Article

Neuron

Distinct Eligibility Traces for LTP and LTD in Cortical Synapses

Highlights

- Hebbian conditioning induces eligibility traces for LTP and LTD in cortical synapses
- β_2ARs and 5-HT$_{2C}$Rs convert the traces into LTP and LTD, respectively
- Anchoring of β_2ARs and 5-HT$_{2C}$ is key for trace conversion
- Temporal properties of the LTP/D traces allow reward-timing prediction

Authors

Kaiwen He, Marco Huertas, Su Z. Hong, XiaoXiu Tie, Johannes W. Hell, Harel Shouval, Alfredo Kirkwood

Correspondence

kirkwood@jhu.edu

In Brief

How is stimulus-evoked activity associated with a time-delayed reward in reinforcement learning? He et al. report on the existence of silent and transient synaptic tags (eligibility traces) that can be converted into long-term changes in synaptic strength by reward-linked neuromodulators.

He et al., 2015, Neuron *88*, 528–538
November 4, 2015 ©2015 Elsevier Inc.
http://dx.doi.org/10.1016/j.neuron.2015.09.037

Distinct Eligibility Traces for LTP and LTD in Cortical Synapses

Kaiwen He,[1] Marco Huertas,[2] Su Z. Hong,[1] XiaoXiu Tie,[1] Johannes W. Hell,[3] Harel Shouval,[2] and Alfredo Kirkwood[1,]*
[1]Mind/Brain Institute, Johns Hopkins University, 3400 North Charles Street, 350 Dunning Hall, Baltimore, MD 21218, USA
[2]Department of Neurobiology and Anatomy, University of Texas at Houston, 6431 Fannin Street, Suite MSB 7.046, Houston, TX 77030, USA
[3]Department of Pharmacology, University of California, Davis, 1544 Newton Court, Davis, CA 95618, USA
*Correspondence: kirkwood@jhu.edu
http://dx.doi.org/10.1016/j.neuron.2015.09.037

SUMMARY

In reward-based learning, synaptic modifications depend on a brief stimulus and a temporally delayed reward, which poses the question of how synaptic activity patterns associate with a delayed reward. A theoretical solution to this so-called distal reward problem has been the notion of activity-generated "synaptic eligibility traces," silent and transient synaptic tags that can be converted into long-term changes in synaptic strength by reward-linked neuromodulators. Here we report the first experimental demonstration of eligibility traces in cortical synapses. We demonstrate the Hebbian induction of distinct traces for LTP and LTD and their subsequent timing-dependent transformation into lasting changes by specific monoaminergic receptors anchored to postsynaptic proteins. Notably, the temporal properties of these transient traces allow stable learning in a recurrent neural network that accurately predicts the timing of the reward, further validating the induction and transformation of eligibility traces for LTP and LTD as a plausible synaptic substrate for reward-based learning.

INTRODUCTION

A central aim of learning in biological organisms is to maximize reward. To achieve this aim, animals must learn what stimuli and actions predict an often-delayed reward and when the reward is likely to arrive. This poses a fundamental question regarding the synaptic mechanisms of learning: How can a delayed reward gate plasticity in synapses that were transiently activated by the predictive stimulus? A theoretical solution proposed decades ago to bridge the temporal gap between stimulus and reward, the so-called credit assignment problem, is the notion that neural activity generates silent and transient "synaptic eligibility traces" that can be transformed into long-term changes in synaptic strength by reward-linked neuromodulators (Crow, 1968; Frémaux et al., 2010; Gavornik et al., 2009; Hull, 1943; Izhikevich, 2007; Klopf, 1982; Sutton and Barto, 1998; Turner et al., 2003; Wörgötter and Porr, 2005).

In most theoretical models of reward-driven learning, synaptic eligibility traces are typically induced in a Hebbian manner by coincident pre- and postsynaptic activity and have half-times on the order of seconds (Frémaux et al., 2010; Izhikevich, 2007; Klopf, 1982; Sutton and Barto, 1998), during which they can be converted into long-term changes by the action of neuromodulators. Although bidirectional synaptic plasticity induced by coincident activity is well established, particularly in the form of spike-timing-dependent plasticity (STDP) (Caporale and Dan, 2008; Richards et al., 2010), the existence of eligibility traces for long-term potentiation (LTP) has been reported in only two studies, neither of them in cortex (Cassenaer and Laurent, 2012; Yagishita et al., 2014).

Recent findings in rodents and humans have implicated primary sensory cortices in reinforced learning (Chubykin et al., 2013; Gardner and Fontanini, 2014; Jaramillo and Zador, 2011; Poort et al., 2015; Seitz et al., 2009; Shuler and Bear, 2006), making them attractive systems to examine the existence of eligibility traces. Historically, neuroplasticity associated with reward has been studied primarily in the dopaminergic system and its projection areas, including basal ganglia and prefrontal cortex, which are involved in detecting reward and orchestrating the appropriate response. However, the process of learning to recognize the reward-predicting stimuli likely involves remodeling in primary sensory cortices as well. Cells in primary sensory cortices can predict essential attributes of the reward, including timing (Poort et al., 2015; Shuler and Bear, 2006) and value (Gardner and Fontanini, 2014).

We examined the existence of eligibility traces in layer II/III pyramidal cells in slices from both visual and prefrontal cortices. An important motivation was the observation in the visual cortex that the Hebbian induction of LTP and long-term depression (LTD) depends crucially on not only glutamate receptors but also neuromodulator receptors coupled to Gs and Gq (Choi et al., 2005; Huang et al., 2012; Yang and Dani, 2014). In reinforcement learning, reward is typically delayed. We therefore tested whether neuromodulators could also act in a retrograde manner to allow synaptic changes when applied after conditioning. In both visual and prefrontal cortices, we demonstrated the Hebbian induction of short-lived eligibility traces that can be converted into either LTP or LTD by specific monoamines. We found that LTP- and LTD-associated traces have different dynamics, and we demonstrated the functional significance of these different dynamics by showing that temporal competition between these eligibility traces produces stable learning that allows a recurrent neural network to predict the arrival time of reward.

RESULTS

Specific Monoamines Transform Synaptic Eligibility Traces Induced by Spike-Timing Conditioning into LTP or LTD

As mentioned earlier, in cortex, unlike other structures such as the hippocampus, the induction of Hebbian plasticity depends critically on the activation of G protein-coupled receptors (GPCRs) such that blockade of these receptors or depletion of the endogenous neuromodulators prevents LTP and LTD (Choi et al., 2005; Huang et al., 2012). Moreover, because of this GPCR dependency, under certain experimental conditions, including ours, the Hebbian induction of synaptic plasticity with spike-timing (ST)-dependent conditioning requires the addition of exogenous neuromodulators (Edelmann and Lessmann, 2013; Huang et al., 2014; Seol et al., 2007; Yang and Dani, 2014). We exploited this fact to directly test the induction of eligibility traces in cortical slices by determining whether ST conditioning can result in LTP or LTD if it is rapidly followed by an application of neuromodulator agonists. The neuromodulators tested were norepinephrine (NE), serotonin, dopamine (DA), and acetylcholine, all of which have been implicated in cortical plasticity. We first focused on the primary visual cortex, where reward-based changes are well established in both primates (including humans) and rodents (Goltstein et al., 2013; Poort et al., 2015; Seitz et al., 2009; Shuler and Bear, 2006). The recordings were done in layer II/III pyramidal cells and involved activation of two independent layer IV to layer II/III pathways, which were conditioned simultaneously with near-coincidental pre- and postsynaptic stimulation (ST conditioning, Figures 1A and 1B). In one pathway, presynaptic stimulation preceded a burst of postsynaptic potentials by 10 ms (pre-post, to promote LTP); in the other one, it occurred 10 ms after the burst (post-pre, to promote LTD). Neuromodulators were pressure ejected from a nearby pipette beginning immediately after ST conditioning and continuing for 10 s. As expected, under control conditions, the ST conditioning elicited no plasticity (Figure 1C, pre-post: p = 0.563; post-pre: p = 0.156), but plasticity was observed when the ST conditioning was immediately followed by pressure ejection of NE (50 μM, 10 s) or serotonin (5-hydroxytryptamine, or 5-HT; 50 μM, 10 s). NE selectively potentiated the pre-post pathway without affecting the post-pre pathway (Figure 1D, pre-post: p = 0.002; post-pre: p = 0.232); conversely, 5-HT selectively depressed the post-pre pathway but not the pre-post pathway (Figure 1E, pre-post: p = 0.160; post-pre: p = 0.002). Pressure ejection of the agonists alone in naive (non-conditioned) pathways had no lasting effect on synaptic strength (NE only: 102.9% ± 5.6%, n = 6; 5-HT only: 102.8% ± 8.5%, n = 5; data not shown), confirming that the monoamine agonists were converting previously induced eligibility traces into changes of synaptic strength.

In contrast to NE and 5-HT, no effect was observed with DA application (50 μM) (Figure 1F, pre-post: p = 0.843; post-pre: p = 1), which is not surprising given that dopaminergic transmission is minimal in the visual cortex. Similarly, application of the cholinergic agonist carbachol (CCh: 250 μM) (Figure 1G, pre-post: p = 0.742; post-pre: p = 0.547) after ST conditioning did not affect the excitatory postsynaptic potentials (EPSPs), even with a long (5 min) puff duration (Figure S1A). However, and confirming previous findings (Kirkwood et al., 1999), the long CCh exposure promoted LTD induction if applied before the ST conditioning (Figure S1B). Thus, only a subset of neuromodulators can transform eligibility traces into LTP and LTD. The induction of the traces, however, is a general phenomenon not restricted to ST conditioning, and it can be achieved by pairing synaptic stimulation (10 Hz, 20 s) with sustained postsynaptic depolarization (-10 mV for LTP and -40 mV for LTD). Conditioning by pairing to -10 mV depolarization produced a modest LTP (109.64 ± 3.59%, n = 8, p = 0.005, data not shown). Consistent with the crucial role of neuromodulators in cortical LTP (Choi et al., 2005; Huang et al., 2012), this LTP was substantially impaired (101.36% ± 4.58%, n = 8, Figure S1C) if the endogenous monoamines were depleted by reserpine injection 1 day before the experiments (Choi et al., 2005; Otmakhova and Lisman, 1996). In these depleted slices, however, LTP developed robustly when NE was puffed on after the conditioning protocol (131.28% ± 7.08%, n = 7, p = 0.006. Figure S1C). Similarly, 10 Hz stimulation paired to -40 mV depolarization alone was not able to induce LTD (106.3% ± 7.0%, n = 9, Figure S1D) in the reserpine-injected mouse. However, it caused prominent LTD when immediately followed by the 5-HT puff (78.2% ± 6.8%, n = 9, p = 0.027, Figure S1D).

To evaluate the generality of the eligibility traces, we extended the studies to layer II/III synapses of the medial prefrontal cortex (mPFC), which is highly innervated by dopaminergic, noradrenergic, and serotonergic fibers and has been implicated in multiple forms of reward-based learning (Kahnt et al., 2011; Ridderinkhof et al., 2004; Rushworth et al., 2011). As in the visual cortex, NE (50 μM, 10 s) transformed the trace in the pre-post pathway into LTP (Figure 2A, p = 0.01) and 5-HT (50 μM, 10 s) transformed the trace in the post-pre pathway into LTD (Figure 2B, p = 0.008). Unlike in the visual cortex, however, DA (50 μM, 10 s) did transform the trace in the pre-post pathway into LTP (Figure 2C, p = 0.01). However, CCh was ineffective in either pathway (Figure 2D, pre-post: p = 0.156; post-pre: p = 0.125). Altogether, these results indicate that eligibility traces for LTP and LTD can be induced in a Hebbian manner and that distinct and specific monoamine neuromodulators can transform these invisible traces into long-term synaptic plasticity throughout many cortical areas.

Endogenous Monoamines Can Transform Synaptic Eligibility Traces

Although puffing neuromodulators at a high concentration yields consistent results, this paradigm may not resemble conditions in vivo. Therefore, we tested a more physiological paradigm for the transformation of eligibility traces by releasing endogenous neuromodulators with optogenetics in TH-ChR2 and Tph2-ChR2 mice, which express channelrhodopsin-2 (ChR2) in adrenergic or dopaminergic (Figure S2) and serotonergic nuclei (Zhao et al., 2011), respectively. Similar to puffing, release of endogenous NE only transformed the LTP eligibility trace (Figure 3A, pre-post: p = 0.039) while endogenous 5-HT only transformed the LTD trace (Figure 3C, post-pre: p = 0.002) in the visual cortex.

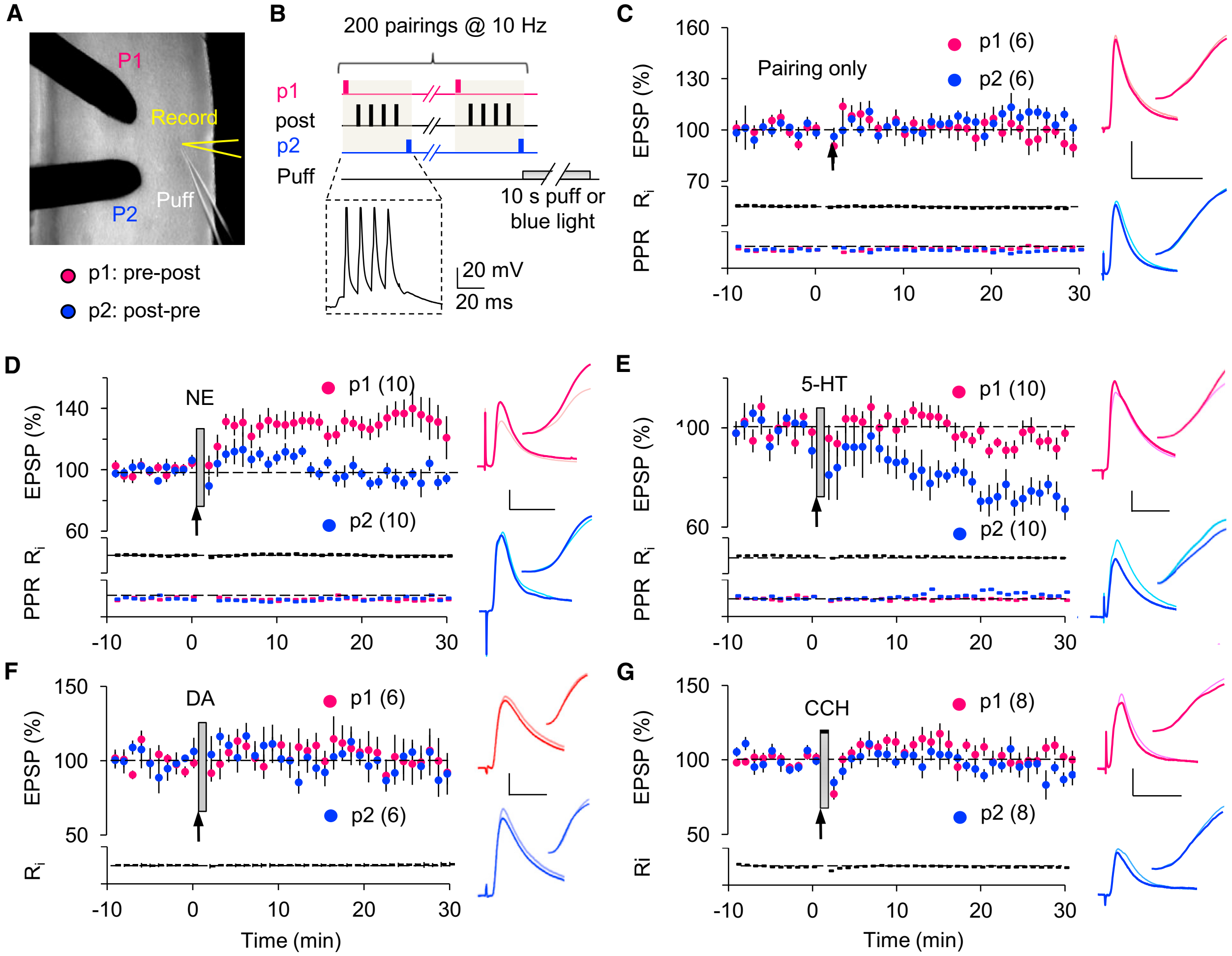

Figure 1. Specific Monoamines Transform STDP-Induced Eligibility Traces into LTP and LTD

(A) Two-pathway whole-cell recording configuration.

(B) Induction of eligibility traces with STDP paradigms. A representative response for two-pathway ST conditioning is shown in the dashed box.

(C) In the visual cortex, ST conditioning alone did not affect synaptic strength in either the pre-post (red dots) or the post-pre (blue dots) pathway.

(D and E) Pressure ejection of NE (50 μM, 10 s, gray bar) immediately after the ST conditioning (arrow) converted LTP eligibility traces in the pre-post pathway (pre-post in D: 132.3% ± 9.0%), while a similar puff of 5-HT (50 μM) transformed LTD traces in the post-pre pathway (post-pre in E: 73.1% ± 4.5%).

(F and G) Eligibility traces were not affected by pressure ejection of either 50 μM DA (F) or 50 μM CCh (G).

The number of experiments is indicated in parentheses. Traces in (C)–(G) are averages of 10 EPSPs of the two pathways (red: pre-post; blue: post-pre) recorded in the same neuron immediately before (thin light-colored line) or 25 min after (thick dark-colored line) conditioning. Scale, 2 mV, 25 ms.

See also Figure S1.

Importantly, the transformation of the long-term potentiation or depression (LTP/D) traces only happened when the monoamines were released after the Hebbian conditioning, not before (Figures 3B and 3D). The requirement for a strict temporal order between the ST conditioning and the phasic release of neuromodulators mirrors the sequential order of stimulus-reward in reinforcement learning.

Reinforced learning in behaving animals occurs over multiple stimulus-reward epochs spaced in time (Chubykin et al., 2013; Seitz et al., 2009). This differs from the protocols we used earlier, which were chosen to demonstrate unequivocally the induction and transformation of the eligibility traces. In this study, we delivered the neuromodulators just once after 200 Hebbian conditionings that were massed into a single induction epoch. To better mirror reinforcement learning, we tested whether optogenetic reinforcement of individual ST-conditioning epochs (40 pre-post or post-pre pairings spaced by 20 s intervals) can also result in LTP/D. In slices from TH-ChR2 mice, 1 s trains of blue light pulses (10 ms at 10 Hz) that were flashed immediately after each pre-post conditioning epoch induced robust LTP (Figure 4A, p1, p = 0.016). Similarly, in the Tph2-ChR2 mice, the blue light flashed immediately after each post-pre conditioning epoch induced LTD (Figure 4B, p1, p = 0.002). In both cases (Figures 4A and 4B), synaptic responses in control pathways that were conditioned with the ST epochs but out of phase with the blue light flashes (10 s gap) did not change (p2, pre-post

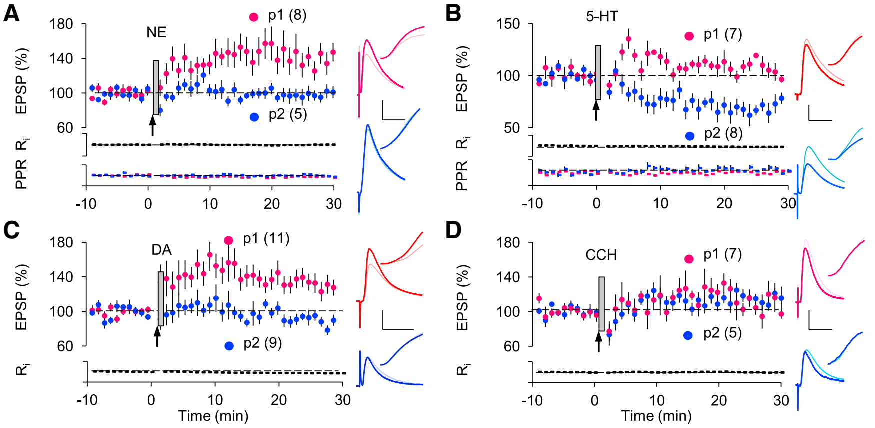

Figure 2. Eligibility Traces in the Prefrontal Cortex

(A) In layer II/III synapses of the mPFC, a 10 s puff of NE (50 μM) transformed the LTP trace (pre-post: 133.1% ± 9.7%).

(B) A puff of 5-HT (50 μM) transformed the LTD trace (post-pre: 72.0% ± 7.3%).

(C) A puff of DA (50 μM) transformed the LTP trace (pre-post: 133.1% ± 9.7%).

(D) A puff of CCh (250 μM) did not affect the EPSPs (pre-post: 113.5% ± 7.4%; post-pre: 116.6% ± 8.6%).

Traces in (A) to (D) are coded as in Figure 1. Scale, 2 mV, 25 ms.

only: p = 0.164, Figure 4A; p2, post-pre only: p = 0.734, Figure 4B). Altogether, these results indicate that the monoamine-mediated transformation of eligibility traces is a physiologically plausible mechanism to encode reward-based learning in vivo.

Transformation of Short-Lived Synaptic Eligibility Traces Requires Anchoring of Monoamine Receptors

Previously we showed that stimulation of the Gs- and Gq-coupled receptors promotes LTP and LTD, respectively (Seol et al., 2007). It was surprising therefore that NE and DA, which stimulate both types of receptors, only affected the eligibility traces for LTP. Only 5-HT acted on the LTD traces. To solve this conundrum, we first set out to identify the relevant neuromodulator receptors using receptor-specific antagonists. One attractive candidate among the adrenoreceptors coupled to Gs were beta 2 adrenergic receptors (β_2ARs), which are enriched in spines and promote LTP (Davare et al., 2001; Qian et al., 2012). We found that the β_2AR antagonist (ICI 118,551, 1 μM) blocked the transformation of the LTP traces by NE (Figure 5A). Moreover, the beta adrenergic receptor agonist isoproterenol (Iso, 50 μM) was sufficient to transform the LTP trace, as was direct elevation of the intracellular cyclic AMP (cAMP) level, which is consistent with the role of β_2AR stimulation in cAMP production (Figure S3). On the other hand, the generic 5-HT_2 antagonist ketanserin (1 μM) blocked the transformation of the LTD trace (99.97% ± 6.75%, n = 7, data not shown). In addition, and consistent with the absence of 5-HT_{2A} receptors in layer II/III (Weber and Andrade, 2010), the specific 5-HT_{2C} receptor (5-HT_{2C}R) antagonist RS 102221 (1 μM) was sufficient to block the transformation of the LTD traces by 5-HT (Figure 5B). Thus, although multiple Gs- and Gq-coupled receptors, including the noradrenergic α1 and the cholinergic m1, may prime the subsequent induction of synaptic plasticity in the visual cortex, our results strongly suggest that β_2AR and 5-HT_{2C}R are mainly responsible for transforming previously induced eligibility traces.

One possible determinant of the specific role of β_2AR and 5-HT_{2C}R in trace transformation is the subcellular location of these receptors. Both receptors can directly interact with PDZ domain-containing proteins such as postsynaptic density protein 95 (PSD-95) and/or MUPP1 (Becamel et al., 2001; Bécamel et al., 2004; Joiner et al., 2010), suggesting that they are anchored at or close to the synapse. Therefore, we tested the effects of disrupting their interaction with PDZ proteins by adding the C-terminal peptides of β_2AR (DSPL: 50 μM) or 5-HT_{2C}R (2C-ct: 50 μM) to the recording electrode (Gavarini et al., 2006; Joiner et al., 2010) (Figures 5C–5F). DSPL, but not the control peptide DAPA (with the −2 and 0 positions changed to alanine), blocked the NE-mediated transformation of the LTP trace (Figure 5D, p = 0.041 between DSPL and DAPA), while the 2C-ct peptide, but not its scrambled control CSSA, prevented the transformation of the LTD eligibility trace (Figure 5F, p = 0.004 between 2C-ct and CSSA). The peptides did not block synaptic plasticity induced by presynaptic stimulation paired with postsynaptic depolarization, which is an effective induction protocol that does not require added neuromodulators (Figure S4; see Experimental Procedures and Huang et al., 2012, for further details). This indicates that the anchoring of receptors was only required for the conversion of the eligibility traces, not for the induction of plasticity. These results suggest that β_2AR and 5-HT_{2C}R needs to be anchored at or close to the synapse to convert transient eligibility traces.

LTP/D Synaptic Eligibility Trace Properties Allow a Network to Learn to Predict Reward Timing

Theoretical considerations suggest that synaptic eligibility traces should be transient, but experimentally little is known about their duration (Yagishita et al., 2014). Moreover, since distinct traces for LTP and LTD have not previously been described either experimentally or theoretically, nothing is known about the temporal properties of LTD traces. We set out to study the duration of the different eligibility traces and found that they have different durations. We show theoretically that these different durations are sufficient for producing stable learning in recurrent networks that learn to predict expected reward times.

To experimentally study the durations of the eligibility traces, we varied the delay between the ST conditioning and the puff of the neuromodulators (Figure 6A, insert). The LTP magnitude was reduced by about half when the agonist puff was delayed by 5 s, and it was gone if the agonist puff was delayed by 10 s (Figure 6; p = 0.007 between Δt = 10 s and Δt = 0 s). The LTD eligibility trace was even shorter, and by 5 s it was absent (Figure 6;

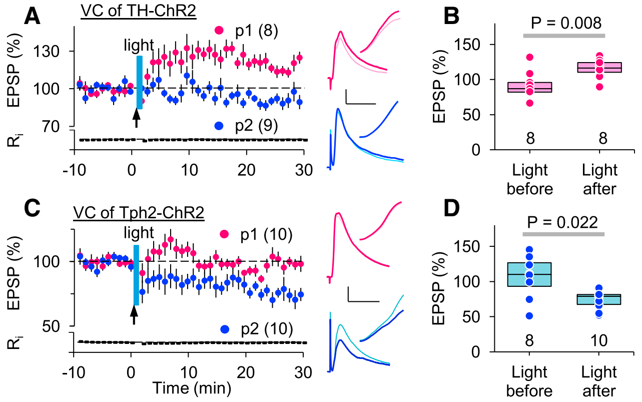

Figure 3. Endogenous Neuromodulators Released Optogenetically Transform Previously Induced Eligibility Traces

(A and C) In the visual cortex, local release of endogenous NE in the TH-ChR2 mouse (A) or 5-HT in the Tph2-ChR2 mouse (C) by optogenetic stimulation (blue bar) transformed the LTP/D eligibility traces generated by ST conditioning (pre-post in A: 115.5% ± 4.4%; post-pre in C: 73.8% ± 8.9%).

(B and D) Neuromodulators only consolidate eligibility traces when phasically released after, not immediately before (no overlap between the light and the conditioning), the ST conditioning (light before in B: 90.7% ± 6.7%; light before in D: 106.2% ± 11%).

Traces in (A) and (C) are coded as in Figure 1. Scale, 2 mV, 25 ms. See also Figure S2.

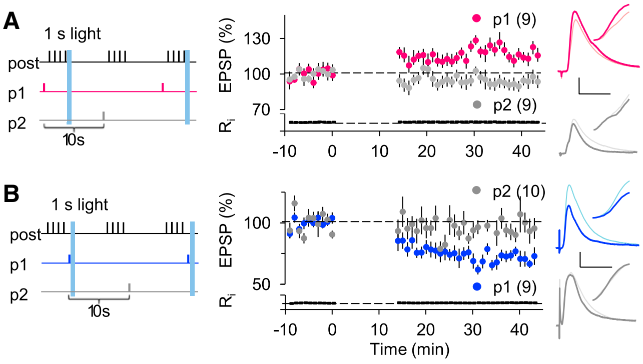

Figure 4. Optogenetic Release of Endogenous Neuromodulators Transforms Eligibility Traces Induced by Spaced Single ST Conditioning

(A and B) Two pathways received 40 ST-conditioning epochs in an alternated manner every 20 s. One pathway (red or blue symbols) was paired with 1 s light (10 light pulses of 10 ms and 700 mA each delivered at 10 Hz). The unpaired pathway (gray symbols) served as a control.

(A) Light stimulation transforms LTP traces induced by pre-post conditioning (red symbols) in slices from TH-ChR2 mice.

(B) Light stimulation transforms LTD traces induced by post-pre conditioning (blue symbols) in slices from the Tph2-ChR2 mice.

Traces in (A) and (B) are coded as in Figure 1. Scale, 2 mV, 25 ms

p = 0.003 between Δt = 5 s and Δt = 0 s). Thus, the eligibility traces are short lived, with the LTD trace substantially shorter than the LTP trace.

In general, learning rules must not only represent the statistics of the environment but also find stable solutions in which synaptic efficacies do not saturate or fall to zero. A possible consequence of having two eligibility traces, one for LTP and one for LTD, is that the balance between LTP and LTD could produce stable learning. Synaptic eligibility traces as observed experimentally are Hebbian and therefore depend on network dynamics, which in turn depend on synaptic efficacies. Here we propose that under certain conditions, the difference observed in the temporal dynamics of the eligibility traces can generate stable reinforcement learning in cortical networks.

We illustrated this process in the context of learning to predict reward timing within a recurrent neural network. Our example is motivated by several experiments in the primary sensory cortex (Chubykin et al., 2013; Gavornik et al., 2009; Goltstein et al., 2013; Shuler and Bear, 2006), in which a stimulus paired with a delayed reward results in cortical cells that remain active until the expected reward time. To this end, we simulated the activity of a recurrent network of excitatory neurons (architecture depicted in Figure 7A). Model details and equations are in Mathematical Model, which implements a learning rule based on two eligibility traces with different dynamics, as observed experimentally (Figure 6). Such a network, as shown previously (Gavornik and Shouval, 2011; Gavornik et al., 2009), can generate long-lasting dynamics that predict the timing of reward by learning the appropriate choice of lateral connection strengths, denoted by the connection matrix L (Figure 7A). Previously, a learning rule based on a single eligibility trace and active inhibition of reward was proposed, but this rule is inconsistent with experimental results (Chubykin et al., 2013; Gavornik and Shouval, 2011; Gavornik et al., 2009; Liu et al., 2015). We replaced the previous learning rule with a rule consistent with the experimental findings discovered here. The learning rule proposed here is based on the following minimal set of assumptions. First, two eligibility traces, one for LTP and one for LTD, are activated in a Hebbian manner. Second, the time constant of the LTP trace is longer than that of the LTD trace. Third, the LTD trace saturates at higher effective values than does the LTP trace. Finally, the change in synaptic weights depends on the difference between the LTP and the LTD traces at the time of reward. These assumptions are implemented mathematically by Equations 1, 2, and 3 in Mathematical Model. The first two assumptions are explicitly demonstrated experimentally in this paper, and the other assumptions are biologically plausible. The network (Figure 7A) was trained by repeatedly pairing a brief feed-forward stimulus (100 ms) with a reward delayed by 1,000 ms. Initially, the network responded only to the presentation of the stimulus (Figure 7B), but over the course of many trials, strengthening of the recurrent synaptic weights (indicated by L in Figure 7A) transformed the network's activity into a sustained response that decayed slowly, spanning the time between the stimulus and the expected reward (Figures 7C and 7D; raster plots in Figure S5). After training, the network exhibited sustained activity that terminated near the expected time of reward, indicating that the network learned to represent the reward timing, similar to what was observed in the rodent visual cortex after a similar training procedure (Chubykin et al., 2013; Shuler and Bear, 2006). This self-limiting sustained network activity results from the temporal competition between the LTP (red) and the LTD (blue) eligibility traces (Figures 7E–7G). Initially, at the time of reward, the LTP eligibility trace

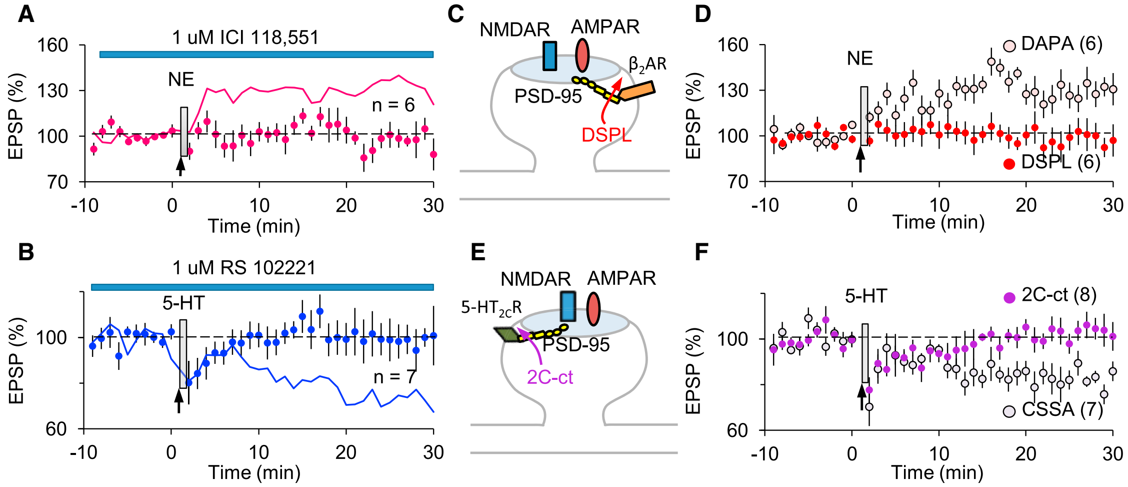

Figure 5. Anchoring of Monoamine Receptors Is Crucial for the Transformation of Transient LTP/D Eligibility Traces

(A) The β_2AR-specific antagonist ICI 118,551 (1 μM) prevents the transformation of the LTP eligibility trace by NE (95.2% ± 5.3%). The magenta line depicts control LTP (data from Figure 1D).

(B) The 5-HT$_{2C}$R-specific antagonist RS 102221 (1 μM) prevents the transformation of the LTD eligibility trace by 5-HT (99.8% ± 8.2%). The blue line depicts control LTD (data from Figure 1E).

(C) β_2AR directly interacts with PSD-95, and its C-terminal peptide DSPL disrupts this interaction.

(D) DSPL, but not the scrambled peptide DAPA, abolished the NE-mediated transformation of the LTP eligibility trace (DSPL: 96.1% ± 8.2%; DAPA: 127.8% ± 7.9%).

(E) The C-terminal peptide 2C-ct prevents the interaction between 5-HT$_{2C}$R and PDZ-containing proteins such as PSD-95.

(F) 2C-ct, but not the control peptide CSSA, blocked transformation of the LTD eligibility trace by 5-HT (2C-ct: 102.9% ± 3.7%; CSSA: 82.6% ± 3.9%).

See also Figures S3 and S4.

(Figure 7E, red) is larger than the LTD-related trace (Figure 7E, blue), resulting in net LTP. The increase in recurrent synaptic efficacies causes reverberations in the network that extend the network activity (Figure 7C). When the network activity is still significantly shorter than the delay to reward, the LTP eligibility trace still dominates (Figure 7F). When the duration of activity in the network approaches the reward time (Figure 7D), the eligibility traces at the time of reward cancel each other out (Figure 7G) and the network dynamics are stabilized. If the network dynamics overshoot the reward time, or if the reward time is modified to a shorter delay, the LTD-related trace would dominate and the network dynamics would become shorter and stabilize at the correct reward interval (Figures S5C1–S5C3). This learning mechanism is robust and can be used to learn the timing for reward arriving over a large range of temporal delays (Figure 7H).

After training, network dynamics do not terminate exactly at the time of reward but decay just before its arrival (Figure 7; Figure S5). The time between the termination of network dynamics and the delivery of reward (defined as *D*) depends on the parameters of the learning rule (Figures S5D and S5E), and this can be approximately characterized by a simple formula (Mathematical Model; Figure S5E).

Figure 6A shows a small potentiation when serotonin is applied with a delay of 5 s for an LTD-inducing protocol. Although this potentiation is not statistically significant, one might pose the question of how this will affect the behavior of the model. We find that at least in the context of the network trained here, this will not have a significant effect because at long delays, the net effect is still LTP. Once the network activity approaches the reward time, LTD will still dominate, resulting in stable learning.

We demonstrated here that reinforcement learning that is based on the competition between the LTP and the LTD traces, which is consistent with our experimental observations, stabilizes learning without the need to include additional reward-inhibiting mechanisms, as assumed previously (Gavornik et al., 2009; Rescorla and Wagner, 1972; Sutton and Barto, 1998).

DISCUSSION

Although it is well established that Hebbian plasticity can account for the remodeling of cortical networks during learning, it has been less clear how Hebbian plasticity can be recruited or gated by reward. We have provided direct physiological support for the theoretical concept of synaptic eligibility traces. We demonstrate that there are two eligibility traces, one for LTP and one for LTD, with different dynamics. The transformation of these transient traces into synaptic plasticity is accomplished by specific monoamine receptors that are anchored at the synapse. The existence of different traces for LTP and LTD may be a general phenomenon, because distinct traces are observable in both visual and prefrontal cortices. The different temporal dynamics of these two generate a self-stabilizing learning rule that allows the cortical network to perform a fundamental computation to learn the expected time of reward. We surmise that Hebbian induction of distinct eligibility traces for LTP and LTD, which can be transformed by specific monoamines, is a

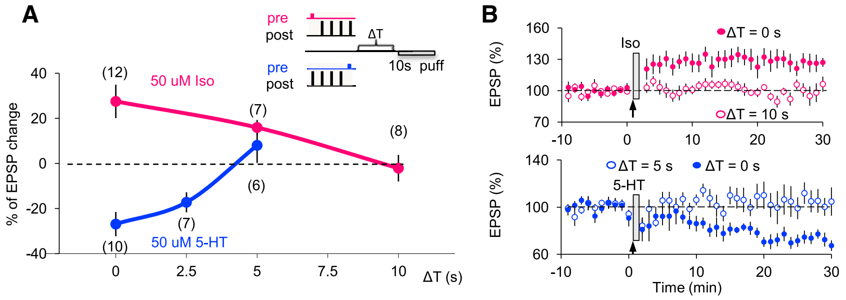

Figure 6. Eligibility Traces for LTP/D Are Transient and Have Different Durations

(A) Magnitude of synaptic changes (measured 30 min after conditioning) evoked when neuromodulators (50 μM Iso for LTP: magenta line and symbols; 50 μM 5-HT for LTD: blue line and symbols) were puffed after the ST conditioning at the specified delays (Δ*t*, in seconds, delay as described in the top right insert). The duration was less than 10 s for the LTP eligibility trace and less than 5 s for the LTD eligibility trace.

(B) Significant LTP (filled magenta circles, top panel) or LTD (filled blue circles, bottom panel) was induced when neuromodulators were puffed immediately after the ST pairings. There was no change in EPSP slope when puffing Iso 10 s after the ST pairings (open magenta circles, top panel) or 5-HT 5 s after the ST pairings (open blue circle, bottom panel).

simple and attractive mechanism that would allow cortical circuits to learn what stimuli and actions predict reward.

The molecular details of eligibility traces remain to be determined. A plausible scenario is that the traces reflect residual activity of kinases and phosphatases that gate AMPA receptor (AMPAR) trafficking in and out of the synapse and that neuromodulators, by phosphorylating AMPARs, are crucial to complement or enhance this process (Huang et al., 2012; Seol et al., 2007). Consistent with this idea, the decay of the LTP trace roughly matches the decay of CaMKII activity at pyramidal cell synapses (Lee et al., 2009). The present results also agree with our previous observation that GPCRs act downstream of NMDA receptor (NMDAR) activation to prime subsequent STDP induction in a pull-push manner, with Gs-coupled receptors promoting LTP over LTD and Gq-coupled receptors promoting LTD over LTP (Huang et al., 2012; Seol et al., 2007). Consistent with this pull-push model, β_2ARs and 5-HT$_{2C}$Rs in the visual cortex, which specifically transform the traces for LTP and LTD, are coupled to Gs and Gq, respectively. Notably, however, while prolonged stimulation of multiple GPCRs can prime LTP and LTD, their corresponding traces are transformed only by β_2ARs and 5-HT$_{2C}$Rs, which are anchored to the synapse. Moreover, brief stimulation of these two receptors can transform previously induced traces but does not promote subsequent plasticity. Thus, our present findings extend the pull-push model, because the anterograde and retrograde actions of the neuromodulators both follow the Gs or Gq rule for LTP or LTD induction. At the same time, the present results reveal that the spatiotemporal profile of neuromodulator activation dictates whether they can support priming or transformation of plasticity.

The principles uncovered in the visual cortex were confirmed in the prefrontal cortex, suggesting that transformation of LTP and LTD traces occurs throughout the cortex, although the specific supporting Gs- and Gq-coupled receptors may vary among cortical regions and layers. For example, DA can convert LTP traces in the frontal but not the visual cortex, and in the visual cortex, acetylcholine puffs can reward input activity in layer V cells (Chubykin et al., 2013) but not layer II/III cells (Figure 1). These discrepancies can be simply explained by the synaptic anchoring of different GCPRs in these cells, although we cannot rule out more complex scenarios related to different mechanisms of synaptic plasticity (Wang and Daw, 2003). A general mechanism of trace transformation is also consistent with the retrograde action of octopamine on STDP in insect olfactory learning (Cassenaer and Laurent, 2012) and with the recent report that in the striatum, Gs-coupled D1 receptors promote structural plasticity akin to LTP in synapses previously conditioned in a Hebbian manner (Yagishita et al., 2014). These previous studies only showed a single eligibility trace, and it remains unclear whether two independent traces are a general phenomenon that applies to these specific systems.

In contrast to previous theories focusing on a single plasticity trace, we uncover distinct and independent traces for LTP and LTD. The observation that the decay of the LTD eligibility trace is about twice as fast as the decay of the LTP trace was initially surprising, because theoretical considerations of unsupervised STDP in neural networks indicate that a larger window for LTD induction confers stability to learning in neural networks (Kempter et al., 2001; Song et al., 2000). To obtain stability, theories of reinforcement learning typically require an additional stopping rule (Gavornik et al., 2009; Rescorla and Wagner, 1972; Sutton and Barto, 1998), which at the physiological level is usually interpreted as inhibition of a reward nucleus. We demonstrated that because of the competition between the two eligibility traces, neural firing in cells within the network naturally stop before the reward time without the need for inhibition of reward. This stability is obtained not because of competition among the different neuromodulators (Boureau and Dayan, 2011) but because of temporal competition between synaptic eligibility traces with different dynamics, and it could in principle be accomplished even if the same neuromodulator was responsible for converting both traces. Such neural dynamics, as observed in vivo (Shuler and Bear, 2006), can enable a cortical network to perform the behaviorally important task of predicting reward times. It would be of interest to explore whether the properties of the two independent eligibility traces, besides predicting timing, can enable learning about other attributes of the reward, like quality and quantity, that are essential for decision making.

EXPERIMENTAL PROCEDURES

Animals

All procedures were approved by the Institutional Animal Care and Use Committee at Johns Hopkins University. *TH-ChR2* mice were produced by crossing

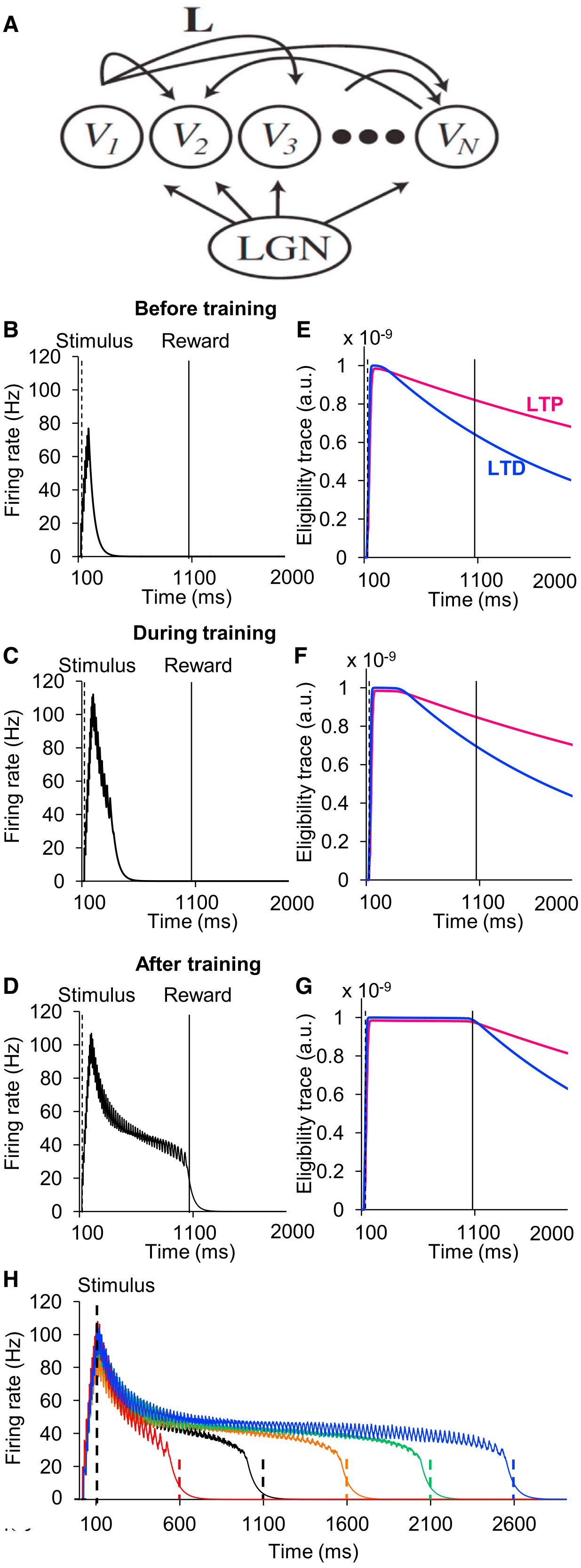

Figure 7. Competition between LTP and LTD Eligibility Traces Results in Stable Reinforcement Learning

(A) Diagram of a recurrent network of excitatory neurons representing cells in the visual cortex driven by feed-forward input from the LGN.

(B–D) Simulated average population firing rate computed from a recurrent network of 100 integrate-and-fire excitatory units. The network is trained to report a 1 s interval after a 100 ms stimulation. Three instances of network dynamics are shown: (B) before training, (C) during training (18 trials), and (D) after training (70 trials).

(E–G) Time evolution of LTP- and LTD-promoting eligibility traces corresponding to the same trials as in (B)–(D). Magenta lines are LTP eligibility traces, and blue lines are LTD eligibility traces. LTP and LTD eligibility traces both increase during the period of network activity (described earlier). LTD traces saturate at higher effective levels. At the beginning of training (E), LTP traces are larger than LTD traces at the time of reward; therefore, LTP is expressed. At the end of training (G), LTP and LTD traces are equal, resulting in no net change in synaptic efficacy.

(H) The model can be trained to predict different reward timings accurately.

See also Figure S5.

THicre homozygote (provided by Dr. Jeremy Nathan) with Floxed-ChR2 (B6; 129S-*Gt(ROSA)26Sor*$^{tm32(CAG-COP4*H134R/EYFP)Hze}$/J, used for data in Figure 3A and Figure S2B, or B6.Cg-*Gt(ROSA)26Sor*$^{tm27.1(CAG-COP4*H134R/tdTomato)Hze}$/J, used in Figure 3B and Figures S2C and S2D (The Jackson Laboratory). A Tph2-ChR2 (B6;SJL-Tg(Tph2-COP4*H134R/EYFP)5Gfng/J) heterozygote breeding pair was purchased from The Jackson Laboratory. Mice used for Figures S1C and S1D were intraperitoneally injected with reserpine (5 mg/kg) 23–24 hr before the experiment. All mice used were bred on a C57BL/6J background and were used at the age of P25–P45, when both LTP and LTD are expressed postsynaptically (Seol et al., 2007).

Slice Preparation

Coronal brain slices containing either the visual or the frontal cortex (300 μm thick) from C57BL/6J or transgenic mice (P25–P45) were prepared as described (Huang et al., 2012). Briefly, slices were cut in ice-cold dissection buffer containing 212.7 mM sucrose, 5 mM KCl, 1.25 mM NaH_2PO_4, 10 mM $MgCl_2$, 0.5 mM $CaCl_2$, 26 mM $NaHCO_3$, and 10 mM dextrose, bubbled with 95% O_2/5% CO_2 (pH 7.4). Slices were transferred to normal artificial cerebrospinal fluid (similar to the dissection buffer except that sucrose is replaced by 119 mM NaCl, $MgCl_2$ is lowered to 1 mM, and $CaCl_2$ is raised to 2 mM) and incubated at 30°C for 30 min and then at room temperature for at least 30 min before recording.

Whole-Cell Current Clamp Recording

Visualized whole-cell recordings were made from layer II/III (>35% depth from the pia) regular-spiking pyramidal neurons. Glass pipette recording electrodes (3–5 MΩ) were filled with solution containing 130 mM (K) gluconate, 10 mM KCl, 0.2 mM EGTA, 10 mM HEPES, 4 mM (Mg) ATP, 0.5 mM (Na) guanosine triphosphate, and 10 mM (Na) phosphocreatine (pH 7.2–7.3, 280–290 mOsm). Only cells with membrane potentials of less than −65 mV, series resistance < 25 MΩ, and input resistance > 85 MΩ were recorded. Cells were discarded if any of these values changed more than 25% during the experiment. Data were filtered at 10 kHz and digitized at 10 kHz using Igor Pro (WaveMetrics).

Electrical Stimulation and Induction of Plasticity

Synaptic responses were evoked in two independent pathways at 0.05 Hz by either alternating or consecutive (300 ms apart) paired-pulse stimulations (0.2 ms, 10–100 μA, 50 ms interval) through two concentric bipolar electrodes (125 μm diameter; FHC) placed ~300 μm apart in the middle of the cortical thickness. Stimulus intensity was adjusted to evoke simple-waveform (2–8 mV), short-onset latency (<4 ms) monosynaptic EPSPs. Input independence was confirmed by the absence of paired-pulse interactions. ST conditioning consisted of 200 pairings (one presynaptic stimulation given either 10 ms before or 10 ms after four consecutive action potentials at 100 Hz in the postsynaptic neuron) delivered at 10 Hz. Action potentials were generated by injecting a 1.2–1.6 nA current for 2 ms. Pairings were followed by one of the following manipulations: a 10 s puff (1–6 psi) of a neuromodulator (Picospritzer, Parker Instrumentation), 50 UV light pulses (Thorlabs; 365 nm light-emitting diode, or LED; 100 ms duration) delivered through the 40X objective at 5 Hz to uncage 4,5-dimethoxy-2-nitrobenzyl adenosine (DMNB)-caged cAMP (Invitrogen), or trains of blue light pulses (Thorlabs 455 nm LED, 10 ms duration) delivered at 10 Hz for 10 s (Figure 3) or 1 s (Figure 4) to activate ChR2. Pairing LTP or LTD in Figure S4 was induced by 150 pairings of presynaptic stimulation with postsynaptic depolarization to 0 or −40 mV, respectively, at 0.75 Hz (each depolarization lasted for 666 ms; presynaptic stimulation was given 100 ms after the onset of depolarization). Pairing LTP or LTD in reserpine-injected mice (Figures S1C and S1D) was induced by pairing 10 Hz presynaptic stimulation with 20 s of postsynaptic depolarization from −70 to −10 mV for LTP and to −40 mV for LTD, with or without 10 s of neuromodulator puffing. The synaptic strength was quantified by measuring the initial slope of the EPSPs.

Iso hydrochloride (50 μM), methoxamine hydrochloride (50 μM), carbamoylcholine chloride (10–500 μM), NE bitartrate (10–50 μM), and ketanserine tartrate salt (1 μM) were purchased from Sigma. Serotonin hydrochloride (5-HT, 50 μM), DA hydrochloride (50 μM), RS 102221 hydrochloride (1 μM), ICI 118,551 hydrochloride (1 μM), and reserpine (5 mg/kg, in 1.5% acetic acid) were purchased from Tocris. DMNB-caged cAMP (100 μM) was purchased from Invitrogen. The membrane-permeable peptide DSPL (11R-QGRNSNTNDSPL) and its active analog DAPA (11R-QGRNSNTNDAPA) were gifts from J.W.H. Synthetic peptides (5-HT$_{2C}$-Ct, VNPSSVVSERISSV; 5-HT$_{2CSSA}$-Ct, VNPSSVVSERISSA, >98% purity) were purchased from GenScript.

Biocytin Staining and Imaging

For imaging locus coeruleus noradrenergic neurons, 5-week-old TH-ChR2 mice were transcardially perfuse with fresh paraformaldehyde (PFA, 4%). Brains were removed and fixed overnight in PFA before being transferred to a sterile solution of 30% sucrose in PBS (pH 7.4) for at least 12 hr. The fixed brain was sectioned into 40 μm coronal slices using a freezing microtome (Leica) and kept at −20°C until use. For imaging-recorded neurons from acute cortical slices of TH-ChR2 mice, biocytin was included into the recording pipette. After recording, slices were fixed in 10% formalin at least overnight before being rinsed in 0.1 M PBS (2x 10 min). Slices were then permeabilized (2% Triton X-100 in 0.1 M PBS) for 1 hr before incubation with 1 μg/ml streptavidin-488 (in 0.1 M PBS containing 1% Triton X-100) overnight at 4°C. Slices were rinsed with 0.1 M PBS (2x 10 min) before being mounted on a glass slide.

Confocal images were taken on a Zeiss laser stimulated microscope 510 with the following objective lenses: 10X/0.45, 20X/0.75, and 40X/1.2.

Data Analysis

Data were analyzed using a custom program (Igor). Data were averaged over the last 5 min of post-induction time and normalized to the last 5 min of baseline, and the Wilcoxon rank-sum test was used for independent data. One-way ANOVAs followed by Tukey's honest significant difference post hoc tests were used to compare the means of more than two samples. Differences were considered to be significant when $p < 0.05$.

Mathematical Model

Learning Rules

Simulations were performed on a recurrent network of excitatory neurons consisting of 100 integrate-and-fire units with all-to-all lateral connections. The network was driven by feed-forward excitatory input representing incoming spikes from the lateral geniculate nucleus (LGN). Model equations describing the dynamics of the neurons are as in Gavornik et al. (2009), except for the learning rule that updates the changes of synaptic weights of the lateral connections. The prolonged network dynamics are due to the positive feedback from lateral connections, and the strength of synaptic efficacies (denoted by the matrix L) determines the duration of activity in the network.

In the current model, two synaptic eligibility traces (previously referred to as proto-weights) (Gavornik et al., 2009), mediating LTP ($T^p{}_{ij}$) and LTD ($T^d{}_{ij}$) separately, evolve in time according to a pair of ordinary differential equations of the form

$$\tau_p \frac{dT^P_{ij}}{dt} = -T^p_{ij} + \epsilon H^p(R_i, R_j)\left(T^p_{max} - T^p_{ij}\right) \qquad \text{(Equation 1)}$$

$$\tau_d \frac{dT^P_{ij}}{dt} = -T^d_{ij} + \epsilon H^d(R_i, R_j)\left(T^d_{max} - T^d_{ij}\right), \qquad \text{(Equation 2)}$$

where τ_p and τ_d are the decay time constants of the corresponding LTP and LTD traces, respectively, and $H^p(R_i,R_j)$ and $H^d(R_i,R_j)$ are Hebbian terms, which in general are different for each trace and can include the effects of the pre- and postsynaptic spike ordering. In the present model, we used the simplest assumption, considering that both Hebbian terms are identical and depend on a product of time-dependent firing rates of postsynaptic (R_i) and presynaptic (R_j) neurons, as in Gavornik et al. (2009). The firing rates are temporal averages computed using an exponential window with a 50 ms decay constant. Each synaptic trace can saturate at a different level, and these levels are determined by the quantities $T^d{}_{max}$ and $T^p{}_{max}$. Finally, ε is a factor scaling the Hebbian term.

We chose a simple rule for updating the synaptic weights, which depends on the difference between these traces and on the delivery of reward:

$$\frac{dL_{ij}}{dt} = \eta\left(T^p_{ij} - T^d_{ij}\right)\delta(t - t_{reward}), \qquad \text{(Equation 3)}$$

where L_{ij} is the magnitude of the synaptic weight between neurons i (postsynaptic) and j (presynaptic), η is the learning rate, and the delta function term indicates that the changes occur at the time of reward (t_{reward}) when neurotransmitter is released. This delta function can easily be replaced by a narrow function near the reward time, representing the presence of a neuromodulator. All these equations were chosen to be as simple as possible rather than to be biophysically precise.

The model assumes a reward signal at time t_{reward} and does not distinguish between the two neuromodulators. By doing this, we implicitly assume that the actual reward activates both neuromodulators simultaneously. One could write a more complex equation with two different neuromodulators acting independently on the two different traces; for our implementation here it would not matter, but it could be useful if we are to consider situations in which one neuromodulator is active and the other is not.

Recurrent Network

The recurrent network is constructed as in Gavornik et al. (2009), and only the learning rule is modified. Each neuron is a conductance-based integrate-and-fire unit following the equations

$$C\frac{d\nu_i}{dt} = g_L(E_L - \nu_i) + g_{E,i}(E_E - \nu_i)$$

and

$$\frac{s_k}{dt} = -\frac{1}{\tau_s}s_k + \rho(1 - s_k)\sum_j \delta\left(t - t^k_j\right), \qquad \text{(Equation 4)}$$

where υ_i represents the membrane potential of the ith neuron, which in this simple model is excitatory (E), and s_k is the synaptic activation of the kth presynaptic neuron. Other parameters are membrane capacitance C; leak and excitatory conductances g_L and $g_{E,i}$, respectively; leak and excitatory reversal potentials E_L and E_E, respectively; percentage change of synaptic activation with input spikes ρ; and time constant for synaptic activation τ_s. The neuron fires an action potential once it reaches threshold (υ_{th}), $\upsilon_i = \upsilon_{th}$, and the membrane potential is then reset to υ_{rest}. The delta function in Equation 4 indicates that these changes occur only at the moment of the arrival of a presynaptic spike at $t^k{}_j$, where the index j indicates that this is the jth spike in neuron k and where $g_{E,i}$ is as follows:

$$g_{E,i} = \sum_k L_{ik}s_k.$$

All parameter values are as in Gavornik et al. (2009).

Equation Derivation

Here we present the derivation of the equation in Figure S5E. After training, network activity decays almost fully before the reward signal is delivered.

The difference between the time that the network decays below a threshold and the reward time is defined as *D* (Figure S5D). The value of *D* can be approximated based on the observation that fixed points are obtained when the two eligibility traces are equal (Equation 3). To calculate this, we make the following approximations: we assume that the network is either fully active or inactive and that when it is fully active both traces are saturated. Combining these crude approximations with Equations 1 and 2, we observe:

$$T^p_{max} e^{-D/\tau_p} = T^d_{max} e^{-D/\tau_d},$$

which can be solved for *D* to yield the following:

$$D = log\left(T^d_{max}/T^p_{max}\right)\frac{\tau_p \tau_d}{\tau_p - \tau_d}.$$

In Figure S5E, this approximate formula is compared to simulation results and yields good agreement, at least for these biophysically plausible parameter ranges.

SUPPLEMENTAL INFORMATION

Supplemental Information includes five figures and can be found with this article online at http://dx.doi.org/10.1016/j.neuron.2015.09.037.

AUTHOR CONTRIBUTIONS

Conceptualization, K.H., H.S., and A.K.; investigation, K.H., M.H., S.Z.H., and X.T.; writing, K.H., H.S., and A.K.; funding acquisition, H.S. and A.K.; resources, J.W.H.

ACKNOWLEDGMENTS

We thank Dr. H.K. Lee for insightful discussions and Drs. M. Bridi and J. Lucas Whitt for comments on the manuscript. Supported by NIH grants R01MH093665 to H.S. (P.I. Shuler) and R01EY012124 to A.K. and by an SLI grant to A.K.

Received: July 7, 2015
Revised: September 1, 2015
Accepted: September 18, 2015
Published: October 22, 2015

REFERENCES

Becamel, C., Figge, A., Poliak, S., Dumuis, A., Peles, E., Bockaert, J., Lubbert, H., and Ullmer, C. (2001). Interaction of serotonin 5-hydroxytryptamine type 2C receptors with PDZ10 of the multi-PDZ domain protein MUPP1. J. Biol. Chem. *276*, 12974–12982.

Bécamel, C., Gavarini, S., Chanrion, B., Alonso, G., Galéotti, N., Dumuis, A., Bockaert, J., and Marin, P. (2004). The serotonin 5-HT$_{2A}$ and 5-HT$_{2C}$ receptors interact with specific sets of PDZ proteins. J. Biol. Chem. *279*, 20257–20266.

Boureau, Y.-L., and Dayan, P. (2011). Opponency revisited: competition and cooperation between dopamine and serotonin. Neuropsychopharmacology *36*, 74–97.

Caporale, N., and Dan, Y. (2008). Spike timing-dependent plasticity: a Hebbian learning rule. Annu. Rev. Neurosci. *31*, 25–46.

Cassenaer, S., and Laurent, G. (2012). Conditional modulation of spike-timing-dependent plasticity for olfactory learning. Nature *482*, 47–52.

Choi, S.-Y., Chang, J., Jiang, B., Seol, G.-H., Min, S.-S., Han, J.-S., Shin, H.S., Gallagher, M., and Kirkwood, A. (2005). Multiple receptors coupled to phospholipase C gate long-term depression in visual cortex. J. Neurosci. *25*, 11433–11443.

Chubykin, A.A., Roach, E.B., Bear, M.F., and Shuler, M.G. (2013). A cholinergic mechanism for reward timing within primary visual cortex. Neuron *77*, 723–735.

Crow, T.J. (1968). Cortical synapses and reinforcement: a hypothesis. Nature *219*, 736–737.

Davare, M.A., Avdonin, V., Hall, D.D., Peden, E.M., Burette, A., Weinberg, R.J., Horne, M.C., Hoshi, T., and Hell, J.W. (2001). A beta2 adrenergic receptor signaling complex assembled with the Ca2+ channel Cav1.2. Science *293*, 98–101.

Edelmann, E., and Lessmann, V. (2013). Dopamine regulates intrinsic excitability thereby gating successful induction of spike timing-dependent plasticity in CA1 of the hippocampus. Front. Neurosci. *7*, 25.

Frémaux, N., Sprekeler, H., and Gerstner, W. (2010). Functional requirements for reward-modulated spike-timing-dependent plasticity. J. Neurosci. *30*, 13326–13337.

Gardner, M.P.H., and Fontanini, A. (2014). Encoding and tracking of outcome-specific expectancy in the gustatory cortex of alert rats. J. Neurosci. *34*, 13000–13017.

Gavarini, S., Bécamel, C., Altier, C., Lory, P., Poncet, J., Wijnholds, J., Bockaert, J., and Marin, P. (2006). Opposite effects of PSD-95 and MPP3 PDZ proteins on serotonin 5-hydroxytryptamine 2C receptor desensitization and membrane stability. Mol. Biol. Cell *17*, 4619–4631.

Gavornik, J.P., and Shouval, H.Z. (2011). A network of spiking neurons that can represent interval timing: mean field analysis. J. Comput. Neurosci. *30*, 501–513.

Gavornik, J.P., Shuler, M.G.H., Loewenstein, Y., Bear, M.F., and Shouval, H.Z. (2009). Learning reward timing in cortex through reward dependent expression of synaptic plasticity. Proc. Natl. Acad. Sci. USA *106*, 6826–6831.

Goltstein, P.M., Coffey, E.B.J., Roelfsema, P.R., and Pennartz, C.M. (2013). In vivo two-photon Ca2+ imaging reveals selective reward effects on stimulus-specific assemblies in mouse visual cortex. J. Neurosci. *33*, 11540–11555.

Huang, S., Treviño, M., He, K., Ardiles, A., Pasquale, Rd., Guo, Y., Palacios, A., Huganir, R., and Kirkwood, A. (2012). Pull-push neuromodulation of LTP and LTD enables bidirectional experience-induced synaptic scaling in visual cortex. Neuron *73*, 497–510.

Huang, S., Rozas, C., Treviño, M., Contreras, J., Yang, S., Song, L., Yoshioka, T., Lee, H.K., and Kirkwood, A. (2014). Associative Hebbian synaptic plasticity in primate visual cortex. J. Neurosci. *34*, 7575–7579.

Hull, C.L. (1943). Principles of Behavior: An Introduction to Behavior Theory (Appleton-Century).

Izhikevich, E.M. (2007). Solving the distal reward problem through linkage of STDP and dopamine signaling. Cereb. Cortex *17*, 2443–2452.

Jaramillo, S., and Zador, A.M. (2011). The auditory cortex mediates the perceptual effects of acoustic temporal expectation. Nat. Neurosci. *14*, 246–251.

Joiner, M.L., Lisé, M.F., Yuen, E.Y., Kam, A.Y., Zhang, M., Hall, D.D., Malik, Z.A., Qian, H., Chen, Y., Ulrich, J.D., et al. (2010). Assembly of a beta2-adrenergic receptor—GluR1 signalling complex for localized cAMP signalling. EMBO J. *29*, 482–495.

Kahnt, T., Grueschow, M., Speck, O., and Haynes, J.-D. (2011). Perceptual learning and decision-making in human medial frontal cortex. Neuron *70*, 549–559.

Kempter, R., Gerstner, W., and van Hemmen, J.L. (2001). Intrinsic stabilization of output rates by spike-based Hebbian learning. Neural Comput. *13*, 2709–2741.

Kirkwood, A., Rozas, C., Kirkwood, J., Perez, F., and Bear, M.F. (1999). Modulation of long-term synaptic depression in visual cortex by acetylcholine and norepinephrine. J. Neurosci. *19*, 1599–1609.

Klopf, A.H. (1982). The Hedonistic Neuron: A Theory of Memory, Learning, and Intelligence (Hemisphere/Taylor & Francis).

Lee, S.-J.R., Escobedo-Lozoya, Y., Szatmari, E.M., and Yasuda, R. (2009). Activation of CaMKII in single dendritic spines during long-term potentiation. Nature *458*, 299–304.

Liu, C.-H., Coleman, J.E., Davoudi, H., Zhang, K., and Shuler, M.G.H. (2015). Selective activation of a putative reinforcement signal conditions cued interval timing in primary visual cortex. Curr. Biol. *25*, 1551–1561.

Otmakhova, N.A., and Lisman, J.E. (1996). D1/D5 dopamine receptor activation increases the magnitude of early long-term potentiation at CA1 hippocampal synapses. J. Neurosci. *16*, 7478–7486.

Poort, J., Khan, A.G., Pachitariu, M., Nemri, A., Orsolic, I., Krupic, J., Bauza, M., Sahani, M., Keller, G.B., Mrsic-Flogel, T.D., and Hofer, S.B. (2015). Learning enhances sensory and multiple non-sensory representations in primary visual cortex. Neuron *86*, 1478–1490.

Qian, H., Matt, L., Zhang, M., Nguyen, M., Patriarchi, T., Koval, O.M., Anderson, M.E., He, K., Lee, H.-K., and Hell, J.W. (2012). β_2-Adrenergic receptor supports prolonged theta tetanus-induced LTP. J. Neurophysiol. *107*, 2703–2712.

Rescorla, R.A., and Wagner, A.R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In Classical Conditioning II: Current Research and Theory, A.H. Black and W.F. Prokasy, eds. (Appleton-Century), pp. 64–99.

Richards, B.A., Aizenman, C.D., and Akerman, C.J. (2010). In vivo spike-timing-dependent plasticity in the optic tectum of *Xenopus laevis*. Front. Synaptic Neurosci. *2*, 7.

Ridderinkhof, K.R., van den Wildenberg, W.P.M., Segalowitz, S.J., and Carter, C.S. (2004). Neurocognitive mechanisms of cognitive control: the role of prefrontal cortex in action selection, response inhibition, performance monitoring, and reward-based learning. Brain Cogn. *56*, 129–140.

Rushworth, M.F.S., Noonan, M.P., Boorman, E.D., Walton, M.E., and Behrens, T.E. (2011). Frontal cortex and reward-guided learning and decision-making. Neuron *70*, 1054–1069.

Seitz, A.R., Kim, D., and Watanabe, T. (2009). Rewards evoke learning of unconsciously processed visual stimuli in adult humans. Neuron *61*, 700–707.

Seol, G.H., Ziburkus, J., Huang, S., Song, L., Kim, I.T., Takamiya, K., Huganir, R., Lee, H.K., and Kirkwood, A. (2007). Neuromodulators control the polarity of spike-timing-dependent synaptic plasticity. Neuron *55*, 919–929.

Shuler, M.G., and Bear, M.F. (2006). Reward timing in the primary visual cortex. Science *311*, 1606–1609.

Song, S., Miller, K.D., and Abbott, L.F. (2000). Competitive Hebbian learning through spike-timing-dependent synaptic plasticity. Nat. Neurosci. *3*, 919–926.

Sutton, R.S., and Barto, A.G. (1998). Reinforcement learning: an introduction. IEEE Trans. Neural Netw. *9*, 1054.

Turner, P.R., O'Connor, K., Tate, W.P., and Abraham, W.C. (2003). Roles of amyloid precursor protein and its fragments in regulating neural activity, plasticity and memory. Prog. Neurobiol. *70*, 1–32.

Wang, X.F., and Daw, N.W. (2003). Long term potentiation varies with layer in rat visual cortex. Brain Res. *989*, 26–34.

Weber, E.T., and Andrade, R. (2010). Htr2a gene and 5-HT(2A) receptor expression in the cerebral cortex studied using genetically modified mice. Front. Neurosci. *4*, 1–12.

Wörgötter, F., and Porr, B. (2005). Temporal sequence learning, prediction, and control: a review of different models and their relation to biological mechanisms. Neural Comput. *17*, 245–319.

Yagishita, S., Hayashi-Takagi, A., Ellis-Davies, G.C., Urakubo, H., Ishii, S., and Kasai, H. (2014). A critical time window for dopamine actions on the structural plasticity of dendritic spines. Science *345*, 1616–1620.

Yang, K., and Dani, J.A. (2014). Dopamine D1 and D5 receptors modulate spike timing-dependent plasticity at medial perforant path to dentate granule cell synapses. J. Neurosci. *34*, 15888–15897.

Zhao, S., Ting, J.T., Atallah, H.E., Qiu, L., Tan, J., Gloss, B., Augustine, G.J., Deisseroth, K., Luo, M., Graybiel, A.M., et al. (2011). Cell type-specific channelrhodopsin-2 transgenic mice for optogenetic dissection of neural circuitry function. Nat. Methods *8*, 745–752.