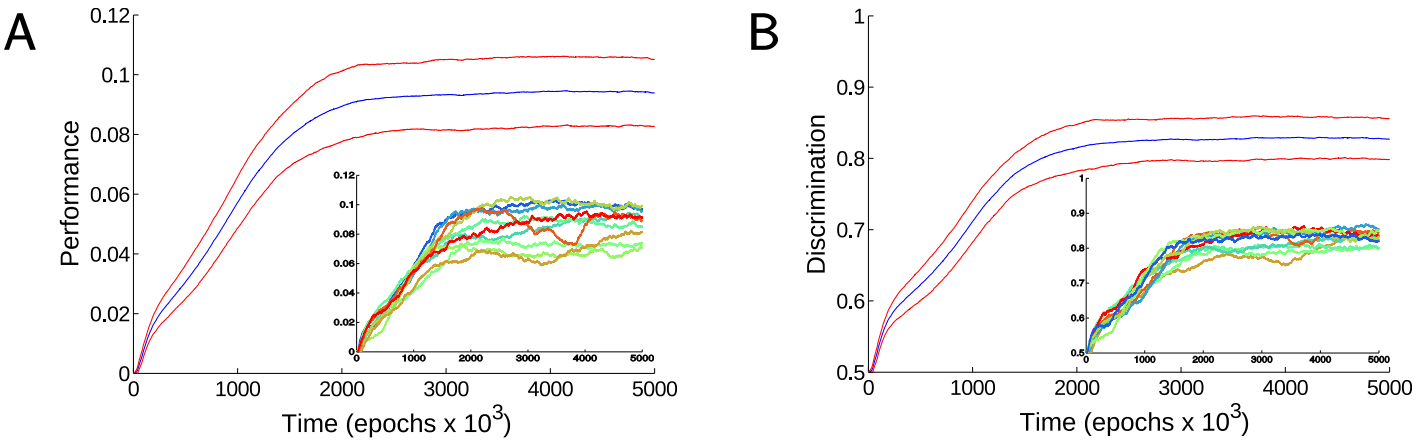

Fig 8. Model performance for complex discrimination task. A) Performance for a task requiring the discrimination of stimuli between vertical and horizontal food pairs. Blue line represents the mean performance and red lines represent the mean ± the standard deviation. Note that asymptotic performance was lower than that for the base model (e.g., Fig 2A) because of the lower overall food density (see Methods) and avoidance of the "bad" food. Inset: The performance of 10 typical trials. B) Average rate of rewarded ("good") food acquired as a percentage of the total rate of food acquired for the same set of trials as in A. Inset shows the same measure for 10 sample trials.

https://doi.org/10.1371/journal.pcbi.1005705.g008

such a way that the majority of middle layer cells had most of their synaptic inputs originating from different pairs of input layer cells (see Fig 7B and 7C).

Our initial analysis revealed that for a fraction of trials after a short period of learning performance could drop to zero and then it would stay very low. We found that for reliable learning and avoiding this problem, the punishment associated with "bad" food (vertical pattern) had to be lower than the reward associated with acquisition of "good" (horizontal pattern) food. When the net sum of reward and punishment was near zero, the "crash" in performance, as described above, could often follow. When the magnitude of reward was set to be twice as large as punishment, a reliable learning with comparable learning times was observed in all trials.

Discussion

We applied a neuronal network of spiking neurons with synaptic plasticity to model a simple foraging task where the virtual agent was required to navigate toward areas of high food concentration. Previously we found that a model with a single plastic layer was able to accomplish this task [30]. In this new study the network employing two plastic layers, combining unrewarded (unsupervised learning) and rewarded (reinforcement learning) spike time dependent plasticity (STDP) was capable of pattern discrimination and also achieved higher performance.

The model simulated a virtual agent moving through a 50 by 50 grid world with "food" particles randomly distributed throughout it. An input layer of the network (7 by 7) received excitation to the neurons representing locations of the food particles in the "visual field"; an output layer (3 by 3) controlled the virtual agent direction of movement. The middle (hidden) layer received plastic (STDP) connections from the input layer and projected plastic (rewarded STDP) connections to the output layer. Multiple homeostatic mechanisms were implemented to maintain synaptic homeostasis. No labeled training data set was necessary or implemented. Over the course of continuous learning, the network developed the ability to move toward concentrations of food and learned to discriminate between "good" food and "bad" food based on the spatial arrangement of the food particles. The noise in the system promoted trial-and-

error learning and helped to select new efficient problem solving strategies. We found that a combination of rewarded and non-rewarded STDP within multiple layers of plasticity was necessary to successfully perform this complex foraging task.

Using STDP as a plasticity rule

While STDP has been previously applied in the network models to explain pattern recognition [36], and rewarded STDP has been shown to be capable of overcoming the distal reward problem [12, 37, 38, 39, 16, 17], our study combined both types of synaptic plasticity in a single model to achieve high learning performance in discrimination and decision making tasks. By punishing the network when incorrect decision (wrong type of food is acquired) was made, the network was required (a) to identify the input pattern (food type) by recognizing the arrangement of the food participles and (b) to learn to move toward the rewarded pattern (horizontal arrangement of food particles) and away from the punished pattern (vertical arrangement of food particles).

It has been shown that STDP is most relevant to cells firing in the mid range frequencies, while other rules, depending primarily on the presynaptic firing rate, may become important for the networks operating in the high frequency range [40, 41, 42]. Combining both rate-based and time-based plasticities gives rise to more complex network connectivity patterns [43, 44] which could potentially be used for more complex tasks. Our model does not employ the rate-based plasticity but future extensions of our model may incorporate a combination of these complementary forms of synaptic learning.

Synaptic balancing

The idea of synaptic strength normalization was formulated shortly after Hebb's seminal work [45]. It was shown that for a successful Hebbian cell assembly to develop, the total sum of input synaptic weights has to be held near constant in order to prevent runaway dynamics [46, 47]. This leads to a strong need for extensive homeostatic regulation to prevent pathological synchronized activity and indeed there is evidence that homeostatic mechanisms are deeply involved in both preventing and causing epileptic conditions [48, 49, 50, 51]. Such normalization rules became an integral part of modeling studies of supervised learning and cortical map formation [52].

In our model, maintaining the balance of synaptic strength within the network was vital to learning and to achieving high model performance. However, not all synaptic rules contributed equally to the final performance of the model. By far the most important was heterosynaptic plasticity which reduces (increases) the strength of all other synapses when one synapse increases (reduces) in strength by homosynaptic plasticity. Such a tendency towards maintaining the net input to a neuron has already been found in experimental studies [53, 54, 26, 27] and has been tested in detailed biophysical models [25, 28]. Its potential mechanism may depend on the intracellular Ca^{2+} dynamics which is influenced by the backpropagating action potentials [55, 56] at the synapses which do not actively participate in homosynaptic plasticity induction [24, 23, 22, 57, 21].

We also found that the model could only operate efficiently when the net network activity remained within a certain range. If the network became overactive, high levels of activity in the output layer made direction selectivity highly inconsistent. If the network activity was too low, not enough input would be received by the output layer to cause spiking, leading to poor performance as well. Control of the long term average network activity can be achieved by homeostatic scaling [58, 59], which adjusted the strength of synaptic inputs up for the neurons with low average firing rates and down for the neurons with high average activity. Whenever

there was variation in food density near the agent, it was important to maintain comparable firing rates in the output neurons. Keeping a functional range of this activity from one trial to the next required that the network activation did not increase in direct proportion to the input activity.

As the highly active and more consistently rewarded synapses regularly increased in strength, it was possible for certain neurons to gain a major control over the network dynamics. Although it was not necessary that all neurons be equally represented, having a small number of neurons exert much greater control over the network had a pronounced negative effect on performance. Output-side synaptic balancing allowed us to reduce inequalities in the representation of the neurons and created competition between the outputs of a neuron as was proposed in [30]. This was particularly useful when a middle-layer neuron had two (or more) strong connections to the output neurons, thus "supporting" two different directions of movement. If such middle-layer neuron received a reward, it would increase strength of both connections—despite the fact that only one of them was actually responsible for correct food acquisition. Output-side balancing allowed only synapses correlated with reward to be strengthened via rewarded STDP traces. While we do not know direct experimental evidence for this mechanism, a similar effect can be potentially achieved by more complex circuits involving reciprocal inhibition between output neurons.

We observed that by normalizing the amount of potentiaton in STDP (based on the history of STDP traces occurring in the synapse past), we increased the speed of initial learning and increased the stability of efficiency gains for the network in the long term. Though not directly observed, this mechanism is consistent with experimental data [32, 33, 34] and might be of interest for additional experimental scrutiny.

Inhibition

Interaction between synaptic excitation and inhibition is a ubiquitous phenomenon found across many different brain circuits [60, 61, 62, 63, 64, 65, 66, 67]; it can be observed during spontaneous activity of the brain [68, 69, 70] and is an integral part of the oscillatory mechanisms in the neuronal networks [71]. To simplify implementation of the inhibitory effects in our model, the neurons of the preceding layer have sent inhibitory connections to the same set of post synaptic cells they excited in the following layer. This essentially provided feedforward inhibition between layers of neurons, using a population of virtual inhibitory neurons that was not explicitly modeled. The total strength of inhibitory connections from any given cell was the same as the total of its excitatory connections. While not biologically realistic, this simplification effectively simulated a feed forward inhibition found in various brain areas [72] where neurons in one layer project to both excitatory and inhibitory neurons in the down stream layer and inhibitory neurons in turn project to excitatory neurons in the same layer [73]. As inhibitory neurons are often more electrically compact [74] and have a faster membrane time constant, this can result in the inhibition arriving to the post synaptic excitatory neurons with only a small delay compared to the excitation. No significant difference was seen in performance of this network as a result of this simplified implementation of inhibition when compared to the previous model where inhibition was implemented explicitly [30].

Noise

Noise is another ubiquitous phenomenon in the neural networks and spans multiple time scales and different physical domains [75]. In contrast with artificially crafted devices where noise is usually detrimental to functioning, biological neural networks are capable of using noise in a productive way to improve efficiency and performance [76, 77]. Examples of this

include enhanced input detection by means of stochastic resonance [78, 79, 80, 81] and probabilistic inference [82, 83]; noise can help with performance of the associative memories [84], smoothing the threshold for action potentials [85, 86], allowing rapid response time for neuronal populations [87, 88] or faithful propagation of firing rates across a layered network [89].

The largest source of the noise in any biological neuronal network is synaptic noise [90]. In our model we employed noise at the level of synaptic currents that represented variability in synaptic vesicle release. Noise was not only tolerated by this model but it was required for its function. Before training there was no meaningful mapping between the input received by the network and its output; the output activity (and therefore direction of movement) was essentially driven by random initial configuration of synaptic weights. Variability (noise) in synaptic release occasionally caused an unexpected output for a given input. When the network output resulted in food acquisition, it was reinforced by a reward. If a synapse that was likely to be associated with a reward was able to surpass the strength of another synapse of the same presynaptic cell, noise in the synaptic output created instances where the "correct" output cell fired and the "incorrect" output cell did not, despite the "incorrect" output cell having a stronger incoming synaptic strength. This effectively allowed for a form of trial and error learning. Without such variety of the actions attempted, it would be difficult to find which goal-oriented strategies are effective and it can be seen as representing a strategy of associating a given stimulus with a given response.

Conclusions

Unlike common machine learning approaches or the currently booming deep network architectures [91] which, in some cases, have already outperformed humans [92, 93, 94, 95, 96], in this study we only applied biologically plausible and experimentally identified plasticity rules. This new study, following our previous work [30], generated specific predictions regarding what biological learning mechanisms are necessary and/or sufficient to accomplish the learning task. The key components for a successfully working model were STDP, synaptic balancing processes keeping the learning stable, dopamine type reward feedback derived from successful outcome of action and noise at the synaptic level allowing trial-and-error learning. Although these components have already been studied, sometimes in great detail, our study links them together (unsupervised STDP for sensory learning and rewarded STDP for decision making in particular) in a new way and shows advantages of interaction between different plasticity mechanisms for complex task solving.

Methods

Environment

Foraging behavior took place in a virtual environment of randomly distributed "food" particles. The environment consists of 50 by 50 grid of locations. Initially, each location was either assigned or not a food particle. When a food particle was acquired as a result of the virtual agent move, it was then assigned randomly to a new spot on the map. This resulted in a continuously changing environment with the same food density. The density of food particles in the environment was set to 10%. The virtual agent was seeing a 7 by 7 grid of squares the ("visual field") centered on its current location and it could move to any adjacent square including diagonally for a total of 8 directions.

Network structure

The network was composed of 842 spiking map based neurons [97, 98], arranged into 3 feed forward layers to mimic a basic biological circuit: a 7 by 7 input layer (I), a 28 by 28 middle (hidden, H) layer, and a 3 by 3 output layer (O) (Fig 1). This structure provides a basic feedforward inhibitory circuit [73] found in many biological structures [99, 73, 100, 101, 102, 103].

Each cell in the middle layer received synaptic inputs from 9 random cells in the input layer. These connections initially had random strengths drawn from a normal distribution. Each cell in the excitatory middle layer (cell H_i) connected to every cell in the output layer (O_j) with synaptic strength W_{ij} or WI_{ij}, respectively. Initially all these connections had uniform strengths and the responses in the output layer were due to the random synaptic variability. Random variability was a property of all synaptic interactions between neurons and it was implemented as variability in the magnitude of the individual synaptic events.

Movement cycle

Simulation time was divided up into epochs of 600 time steps, each roughly equivalent to 300 ms. At the start of each epoch the virtual agent received input corresponding to locations of nearby food within the input (7x7) layer. Thus 48 of the 49 cells receive input from a unique position relative to the virtual agent location. At the end of the epoch the virtual agent made one move based on the activity of the output layer. If the virtual agent moved to a grid square with "food", the "food" was removed from that square and assigned to a randomly selected new square.

Each epoch was of sufficient duration for the network to receive inputs, produce outputs, and return to a resting state. Input layer neurons representing positions in the environment containing food received brief pulse of excitatory stimulation that triggered a spike; this stimulation was evoked at the start of each movement cycle (epoch). Output was chosen and the virtual agent moved at the end of the epoch.

The activity of the output layer of the network controlled direction of virtual agent's movement. Each of the output layer cells was mapped to a specific direction. The output layer cell (O_j) that spiked the greatest number of times during the first half of an epoch defined the direction of movement on that epoch. If there was a tie, direction was chosen randomly from the tied outputs. If no cells in the output layer fired, the virtual agent continued in the direction it traveled during the previous epoch.

There was 1% chance on every move that the virtual agent would ignore any output and instead move in a random direction. This random variability prevented infinite loops of virtual agent's motion during the learning process. Synaptic noise was not sufficient to break out of all movement loops as some loops were the result of forming strong connections that would mediate the same spiking pattern regardless of the noise. Other times the probability of escape from a loop due to the noise was simply low enough that it would take a significant amount of time to break the loop. While biological systems could utilize different mechanisms to achieve the same goal, the method we implemented was efficient and accomplished the goal.

In the model where no pattern recognition was required, the chance of random movement started at 0.5% but increased by 0.5% for each move in which no food was obtained. This value was reset to its starting value whenever food was obtained.

For more complex task where pattern recognition was also required, high performing solutions to the task included both approach and avoidance behavior. As such, the network was far more susceptible to becoming stuck in a movement cycle, usually when the virtual agent became surrounded by "toxic" food. In order to break these cycles the rate of random motion was gradually increased by 1% per move in which the agent did not obtain food.

Synaptic plasticity

Synaptic plasticity closely followed the rules introduced in [30]. A rewarded STDP paradigm [12, 37, 38, 104] was implemented between layers H and O and a non-reward modulated STDP paradigm between I and H. A spike in a post-synaptic cell (O_j of the output layer) that directly followed a spike in a pre-synaptic cell (H_i of the hidden layer) created a "*pre before post*" event. Each new post synaptic cell spike was compared to all pre synaptic spikes within the time window and each new pre synaptic spike was compared to all postsynaptic spikes within the window.

The value of an STDP event (trace) was calculated using the following equation [10, 105]:

$$\begin{aligned} p &= \frac{-|t_r - t_p|}{T_c}, \\ tr_k &= Ke^p \end{aligned} \tag{1}$$

where t_r and t_p are the times at which the pre and post synaptic spiking events occurred respectively, T_c is the time constant and is equal to 40 ms. K is equal to —0.04 in the case of a *post before pre* event and 0.04 in the case of a *pre before post* event.

STDP event was immediately applied to the respective synapse W_{ij} between neurons I_i and H_j. In contrast, for synapses between neurons H_i and O_j the events were stored as traces for later use. Each trace remained stored for 6 epochs after its creation and then was erased. While still stored, STDP trace had an effect whenever there was a rewarding or punishing event. If the network was rewarded or punished the new synaptic strength of the synapse W_{ij} was described as:

$$\begin{aligned} W_{ij}(n+1) &= W_{ij}(n)\prod_k^{traces}\left(1 + \frac{W_{i0}}{W_i} * \Delta_k\right), \\ \Delta_k &= S_{rp} \cdot \frac{tr_k}{t - t_k + c} \cdot \frac{Sum_{tr}(n+1)}{Avg_{tr}(n+1)}, \\ Sum_{tr}(n+1) &= \sum_k^{traces} \frac{tr_k}{t - t_k + c}, \\ Avg_{tr}(n+1) &= Avg_{tr}(n)(1-\delta) + \delta Sum_{tr}(n+1), \end{aligned} \tag{2}$$

where t is current time step, S_{rp} is a scaling factor for reward/punishment, tr_k is magnitude of trace (defined in Eq (2)), t_k is time of the trace event, c is a constant (=1 epoch) used for decreasing sensitivity to very recent spikes, $W_i = \sum_j W_{ij}$ is a total synaptic strength of all connections from specific cell H_i to all cells O_j of the output layer, W_{i0} is a constant that is set to the value of W_i at the at the beginning of the simulation ("target value"). The term W_{i0}/W_i helped to keep the output weight sum close to the initial target value. The values for Avg_{tr} and Sum_{tr} were almost always positive in our simulations due to the feed forward architecture that we used. We should note, that for a more general model with feedback loops it would be more appropriate to handle negative and positive traces separately.

The network was rewarded when the virtual agent moved to a "food" location and $S_{rp} = 1$. In case of pattern recognition model it was punished when it moved to a location with a toxic food and $S_{rp} = -0.5$. There was also smaller punishment applied when no food is obtained, $S_{rp} = -0.1$ in the base model and $S_{rp} = -0.01$ in the pattern recognition model. The effect of

these rules was that the cells with lower total output strength increased their output strength more easily.

To ensure that all the output neurons maintained a relatively constant long term firing rate, the model incorporated homeostatic synaptic scaling, which takes place every epoch (= 600 time steps). The total strength of synaptic inputs $W_j = \Sigma_i W_{ij}$ to a given output cell O_j was set to be equal at each time step to the target synaptic input $W_j = W_{j0}$—a slow variable that varied over many epochs and depended on the activity of the cell O_j and activity of its all pre-synaptic cells. If a cell O_j consistently fired below the target rate, the W_{j0} was increased by $D_{tar} = 0.0001$. If the cell responded above its target firing rate the W_{j0} was gradually reduced:

$$W_{j(n+1)} = \begin{cases} W_{j(n)} * (1 + D_{tar}) & \text{spike rate} < \text{target rate} \\ W_{j(n)} * (1 - D_{tar}) & \text{spike rate} > \text{target rate} \end{cases} \quad (3)$$

To ensure that the net synaptic input W_j to any neuron was unaffected by plasticity events of the individual connections at the individual time steps and equal to W_{j0}, we implemented scaling process that occurs after each STDP event. When any excitatory connection increased in strength, all the other excitatory connections incoming to that cell decreased in strength by a "*scale factor*" S_f to keep $W_j = W_{j0}$:

$$W_{ij(n+1)} = W_{ijn} S_f$$

$$S_f = \frac{W_{j0}}{\sum_i W_{ijn}} \quad (4)$$

W_{ijn} are synaptic weights right after STDP event but before scaling, $W_{ij(n+1)}$ are synaptic weights after scaling, W_{j0} is W_{i0} from Eq (2).

Map based neuronal models

The underlying reduced model of fast spiking neuron was identical to the model used in [30] and can be described by the following set of difference equations [97, 106, 107]:

$$V_{n+1} = f_\alpha(V_n, I_n + \beta_n), \quad (5)$$

$$I_{n+1} = I_n - \mu(V_n + 1) + \mu\sigma + \mu\sigma_n,$$

where V_n is the membrane voltage, I_n is a slow dynamical variable describing the effects of slow conductances, and n is a discrete time step (~ 0.5 msec). Slow temporal evolution of I_n was achieved by using small values of the parameter $\mu \ll 1$. Input variables β_n and σ_n were used to incorporate external current I_n^{ext} (e.g., synaptic input): $\beta_n = \beta^e I_n^{ext}$, $\sigma_n = \sigma^e I_n^{ext}$. Parameter values were set to $\sigma = 0.06$, $\beta^e = 0.133$, $\sigma^e = 1$, $\mu = 0.0005$. The nonlinearity $f_\alpha(V, I)$ was designed in the form of a piece-wise continuous function:

$$f_\alpha(V_{n,I_n}) = \begin{cases} \alpha(1 - V_n)^{-1} + I_n, & V_n \le 0 \\ \alpha + I_n, & 0 < V_n < \alpha + I_n \;\&\; V_{n-1} \le 0 \\ -1, & \alpha + I_n \le V_n \text{ or } V_{n-1} > 0 \end{cases}, \quad (6)$$

where $\alpha = 3.65$. To convert the dimensionless "membrane potential" V to the physiological membrane potential V_{ph}, the following equation was applied: $V_{ph} = 50V - 15$ [mV] [98].

This model is very computationally efficient and, despite its intrinsic low dimensionality, produces a rich repertoire of dynamics capable to mimic the dynamics of the Hodgkin-Huxley type neurons both at the single cell level and in the context of network dynamics [97, 107]. A fast spiking neuron model was chosen to simulate the neurons.

To model synaptic interconnections, we used conventional first order kinetic models of synaptic conductances rewritten in the form of difference equations:

$$g^{syn}_{(n+1)} = \gamma g^{syn}_{n} + \begin{cases} (1 + XR) g_{syn} / W_j, & spike_{pre}, \\ 0, & otherwise, \end{cases} \quad (7)$$

and the synaptic current computed as: $I^{syn}_{n} = -g^{syn}_{n}(V^{post}_{n} - V_{rp})$.

Here g_{syn} is the strength of synaptic coupling, modulated by the target rate W_j of receiving cell j, indices *pre* and *post* stand for the presynaptic and postsynaptic variables, respectively. The first condition, "$spike_{pre}$", is satisfied when presynaptic spikes are generated. Parameter γ controls the relaxation rate of synaptic conductance after a presynaptic spike is received ($0 \leq \gamma < 1$). The parameter R is the coefficient of variability in synaptic release. The standard value of R is 0.12. X is a randomly generated number between -1 and 1. Parameter V_{rp} defines the reversal potential and, therefore, the type of synapse: excitatory or inhibitory. The term (1+XR) introduces a variability in synaptic release such that the effect of any synaptic interaction has an amplitude that is pulled from a flat distribution ranging from 1+R to 1-R times the average value of the synapse.

Supporting information

S1 Fig. Average synaptic strengths of the input layer neurons. Input layer neurons ("visual field") were divided into 3 groups based on the distance from the center (agent position). For each group, we measured the average total strength of the connections to the middle layer cells. Left: baseline model; Right: output balancing disabled. X-axis is a distance from the center. Y-axis is the average connection strength, averaged over 20 independent trials. Output balancing helped to keep the average synaptic strengths for all three groups of neurons in the same range, while the network without output balancing developed large differences between the groups. Even distribution of the outputs helped the model to learn equally the information about distant and nearby food, yielding better results in the overall performance as shown in the Fig 3.
(PDF)

S2 Fig. Effect of the output layer neurons' excitability on the outcome of the decision making. Left: number of non-zero ties between output neurons, middle: number of the epochs with no response of the output neurons (all cells in the output layer remained silent during epoch), right: both ties and zeros counted together. Each dot is average of 10 independent trials. The mean output layer firing rate was considered as a measure of excitability; 1. 6 Hz was the default output firing rate. Note, that the number of non-zero ties was small and the number of epochs with no response was high for the low excitability (< 1 Hz), because the excitability was too low to reach the spiking threshold and the output layer commonly remained silent. Decreasing number of non-zero ties for high excitability (>10Hz) was observed because the likelihood of exact tight became low as the number of spikes generated by the output neurons increased.
(PDF)

S3 Fig. Effect of inhibition on the output layer activity. Left, the histogram shows baseline activity (inhibition enabled). X-axis gives total number of spikes in the output layer during

single epochs; Y-axis—number of epochs for each class of firing (distribution). Next two histograms show the network activity when no input->middle layer or no middle->output inhibition, respectively, was implemented. All data are from 10 independent trials for each scenario.
(PDF)

S4 Fig. The role of fan-in to the middle layer. Performance of the model with respect to the varying fan-in from the input to the middle layer cells. Each point is a final performance averaged from 10 independent trials, each trial running for 2.10^4 epochs. Left: Simple task. From fan-in > 4, the model reached performance close to the optimum level. Right: Complex task. The learning performance was almost zero for fan-in < 4, gradually improving until it peaked around 8-9 and then decreasing slowly for even higher fan-in connectivity.
(PDF)

S5 Fig. Structure of synaptic connectivity after learning a simple foraging task. Strengths of the synaptic inputs (left) and outputs (right) of a typical middle layer neuron after successful training. The network has been trained to move toward any (single) food particle. Red represents the highest synaptic strength while blue represents the lowest strength. A characteristic middle layer cell that became responsive to a single food particle in the top right of the visual field (left) and excited the top right output cell (right) which moved the virtual agent toward the food particle.
(PDF)

Author Contributions

Conceptualization: Pavel Sanda, Steven Skorheim, Maxim Bazhenov.

Data curation: Pavel Sanda.

Formal analysis: Pavel Sanda, Steven Skorheim.

Funding acquisition: Maxim Bazhenov.

Investigation: Pavel Sanda, Steven Skorheim.

Methodology: Pavel Sanda, Steven Skorheim, Maxim Bazhenov.

Project administration: Maxim Bazhenov.

Resources: Maxim Bazhenov.

Software: Pavel Sanda, Steven Skorheim.

Supervision: Maxim Bazhenov.

Validation: Pavel Sanda, Steven Skorheim.

Visualization: Pavel Sanda, Steven Skorheim.

Writing – original draft: Pavel Sanda, Steven Skorheim.

Writing – review & editing: Pavel Sanda, Steven Skorheim, Maxim Bazhenov.

References

1. Watanabe T, Sasaki Y. Perceptual learning: toward a comprehensive theory. Annu Rev Psychol. 2015; 66:197–221. https://doi.org/10.1146/annurev-psych-010814-015214 PMID: 25251494
2. Platt ML, Glimcher PW. Neural correlates of decision variables in parietal cortex. Nature. 1999; 400(6741):233–238. https://doi.org/10.1038/22268 PMID: 10421364

3. Schultz W. Neuronal reward and decision signals: from theories to data. Physiol Rev. 2015; 95(3): 853–951. https://doi.org/10.1152/physrev.00023.2014 PMID: 26109341
4. Law CT, Gold JI. Reinforcement learning can account for associative and perceptual learning on a visual-decision task. Nat Neurosci. 2009; 12(5):655–663. https://doi.org/10.1038/nn.2304 PMID: 19377473
5. Frankó E, Seitz AR, Vogels R. Dissociable neural effects of long-term stimulus–reward pairing in macaque visual cortex. J Cogn Neurosci. 2010; 22(7):1425–1439. https://doi.org/10.1162/jocn.2009.21288 PMID: 19580385
6. Seitz AR, Watanabe T. Psychophysics: Is subliminal learning really passive? Nature. 2003; 422(6927):36–36. https://doi.org/10.1038/422036a PMID: 12621425
7. Yao H, Dan Y. Stimulus timing-dependent plasticity in cortical processing of orientation. Neuron. 2001; 32(2):315–323. https://doi.org/10.1016/S0896-6273(01)00460-3 PMID: 11684000
8. Li N, DiCarlo JJ. Unsupervised natural experience rapidly alters invariant object representation in visual cortex. Science. 2008; 321(5895):1502–1507. https://doi.org/10.1126/science.1160028 PMID: 18787171
9. Brown TH, Kairiss EW, Keenan CL. Hebbian synapses: biophysical mechanisms and algorithms. Annu Rev Neurosci. 1990; 13(1):475–511. https://doi.org/10.1146/annurev.ne.13.030190.002355 PMID: 2183685
10. Bi Gq, Poo Mm. Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. J Neurosci. 1998; 18(24):10464–10472. PMID: 9852584
11. Frey U, Morris RG. Synaptic tagging and long-term potentiation. Nature. 1997; 385(6616):533–536. https://doi.org/10.1038/385533a0 PMID: 9020359
12. Izhikevich EM. Solving the distal reward problem through linkage of STDP and dopamine signaling. Cereb Cortex. 2007; 17(10):2443–2452. https://doi.org/10.1093/cercor/bhl152 PMID: 17220510
13. Sajikumar S, Frey JU. Late-associativity, synaptic tagging, and the role of dopamine during LTP and LTD. Neurobiol Learn Mem. 2004; 82(1):12–25. https://doi.org/10.1016/j.nlm.2004.03.003 PMID: 15183167
14. Seamans JK, Yang CR. The principal features and mechanisms of dopamine modulation in the prefrontal cortex. Prog Neurobiol. 2004; 74(1):1–57. https://doi.org/10.1016/j.pneurobio.2004.05.006 PMID: 15381316
15. Nitz DA, Kargo WJ, Fleischer J. Dopamine signaling and the distal reward problem. NeuroReport. 2007; 18(17):1833–1836. https://doi.org/10.1097/WNR.0b013e3282f16d86 PMID: 18090321
16. Chadderdon GL, Neymotin SA, Kerr CC, Lytton WW. Reinforcement learning of targeted movement in a spiking neuronal model of motor cortex. PLoS One. 2012; 7(10):e47251. https://doi.org/10.1371/journal.pone.0047251 PMID: 23094042
17. Neymotin SA, Chadderdon GL, Kerr CC, Francis JT, Lytton WW. Reinforcement learning of two-joint virtual arm reaching in a computer model of sensorimotor cortex. Neural Comput. 2013; 25(12):3263–3293. https://doi.org/10.1162/NECO_a_00521 PMID: 24047323
18. Abbott LF, Nelson SB. Synaptic plasticity: taming the beast. Nat Neurosci. 2000; 3:1178–1183. PMID: 11127835
19. Watt AJ, Desai NS. Homeostatic plasticity and STDP: keeping a neuron's cool in a fluctuating world. Front Synaptic Neurosci. 2010; 2(5):1–16.
20. Turrigiano G. Homeostatic synaptic plasticity: local and global mechanisms for stabilizing neuronal function. Cold Spring Harb Perspect Biol. 2012; 4(1):a005736. https://doi.org/10.1101/cshperspect.a005736 PMID: 22086977
21. Schuman EM, Madison DV. Locally distributed synaptic potentiation in the hippocampus. Science. 1994; 263(5146):532–536. https://doi.org/10.1126/science.8290963 PMID: 8290963
22. Kossel A, Bonhoeffer T, Bolz J. Non-Hebbian synapses in rat visual cortex. NeuroReport. 1990; 1(2): 115–118. https://doi.org/10.1097/00001756-199010000-00008 PMID: 2129865
23. Bonhoeffer T, Staiger V, Aertsen A. Synaptic plasticity in rat hippocampal slice cultures: local "Hebbian" conjunction of pre-and postsynaptic stimulation leads to distributed synaptic enhancement. Proc Natl Acad Sci USA. 1989; 86(20):8113–8117. https://doi.org/10.1073/pnas.86.20.8113 PMID: 2813381
24. Lynch GS, Dunwiddie T, Gribkoff V. Heterosynaptic depression: a postsynaptic correlate of long-term potentiation. Nature. 1977; 266(5604):737–739. https://doi.org/10.1038/266737a0 PMID: 195211
25. Chen JY, Lonjers P, Lee C, Chistiakova M, Volgushev M, Bazhenov M. Heterosynaptic plasticity prevents runaway synaptic dynamics. J Neurosci. 2013; 33(40):15915–15929. https://doi.org/10.1523/JNEUROSCI.5088-12.2013 PMID: 24089497

26. Chistiakova M, Bannon NM, Bazhenov M, Volgushev M. Heterosynaptic plasticity multiple mechanisms and multiple roles. Neuroscientist. 2014; 20(5):483–498. https://doi.org/10.1177/1073858414529829 PMID: 24727248

27. Chistiakova M, Bannon NM, Chen JY, Bazhenov M, Volgushev M. Homeostatic role of heterosynaptic plasticity: models and experiments. Front Comput Neurosci. 2015; 9(89):1–22.

28. Volgushev M, Chen JY, Ilin V, Goz R, Chistiakova M, Bazhenov M. Partial Breakdown of Input Specificity of STDP at Individual Synapses Promotes New Learning. J Neurosci. 2016; 36(34):8842–8855. https://doi.org/10.1523/JNEUROSCI.0552-16.2016 PMID: 27559167

29. Law CT, Gold JI. Neural correlates of perceptual learning in a sensory-motor, but not a sensory, cortical area. Nat Neurosci. 2008; 11(4):505–513. https://doi.org/10.1038/nn2070 PMID: 18327253

30. Skorheim S, Lonjers P, Bazhenov M. A spiking network model of decision making employing rewarded STDP. PLoS One. 2014; 9(3):e90821. https://doi.org/10.1371/journal.pone.0090821 PMID: 24632858

31. Shepherd Gordon M. The Synaptic Organization of the Brain. Oxford University Press; 2004.

32. van Rossum MC, Bi GQ, Turrigiano GG. Stable Hebbian learning from spike timing-dependent plasticity. J Neurosci. 2000; 20(23):8812–8821. PMID: 11102489

33. Froemke RC, Dan Y. Spike-timing-dependent synaptic modification induced by natural spike trains. Nature. 2002; 416(6879):433–438. https://doi.org/10.1038/416433a PMID: 11919633

34. Guyonneau R, VanRullen R, Thorpe SJ. Neurons tune to the earliest spikes through STDP. Neural Comput. 2005; 17(4):859–879. https://doi.org/10.1162/0899766053429390 PMID: 15829092

35. Poggio T, Bizzi E. Generalization in vision and motor control. Nature. 2004; 431(7010):768–774. https://doi.org/10.1038/nature03014 PMID: 15483597

36. Masquelier T, Guyonneau R, Thorpe SJ. Competitive STDP-based spike pattern learning. Neural Comput. 2009; 21(5):1259–1276. https://doi.org/10.1162/neco.2008.06-08-804 PMID: 19718815

37. Farries MA, Fairhall AL. Reinforcement learning with modulated spike timing–dependent synaptic plasticity. J Neurophysiol. 2007; 98(6):3648–3665. https://doi.org/10.1152/jn.00364.2007 PMID: 17928565

38. Florian RV. Reinforcement learning through modulation of spike-timing-dependent synaptic plasticity. Neural Comput. 2007; 19(6):1468–1502. https://doi.org/10.1162/neco.2007.19.6.1468 PMID: 17444757

39. Vasilaki E, Frémaux N, Urbanczik R, Senn W, Gerstner W. Spike-based reinforcement learning in continuous state and action space: when policy gradient methods fail. PLoS Comput Biol. 2009; 5(12): e1000586. https://doi.org/10.1371/journal.pcbi.1000586 PMID: 19997492

40. Sjöström PJ, Turrigiano GG, Nelson SB. Rate, timing, and cooperativity jointly determine cortical synaptic plasticity. Neuron. 2001; 32(6):1149–1164. https://doi.org/10.1016/S0896-6273(01)00542-6 PMID: 11754844

41. Froemke RC, Tsay IA, Raad M, Long JD, Dan Y. Contribution of individual spikes in burst-induced long-term synaptic modification. J Neurophysiol. 2006; 95(3):1620–1629. https://doi.org/10.1152/jn.00910.2005 PMID: 16319206

42. Butts DA, Kanold PO, Shatz CJ. A burst-based "Hebbian" learning rule at retinogeniculate synapses links retinal waves to activity-dependent refinement. PLoS Biol. 2007; 5(3):e61. https://doi.org/10.1371/journal.pbio.0050061 PMID: 17341130

43. Clopath C, Büsing L, Vasilaki E, Gerstner W. Connectivity reflects coding: a model of voltage-based STDP with homeostasis. Nat Neurosci. 2010; 13(3):344–352. https://doi.org/10.1038/nn.2479 PMID: 20098420

44. Clopath C, Gerstner W. Voltage and Spike Timing Interact in STDP–A Unified Model. Front Synaptic Neurosci. 2010; 2(25):1–11.

45. Hebb D. The Organization of Behavior: A Neuropsychological Theory. Wiley, New York; 1949.

46. Rochester N, Holland J, Haibt L, Duda W. Tests on a cell assembly theory of the action of the brain, using a large digital computer. IRE Trans Info Theory. 1956; 2(3):80–93. https://doi.org/10.1109/TIT.1956.1056810

47. Zenke F, Gerstner W. Hebbian plasticity requires compensatory processes on multiple timescales. Phil Trans R Soc B. 2017; 372(1715):20160259. https://doi.org/10.1098/rstb.2016.0259 PMID: 28093557

48. Houweling AR, Bazhenov M, Timofeev I, Steriade M, Sejnowski TJ. Homeostatic synaptic plasticity can explain post-traumatic epileptogenesis in chronically isolated neocortex. Cereb Cortex. 2005; 15(6):834–845. https://doi.org/10.1093/cercor/bhh184 PMID: 15483049

49. Bazhenov M, Houweling AR, Timofeev I, Sejnowski TJ. Homeostatic Plasticity and Post-Traumatic Epileptogenesis. In: Soltesz I, Staley K, editors. Computational Neuroscience in Epilepsy. Academic Press; 2008. p. 259.

50. Volman V, Sejnowski TJ, Bazhenov M. Topological basis of epileptogenesis in a model of severe cortical trauma. J Neurophysiol. 2011; 106(4):1933–1942. https://doi.org/10.1152/jn.00458.2011 PMID: 21775725

51. González OC, Krishnan GP, Chauvette S, Timofeev I, Sejnowski T, Bazhenov M. Modeling of Age-Dependent Epileptogenesis by Differential Homeostatic Synaptic Scaling. J Neurosci. 2015; 35(39): 13448–13462. https://doi.org/10.1523/JNEUROSCI.5038-14.2015 PMID: 26424890

52. Markram H, Gerstner W, Sjöström P. A history of spike-timing-dependent plasticity. Front Synaptic Neurosci. 2011; 3(4):1–24.

53. Royer S, Paré D. Conservation of total synaptic weight through balanced synaptic depression and potentiation. Nature. 2003; 422(6931):518–522. https://doi.org/10.1038/nature01530 PMID: 12673250

54. Chistiakova M, Volgushev M. Heterosynaptic plasticity in the neocortex. Exp Brain Res. 2009; 199(3-4):377–390. https://doi.org/10.1007/s00221-009-1859-5 PMID: 19499213

55. Yuste R, Denk W, et al. Dendritic spines as basic functional units of neuronal integration. Nature. 1995; 375(6533):682–684. https://doi.org/10.1038/375682a0 PMID: 7791901

56. Schiller J, Schiller Y, Clapham DE. NMDA receptors amplify calcium influx into dendritic spines during associative pre-and postsynaptic activation. Nat Neurosci. 1998; 1(2):114–118. https://doi.org/10.1038/363 PMID: 10195125

57. Engert F, Bonhoeffer T. Synapse specificity of long-term potentiation breaks down at short distances. Nature. 1997; 388(6639):279–284. https://doi.org/10.1038/40870 PMID: 9230437

58. Turrigiano GG, Leslie KR, Desai NS, Rutherford LC, Nelson SB. Activity-dependent scaling of quantal amplitude in neocortical neurons. Nature. 1998; 391(6670):892–896. https://doi.org/10.1038/36103 PMID: 9495341

59. Watson BO, Levenstein D, Greene JP, Gelinas JN, Buzsáki G. Network Homeostasis and State Dynamics of Neocortical Sleep. Neuron. 2016; 90(4):839–852. https://doi.org/10.1016/j.neuron.2016.03.036 PMID: 27133462

60. Tan AY, Zhang LI, Merzenich MM, Schreiner CE. Tone-evoked excitatory and inhibitory synaptic conductances of primary auditory cortex neurons. J Neurophysiol. 2004; 92(1):630–643. https://doi.org/10.1152/jn.01020.2003 PMID: 14999047

61. Wehr M, Zador AM. Balanced inhibition underlies tuning and sharpens spike timing in auditory cortex. Nature. 2003; 426(6965):442–446. https://doi.org/10.1038/nature02116 PMID: 14647382

62. Wu GK, Arbuckle R, Liu Bh, Tao HW, Zhang LI. Lateral sharpening of cortical frequency tuning by approximately balanced inhibition. Neuron. 2008; 58(1):132–143. https://doi.org/10.1016/j.neuron.2008.01.035 PMID: 18400169

63. Mariño J, Schummers J, Lyon DC, Schwabe L, Beck O, Wiesing P, et al. Invariant computations in local cortical networks with balanced excitation and inhibition. Nat Neurosci. 2005; 8(2):194–201. https://doi.org/10.1038/nn1391 PMID: 15665876

64. Assisi C, Stopfer M, Laurent G, Bazhenov M. Adaptive regulation of sparseness by feedforward inhibition. Nat Neurosci. 2007; 10(9):1176–1184. https://doi.org/10.1038/nn1947 PMID: 17660812

65. Poo C, Isaacson JS. Odor representations in olfactory cortex:"sparse" coding, global inhibition, and oscillations. Neuron. 2009; 62(6):850–861. https://doi.org/10.1016/j.neuron.2009.05.022 PMID: 19555653

66. Stopfer M, Bhagavan S, Smith BH, Laurent G. Impaired odour discrimination on desynchronization of odour-encoding neural assemblies. Nature. 1997; 390(6655):70–74. https://doi.org/10.1038/36335 PMID: 9363891

67. Anderson JS, Carandini M, Ferster D. Orientation tuning of input conductance, excitation, and inhibition in cat primary visual cortex. J Neurophysiol. 2000; 84(2):909–926. PMID: 10938316

68. Atallah BV, Scanziani M. Instantaneous modulation of gamma oscillation frequency by balancing excitation with inhibition. Neuron. 2009; 62(4):566–577. https://doi.org/10.1016/j.neuron.2009.04.027 PMID: 19477157

69. Okun M, Lampl I. Instantaneous correlation of excitation and inhibition during ongoing and sensory-evoked activities. Nat Neurosci. 2008; 11(5):535–537. https://doi.org/10.1038/nn.2105 PMID: 18376400

70. Dehghani N, Peyrache A, Telenczuk B, Le Van Quyen M, Halgren E, Cash SS, et al. Dynamic Balance of Excitation and Inhibition in Human and Monkey Neocortex. Sci Rep. 2016; 6(23176).

71. Mann EO, Paulsen O. Role of GABAergic inhibition in hippocampal network oscillations. Trends Neurosci. 2007; 30(7):343–349. https://doi.org/10.1016/j.tins.2007.05.003 PMID: 17532059

72. Isaacson JS, Scanziani M. How inhibition shapes cortical activity. Neuron. 2011; 72(2):231–243. https://doi.org/10.1016/j.neuron.2011.09.027 PMID: 22017986

73. Bazhenov M, Stopfer M. Forward and back: motifs of inhibition in olfactory processing. Neuron. 2010; 67(3):357–358. https://doi.org/10.1016/j.neuron.2010.07.023 PMID: 20696373

74. Shepherd G, Grillner S. Handbook of brain microcircuits. Oxford University Press; 2010.

75. Faisal AA, Selen LP, Wolpert DM. Noise in the nervous system. Nat Rev Neurosci. 2008; 9(4): 292–303. https://doi.org/10.1038/nrn2258 PMID: 18319728

76. Ermentrout GB, Galán RF, Urban NN. Reliability, synchrony and noise. Trends Neurosci. 2008; 31(8):428–434. https://doi.org/10.1016/j.tins.2008.06.002 PMID: 18603311

77. McDonnell MD, Ward LM. The benefits of noise in neural systems: bridging theory and experiment. Nat Rev Neurosci. 2011; 12(7):415–426. https://doi.org/10.1038/nrn3061 PMID: 21685932

78. Longtin A, Bulsara A, Moss F. Time-interval sequences in bistable systems and the noise-induced transmission of information by sensory neurons. Phys Rev Lett. 1991; 67(5):656–659. https://doi.org/10.1103/PhysRevLett.67.656 PMID: 10044954

79. Collins J, Chow CC, Imhoff TT, et al. Stochastic resonance without tuning. Nature. 1995; 376(6537):236–238. https://doi.org/10.1038/376236a0 PMID: 7617033

80. Douglass JK, Wilkens L, Pantazelou E, Moss F, et al. Noise enhancement of information transfer in crayfish mechanoreceptors by stochastic resonance. Nature. 1993; 365(6444):337–340. https://doi.org/10.1038/365337a0 PMID: 8377824

81. Stocks N, Mannella R. Generic noise-enhanced coding in neuronal arrays. Phys Rev E. 2001; 64(3): 030902. https://doi.org/10.1103/PhysRevE.64.030902

82. Pecevski D, Buesing L, Maass W. Probabilistic inference in general graphical models through sampling in stochastic networks of spiking neurons. PLoS Comput Biol. 2011; 7(12):e1002294. https://doi.org/10.1371/journal.pcbi.1002294 PMID: 22219717

83. Maass W. Noise as a resource for computation and learning in networks of spiking neurons. Proc IEEE. 2014; 102(5):860–880. https://doi.org/10.1109/JPROC.2014.2310593

84. Karbasi A, Salavati AH, Shokrollahi A, Varshney LR. Noise-enhanced associative memories. In: Burges CJC, Bottou L, Welling M, Ghahramani Z, Weinberger KQ, editors. Advances in Neural Information Processing Systems 26. MIT Press; 2013. p. 1682–1690.

85. Anderson JS, Lampl I, Gillespie DC, Ferster D. The contribution of noise to contrast invariance of orientation tuning in cat visual cortex. Science. 2000; 290(5498):1968–1972. https://doi.org/10.1126/science.290.5498.1968 PMID: 11110664

86. Sanda P, Marsalek P. Stochastic interpolation model of the medial superior olive neural circuit. Brain Res. 2012; 1434:257–265. https://doi.org/10.1016/j.brainres.2011.08.048 PMID: 21920505

87. van Vreeswijk C, Sompolinsky H. Chaos in neuronal networks with balanced excitatory and inhibitory activity. Science. 1996; 274(5293):1724–1726. https://doi.org/10.1126/science.274.5293.1724 PMID: 8939866

88. Silberberg G, Bethge M, Markram H, Pawelzik K, Tsodyks M. Dynamics of population rate codes in ensembles of neocortical neurons. J Neurophysiol. 2004; 91(2):704–709. https://doi.org/10.1152/jn.00415.2003 PMID: 14762148

89. van Rossum MC, Turrigiano GG, Nelson SB. Fast propagation of firing rates through layered networks of noisy neurons. J Neurosci. 2002; 22(5):1956–1966. PMID: 11880526

90. Destexhe A, Rudolph-Lilith M. Neuronal noise. vol. 8 of Springer Series in Computational Neuroscience. Springer; 2012.

91. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015; 521(7553):436–444. https://doi.org/10.1038/nature14539 PMID: 26017442

92. Cireşan D, Meier U, Schmidhuber J. Multi-column deep neural networks for image classification. In: Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on. IEEE; 2012. p. 3642–3649.

93. Mnih V, Kavukcuoglu K, Silver D, Rusu AA, Veness J, Bellemare MG, et al. Human-level control through deep reinforcement learning. Nature. 2015; 518(7540):529–533. https://doi.org/10.1038/nature14236 PMID: 25719670

94. Silver D, Huang A, Maddison CJ, Guez A, Sifre L, Van Den Driessche G, et al. Mastering the game of Go with deep neural networks and tree search. Nature. 2016; 529(7587):484–489. https://doi.org/10.1038/nature16961 PMID: 26819042

95. Moravčík M, Schmid M, Burch N, Lisỳ V, Morrill D, Bard N, et al. DeepStack: Expert-level artificial intelligence in heads-up no-limit poker. Science. 2017; 356(6337):508–513. https://doi.org/10.1126/science.aam6960 PMID: 28254783

96. van Seijen H, Fatemi M, Romoff J, Laroche R, Barnes T, Tsang J. Hybrid Reward Architecture for Reinforcement Learning. arXiv preprint arXiv:170604208. 2017;.

97. Rulkov N, Timofeev I, Bazhenov M. Oscillations in large-scale cortical networks: map-based model. J Comput Neurosci. 2004; 17(2):203–223. https://doi.org/10.1023/B:JCNS.0000037683.55688.7e PMID: 15306740

98. Rulkov NF, Bazhenov M. Oscillations and synchrony in large-scale cortical network models. J Biol Phys. 2008; 34(3-4):279–299. https://doi.org/10.1007/s10867-008-9079-y PMID: 19669478

99. Bruno RM. Synchrony in sensation. Curr Opin Neurobiol. 2011; 21(5):701–708. https://doi.org/10.1016/j.conb.2011.06.003 PMID: 21723114

100. Silberberg G. Polysynaptic subcircuits in the neocortex: spatial and temporal diversity. Curr Opin Neurobiol. 2008; 18(3):332–337. https://doi.org/10.1016/j.conb.2008.08.009 PMID: 18801433

101. Pouille F, Scanziani M. Enforcement of temporal fidelity in pyramidal cells by somatic feed-forward inhibition. Science. 2001; 293(5532):1159–1163. https://doi.org/10.1126/science.1060342 PMID: 11498596

102. Dong H, Shao Z, Nerbonne JM, Burkhalter A. Differential depression of inhibitory synaptic responses in feedforward and feedback circuits between different areas of mouse visual cortex. J Comp Neurol. 2004; 475(3):361–373. https://doi.org/10.1002/cne.20164 PMID: 15221951

103. Shao Z, Burkhalter A. Different balance of excitation and inhibition in forward and feedback circuits of rat visual cortex. J Neurosci. 1996; 16(22):7353–7365. PMID: 8929442

104. Legenstein R, Pecevski D, Maass W. A learning theory for reward-modulated spike-timing-dependent plasticity with application to biofeedback. PLoS Comput Biol. 2008; 4(10):e1000180. https://doi.org/10.1371/journal.pcbi.1000180 PMID: 18846203

105. Markram H, Lübke J, Frotscher M, Sakmann B. Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs. Science. 1997; 275(5297):213–215. https://doi.org/10.1126/science.275.5297.213 PMID: 8985014

106. Rulkov NF. Modeling of spiking-bursting neural behavior using two-dimensional map. Phys Rev E. 2002; 65(4):041922. https://doi.org/10.1103/PhysRevE.65.041922

107. Bazhenov M, Rulkov NF, Fellous JM, Timofeev I. Role of network dynamics in shaping spike timing reliability. Phys Rev E. 2005; 72(4):041903. https://doi.org/10.1103/PhysRevE.72.041903