Sequential neuromodulation of Hebbian plasticity offers mechanism for effective reward-based navigation

Zuzanna Brzosko et al
Figure 1. Acetylcholine biases STDP toward depression. (a) In control condition, the pre-before-post pairing protocol with Δt = +10 ms induced t-LTP (red), whereas in the presence of 1 μM acetylcholine (ACh), the same pairing protocol induced t-LTD (black). In the absence of the pairing protocol, Figure 1 continued on next page.
Figure 1 continued

ACh had no effect on baseline EPSPs (blue). Traces show an EPSP before (1) and 40 min after pairing (2) for each condition. (b) Summary of results (mean ± s.e.m.). (c) Application of muscarinic ACh receptor antagonist, 100 nM atropine, at the beginning of the recordings prevented ACh-facilitated t-LTD ($\Delta t = +10$ ms; black) and pre-before-post pairing resulted in significant t-LTP (red). Traces are presented as in a. (d) Summary of results. (e) Summary of the STDP induction with various spike timing intervals ($\Delta t$ in ms) in control condition (− ACh; red) and in the presence of ACh (+ ACh; black). Each data point is the group mean percentage change from baseline of the EPSP slope. Error bars represent s.e.m. Significant difference (*p<0.05, **p<0.01, ***p<0.001) compared with the baseline or between the indicated two groups (two-tailed Student’s t-test). The total numbers of individual cells (blue data points) are shown in parentheses.

DOI: 10.7554/eLife.27756.002

The following source data is available for figure 1:

Source data 1. Source data for Figure 1.
DOI: 10.7554/eLife.27756.003

Source data 2. Source data for Figure 1—figure supplement 1.
DOI: 10.7554/eLife.27756.004

Source data 3. Source data for Figure 1—figure supplement 2.
DOI: 10.7554/eLife.27756.005
Figure 1—figure supplement 1. Neuromodulation of STDP by dopamine and co-application of dopamine and acetylcholine. (a) Dopamine (DA) biases STDP toward potentiation: (ai) post-before-pre pairing protocol with Δt = −20 ms (data from Figure 1e in Brzosko et al., 2015) and (aii) pre-before-post pairing protocol with Δt = +10 ms. (b) Co-application of DA and acetylcholine (ACh) further biases STDP toward potentiation: (bi) post-before-pre protocol with Δt = −20 ms and (bii) pre-before-post protocol with Δt = +10 ms. (c) Summary of spike timing window for potentiation with co-application of DA and ACh. Significant differences are indicated by * (p < 0.05) and ** (p < 0.01).
Figure 1—figure supplement 1 continued

post pairing protocol with $\Delta t = +10$ ms. Traces show an EPSP before (1) and 40 min after pairing (2). (b) Co-application of DA and acetylcholine (ACh) yields depression. (bi) post-before-pre pairing protocol with $\Delta t = -20$ ms and (bii) pre-before-post pairing protocol with $\Delta t = +10$ ms produces initial synaptic depression which reverts back to baseline. Traces are presented as in a. (c) Summary of results. Each data point is the mean percentage change from baseline of the EPSP slope. Error bars represent s.e.m. Significant difference (*p<0.05, **p<0.01) compared with the baseline (two-tailed Student’s t-test). The total numbers of individual cells are shown in parentheses.

DOI: 10.7554/eLife.27756.006
Figure 1—figure supplement 2. Low concentration of acetylcholine prevents development of t-LTP. (a) In control condition, the pre-before-post pairing protocol with $\Delta t = +10$ ms induced t-LTP (red). Application of 100 nM acetylcholine (ACh) prevented development of significant t-LTP (blue). Traces show an EPSP before (1) and 40 min after pairing (2). (b) Summary of results. Error bars represent s.e.m. Significant difference (*p<0.05) compared with the baseline or between the indicated two groups (two-tailed Student’s t-test). The total numbers of individual cells (blue data points) are shown in parentheses.

DOI: 10.7554/eLife.27756.007
Figure 2. Dopamine retroactively converts acetylcholine-facilitated t-LTD into t-LTP. (a) Application of muscarinic acetylcholine receptor antagonist, 100 nM atropine, following acetylcholine (ACh; 1 μM) washout after the pre-
Figure 2 continued

before-post pairing protocol with \( \Delta t = +10 \text{ ms} \) did not affect the development of ACh-facilitated t-LTD (black). Dopamine (DA; 100 \( \mu \text{M} \)) applied, together with atropine, immediately after ACh washout at the end of the same pairing protocol, converted ACh-facilitated t-LTD into t-LTP (red). Traces show an EPSP before (1) and 40 min after pairing (2) in the two conditions. (b) Forskolin (50 \( \mu \text{M} \)), applied together with atropine, converted ACh-facilitated t-LTD into t-LTP, mimicking the effect of DA. Traces are presented as in a. (c) Summary of results from a and b. In the absence of the pairing protocol, atropine had no significant effect on baseline EPSPs (blue). Error bars represent s.e.m. Significant difference (*p<0.05) compared with the baseline or between the indicated two groups (two-tailed Student’s t-test). The total numbers of individual cells (blue data points) are shown in parentheses.

DOI: 10.7554/eLife.27756.008

The following source data is available for figure 2:

Source data 1. Source data for Figure 2.
DOI: 10.7554/eLife.27756.009

Source data 2. Source data for Figure 2—figure supplement 1.
DOI: 10.7554/eLife.27756.010
Figure 2—figure supplement 1. Acetylcholine applied after the pairing protocol does not affect t-LTP. Application of acetylcholine after the pre-before-post pairing protocol with Δt = +10 ms did not affect the development of t-LTP (n = 5). Traces show an EPSP before (1) and 40 min after (2) pairing. Error bars represent s.e.m.

DOI: 10.7554/eLife.27756.011
**Figure 3.** From plasticity to behavior: A computational model. (a) Schematic diagram of synaptic and behavioral timescales in reward-related learning. During Exploration, the activity-dependent modification of synaptic strength due to spike timing-dependent plasticity (STDP) depends on the coordinated spiking between presynaptic and postsynaptic neurons on a millisecond time scale. STDP develops gradually on a scale of minutes. Increased cholinergic tone (ACh) during Exploration facilitates synaptic depression. When Reward, signaled via dopamine (DA), follows Exploration with a Delay of seconds to minutes, synaptic depression is converted into potentiation. (b) Computational model. (bi) Symmetric STDP learning windows incorporated in the model, where acetylcholine biases STDP toward synaptic depression, while subsequent application of dopamine converts this depression into potentiation. (bii) The position of the agent in the field, \( x(t) \), is coded by place cells and its moves are determined by the activity of action neurons. STDP is implemented in the feed-forward connections between place cells and action neurons. Place cells become active with the proximity of the agent (active neurons in red: the darker, the higher their firing rate). Place cells are connected to action neurons through excitatory synapses \( w_{\text{feed}} \) (the darker, the stronger the connection). Action neurons are connected with each other: recurrent synaptic weights \( w_{\text{lat}} \) are excitatory (red) when action neurons have similar tuning, or inhibitory (blue) otherwise. Thus, the activation of action neurons is dependent on both the feed-forward and recurrent connections. (biii) Each action neuron \( j \) codes for a different direction \( a_j \) (large arrow’s direction) and has a different firing rate \( p_j \) (large arrow's color: the darker, ...
Figure 3 continued

The action to take $a(t)$ (black arrow) is the average of all directions, weighted by their respective firing rate. (biv) The agent takes action $a(t)$. Therefore, it moves to $x(t + \Delta t) = x(t) + a(t)$.

DOI: 10.7554/eLife.27756.012
Figure 4. Temporally sequenced cholinergic and dopaminergic modulation of STDP yields effective navigation toward changing reward locations. (a) Learning of an initial reward location (trials 1–20; 1000 simulations in each trial) shows a modest improvement in learning when cholinergic depression is included in the model. (i) Example trajectories. The agent starts from the center of the open field (red dot) and learns the reward location (closed black circle) with (+ ACh; brown) and without (− ACh; green) cholinergic depression built into the model. Trials are coded from light to dark, according to their temporal order (early = light, late = dark). (ii) Color scheme. (iii) Reward discovery. The graph shows percent cumulative distribution of trials in which the reward location is discovered for the first time. (iv) Learning curve presented as a percentage of successful simulations over successive trials. (v) Average time to reward in each successful trial. Unsuccessful trials, in which the agent failed to find the reward, were excluded. (vi) Percentage of successful simulations in trial 5, under conditions with different magnitudes of dopaminergic effect (learning windows in the top-left corner). Decreasing the magnitude of dopaminergic potentiation significantly affects learning (p<0.001, two-sample Student’s t-test: Small vs. Medium and Small vs. Large). Under Medium and Large conditions, the agent performs similarly most likely due to a saturation effect (p>0.05, two-sample Student’s t-test: Medium vs. Large). (b) Learning of a displaced reward location is facilitated when cholinergic depression is included in the model. (i) Example trajectories (trials 21–40; 1000 simulations in each trial). The agent learns a novel reward location (closed circle; previously exploited reward = open circle). Trajectories presented as in ai. Comparison of control (− ACh) and test (+ ACh) simulations: (ii) visits to previous reward location (%); (iii) trial number at novel reward discovery; (iv) successful reward collection over successive trials (%); (v) average time to reward over trials. (vi) Percentage of visits to the old reward Figure 4 continued on next page
location in trial 25, under conditions with different magnitudes of cholinergic depression (learning windows in the top-right corner). Increasing the magnitude of acetylcholine effect yields faster unlearning ($p<0.001$, two-sample Student’s t-test: Small vs. Medium, Medium vs. Large and Small vs. Large). The graphs (biii-bv) are presented as in a. The shaded area (aiv-v and bii, biv-v) represents the 95% confidence interval of the sample mean.

DOI: 10.7554/eLife.27756.013
Figure 4—figure supplement 1. Exploration following reward displacement. The mean firing rate of place cells (average over time and simulations) mapped onto the open field. When the reward location is displaced in Trial 21, in both test (+ ACh, right) and control (− ACh, left) simulations, agents still navigate toward the initial reward location (warmer colors in the upper right quadrant). By Trial 26, cholinergic depression in test simulations allows unlearning of the old reward location, which results in transient enhancement of exploration compared to control simulations. By Trial 31, agents with acetylcholine-modulated plasticity are able to successfully navigate to the novel goal area (lower left quadrant) whilst in control simulations most agents are not.

DOI: 10.7554/eLife.27756.014
**a Initial reward location**

i

- $A_{DA} = 0.1$
- $A_{DA} = 1.0$
- $A_{DA} = 3.0$

**b Novel reward location**

i

**Figure 4—figure supplement 2.** The magnitude of dopamine effect affects learning. (a) The agents are subject to the same learning paradigm as in Figure 4. Small magnitude of dopamine effect slows down learning and decreases performance. (ai) The parameters used for Small, Medium and Large amplitudes of dopaminergic effect are specified in the legend; darker colors correspond to larger amplitudes. (aii) Learning curve presented as a percentage of successful simulations over successive trials (trials 1–20; 1000 simulation). Decreasing the magnitude of dopamine effect leads to slower learning. In Medium and Large conditions, the agent’s performance is similar most likely due to a saturation effect. (b) Learning of a displaced reward location is only marginally slower in Small condition. (bi) Over the trials, the percentage of visits to the previously rewarded location decreases for Medium and Large conditions (trials 21–40; 1000 simulation). However, Small condition agents present an initial advantage due to weaker learning in the first phase of the experiment. (bii) Learning of the novel reward location is slightly faster for Medium and High conditions. The shaded area (aii and bi-ii) represents the 95% confidence interval of the sample mean.

DOI: 10.7554/eLife.27756.015
Figure 4—figure supplement 3. The magnitude of acetylcholine effect affects unlearning. (a) The agents are subject to the same learning paradigm as in Figure 4. Difference in the magnitude of cholinergic effect does not affect performance. (ai) The parameters used for Small, Medium and Large magnitudes of acetylcholine effect are specified in the legend; darker colors correspond to larger magnitudes of cholinergic effect. (aii) Learning curve presented as a percentage of successful simulations over successive trials (trials 1–20; 1000 simulation). The initial difference in performance is due to reduced speed, a consequence of the online depression. (b) Learning of a displaced reward location becomes more efficient with larger magnitudes of acetylcholine effect. (bi) The percentage of visits to the previously rewarded location is lower for conditions with larger magnitudes of acetylcholine effect (trials 21–40; 1000 simulation). (bii) Learning of the novel reward location is also faster for Medium and Large conditions. The shaded area (aii and bi-ii) represents the 95% confidence interval of the sample mean.

DOI: 10.7554/eLife.27756.016
**Figure 4**—figure supplement 4. The integral of the asymmetric STDP learning window determines the performance of the agent. (a) The agents are subject to the same learning paradigm as in Figure 4. Learning is successful only when the integral of the STDP window is positive. (ai) Learning windows with negative, zero and positive integrals. Parameters and color scheme are as specified at the top. (a(ii) Learning curve presented as a percentage of successful simulations over successive trials (trials 1–20; 1000 simulations). Only the agents using a learning rule with a net positive integral of the STDP window learn successfully. (b) Learning of a displaced reward location is not achieved successfully by any of the STDP learning rules. (b(i)) The percentage of visits to the previous reward area is low only for agents with a negative integral of the STDP window (trials 21–40; 1000 simulations). This is because unlearning occurred in the first phase of the experiments. (b(ii) Agents with a positive integral of the STDP window only partially learn the new reward location, but do not effectively unlearn the previous reward location (as shown in b(i)). Agents with a negative integral of the STDP window unlearn both the old and the new reward areas. The shaded area (aii and bi-ii) represents the 95% confidence interval of the sample mean.

DOI: 10.7554/eLife.27756.017