frontiers in Neural Circuits

REVIEW
published: 31 July 2018
doi: 10.3389/fncir.2018.00053

Eligibility Traces and Plasticity on Behavioral Time Scales: Experimental Support of NeoHebbian Three-Factor Learning Rules

Wulfram Gerstner, Marco Lehmann, Vasiliki Liakoni, Dane Corneil and Johanni Brea*

School of Computer Science and School of Life Sciences, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

Most elementary behaviors such as moving the arm to grasp an object or walking into the next room to explore a museum evolve on the time scale of seconds; in contrast, neuronal action potentials occur on the time scale of a few milliseconds. Learning rules of the brain must therefore bridge the gap between these two different time scales. Modern theories of synaptic plasticity have postulated that the co-activation of pre- and postsynaptic neurons sets a flag at the synapse, called an eligibility trace, that leads to a weight change only if an additional factor is present while the flag is set. This third factor, signaling reward, punishment, surprise, or novelty, could be implemented by the phasic activity of neuromodulators or specific neuronal inputs signaling special events. While the theoretical framework has been developed over the last decades, experimental evidence in support of eligibility traces on the time scale of seconds has been collected only during the last few years. Here we review, in the context of three-factor rules of synaptic plasticity, four key experiments that support the role of synaptic eligibility traces in combination with a third factor as a biological implementation of neoHebbian three-factor learning rules.

Keywords: eligibility trace, hebb rule, reinforcement learning, neuromodulators, surprise, synaptic tagging, synaptic plasticity, behavioral learning

OPEN ACCESS

Edited by:
Edward S. Ruthazer,
McGill University, Canada

Reviewed by:
Blake A. Richards,
University of Toronto Scarborough, Canada
Joel Zylberberg,
University of Colorado Anschutz Medical Campus, United States

****Correspondence:***
Wulfram Gerstner
wulfram.gerstner@epfl.ch

Received: *11 January 2018*
Accepted: *19 June 2018*
Published: *31 July 2018*

Citation:
Gerstner W, Lehmann M, Liakoni V, Corneil D and Brea J (2018) Eligibility Traces and Plasticity on Behavioral Time Scales: Experimental Support of NeoHebbian Three-Factor Learning Rules. Front. Neural Circuits 12:53. doi: 10.3389/fncir.2018.00053

1. INTRODUCTION

Humans are able to learn novel behaviors such as pressing a button, swinging a tennis racket, or braking at a red traffic light; they are also able to form memories of salient events, learn to distinguish flowers, and to establish a mental map when exploring a novel environment. Memory formation and behavioral learning is linked to changes of synaptic connections (Martin et al., 2000). Long-lasting synaptic changes, necessary for memory, can be induced by Hebbian protocols that combine the activation of presynaptic terminals with a manipulation of the voltage or the firing state of the postsynaptic neuron (Lisman, 2003). Traditional experimental protocols of long-term potentiation (LTP) (Bliss and Lømo, 1973; Bliss and Collingridge, 1993), long-term depression (LTD) (Levy and Stewart, 1983; Artola and Singer, 1993) and spike-timing dependent plasticity (STDP) (Markram et al., 1997; Zhang et al., 1998; Sjöström et al., 2001) neglect that additional factors such as neuromodulators or other gating signals might be necessary to permit synaptic changes (Gu, 2002; Reynolds and Wickens, 2002; Hasselmo, 2006). Early STDP experiments that involved neuromodulators mainly focused on tonic bath application of modulatory factors (Pawlak et al., 2010) with the exception of one study in locusts Cassenaer and Laurent (2012). However, from the perspective of formal learning theories, to be reviewed below, the timing of modulatory factors is just as crucial (Schultz and Dickinson, 2000; Schultz, 2002). From the theoretical perspective, STDP under the control of neuromodulators leads to the framework of three-factor learning rules (Xie and Seung, 2004; Legenstein et al., 2008; Vasilaki et al., 2009) where an eligibility trace represents the Hebbian idea of co-activation of pre- and postsynaptic neurons (Hebb, 1949) while modulation of plasticity by additional gating signals is represented generically by a "third factor" (Crow, 1968; Barto, 1985; Legenstein et al., 2008). Such a third factor could represent variables such as "reward minus expected reward" (Williams, 1992; Schultz, 1998; Sutton and Barto, 1998) or the saliency of an unexpected event (Ljunberg et al., 1992; Redgrave and Gurney, 2006).

In an earlier paper (Frémaux and Gerstner, 2016) we reviewed the theoretical literature of, and experimental support for, three-factor rules available by the end of 2013. During recent years, however, the experimental procedures advanced significantly and provided direct physiological evidence of eligibility traces and three-factor learning rules for the first time, making an updated review of three-factor rules necessary. In the following, we—a group of theoreticians—review five experimental papers indicating support of eligibility traces in striatum (Yagishita et al., 2014), cortex (He et al., 2015), and hippocampus (Brzosko et al., 2015, 2017; Bittner et al., 2017). We will close with a few remarks on the paradoxical nature of theoretical predictions in the field of computational neuroscience.

2. HEBBIAN RULES VS. THREE-FACTOR RULES

Learning rules describe the change of the strength of a synaptic contact between a presynaptic neuron j and a postsynaptic neuron i. The strength of an excitatory synaptic contact can be defined by the amplitude of the postsynaptic potential which is closely related to the spine volume and the number of AMPA receptors (Matsuzaki et al., 2001). Synapses contain complex molecular machineries (Lisman, 2003, 2017; Redondo and Morris, 2011; Huganir and Nicoll, 2013), but for the sake of transparency of the arguments, we will keep the mathematical notation as simple as possible and characterize the synapse by two variables only: the first one is the synaptic strength w_{ij}, measured as spine volume or amplitude of postsynaptic potential, and the second one is a synapse-internal variable e_{ij} which is not directly visible in standard electrophysiological experiments. In our view, the internal variable e_{ij} represents a metastable transient state of interacting molecules in the spine head or a multi-molecular substructure in the postsynaptic density which serves as a synaptic flag indicating that the synapse is ready for an increase or decrease of its spine volume (Bosch et al., 2014). The precise biological nature of e_{ij} is not important to understand the theories and experiments that are reviewed below. We refer to e_{ij} as the "synaptic flag" or the "eligibility trace" and to w_{ij} as the "synaptic weight," or "strength" of the synaptic contact. A change of the synaptic flag indicates a 'candidate weight change' (Frémaux et al., 2010) whereas a change of w_{ij} indicates an actual, measurable, change of the synaptic weight. Before we turn to three-factor rules, let us discuss conventional models of Hebbian learning.

2.1. Hebbian Learning Rules

Hebbian learning rules are the mathematical summary of the outcome of experimental protocols inducing long-term potentiation (LTP) or long-term depression (LTD) of synapses. Suitable experimental protocols include strong extracellular stimulation of presynaptic fibers (Bliss and Lømo, 1973; Levy and Stewart, 1983), manipulation of postsynaptic voltage in the presence of presynaptic spike arrivals (Artola and Singer, 1993), or spike-timing dependent plasticity (STDP) (Markram et al., 1997; Sjöström et al., 2001). In all mathematical formulations of Hebbian learning, the synaptic flag variable e_{ij} is sensitive to the *combination* of presynaptic spike arrival and a postsynaptic variable, such as the voltage at the location of the synapse. Under a Hebbian learning rule, repeated presynaptic spike arrivals at a synapse of a neuron at rest do not cause a change of the synaptic variable. Similarly, an elevated postsynaptic potential in the absence of a presynaptic spike does not cause a change of the synaptic variable. Thus, Hebbian learning always needs two factors for a synaptic change: a factor caused by a presynaptic signal such as glutamate; and a factor that depends on the state of the postsynaptic neuron.

What are these factors? We can think of the presynaptic factor as the time course of glutamate available in the synaptic cleft or bound to the postsynaptic membrane. Note that the term "presynaptic factor" that we will use in the following *does not imply* that the physical location of the presynaptic factor is inside the presynaptic terminal–the factor could very well be located in the postsynaptic membrane as long as it *only* depends on the amount of available neurotransmitters. The postsynaptic factor might be related to calcium in the synaptic spine (Shouval et al., 2002; Rubin et al., 2005), a calcium-related second messenger molecule (Graupner and Brunel, 2007), or simply the voltage at the site of the synapse (Brader et al., 2007; Clopath et al., 2010).

We remind the reader that we always use the index j to refer to the presynaptic neuron and the index i to refer to the postsynaptic one. For the sake of simplicity, let us call the presynaptic factor x_j (representing the activity of the presynaptic neuron or the amount of glutamate in the synaptic cleft) and the postsynaptic factor y_i (representing the state of the postsynaptic neuron). In a Hebbian learning rule, changes of the synaptic flag e_{ij} need both x_j and y_i

$$\frac{d}{dt} e_{ij} = \eta\, f_j(x_j)\, g_i(y_i) - e_{ij}/\tau_e \qquad (1)$$

where η is the constant learning rate, τ_e is a decay time constant, $f_j(x_j)$ is an (often linear) function of the presynaptic activity x_j and $g_i(y_i)$ is some arbitrary, potentially nonlinear, function of the postsynaptic variable y_i. The index j of the function f_j and the index i of the function g_i indicate that the rules for changing a synaptic flag can depend on the type of pre- and postsynaptic neurons, on the cortical layer and area, but also on some heterogeneity of parameters between one neuron and the next.

According to Equation 1 the synaptic flag e_{ij} acts as a *correlation detector* between presynaptic activity x_j and the state of the postsynaptic neuron y_i. In some models, there is no decay or the decay is considered negligible on the time scale of one experiment ($\tau_e \rightarrow \infty$). The flag variable e_{ij} could be related to a calcium-based coincidence detection mechanism in the spine such as CaMKII (Lisman, 1989; Shouval et al., 2002) or a metastable state of the molecular machinery in the synapse (Bosch et al., 2014).

Let us discuss two examples. In the Bienenstock-Cooper Munro (BCM) model of developmental cortical plasticity (Bienenstock et al., 1982) the presynaptic factor x_j is the firing rate of the presynaptic neuron and $g(y_i) = (y_i - \theta)\, y_i$ is a quadratic function with y_i the postsynaptic firing rate and θ a threshold rate. Thus, if both pre- and postsynaptic neurons fire together at a high rate $x_j = y_i > \theta$ then the synaptic flag e_{ij} increases. In the BCM model, just like in most other conventional models, a change of the synaptic flag (i.e., an internal state of the synapse) leads instantaneously to a change of the weight $e_{ij} \longrightarrow w_{ij}$ so that an experimental protocol results immediately in a measurable weight change. With the BCM rule and other similar rules (Oja, 1982; Miller and MacKay, 1994), the synaptic weight increases if both presynaptic and postsynaptic neuron are highly active, implementing the slogan "fire together, wire together" (Lowel and Singer, 1992; Shatz, 1992) cf. **Figure 1Ai**.

As a second example, we consider the Clopath model (Clopath et al., 2010). In this model, there are two correlation detectors implemented as synaptic flags e_{ij}^{+} and e_{ij}^{-} for LTP and LTD, respectively. The synaptic flag e_{ij}^{+} for LTP uses a presynaptic factor x_j^{+} (related to the amount of glutamate available in the synaptic cleft) which increases with each presynaptic spike and decays back to zero over the time of a few milliseconds (Clopath et al., 2010). The postsynaptic factor for LTP depends on the postsynaptic voltage y_i via a function $g(y_i) = a_{+}[y_i - \theta_{+}]\bar{y}_i$ where a_{+} is a positive constant, θ_{+} a voltage threshold, square brackets denote the rectifying piecewise linear function, and $\bar{y}_i$ a running average of the voltage with a time constant of tens of milliseconds. An analogous, but simpler, combination of presynaptic spikes and postsynaptic voltage defines the second synaptic flag e_{ij}^{-} for LTD (Clopath et al., 2010). The total change of the synaptic weight is the combination of the two synaptic flags for LTP and LTD: $dw_{ij}/dt = de_{ij}^{+}/dt - de_{ij}^{-}/dt$. Note that, since both synaptic flags e_{ij}^{+} and e_{ij}^{-} depend on the postsynaptic voltage, postsynaptic spikes are not a necessary condition for changes, in agreement with voltage-dependent protocols (Artola and Singer, 1993; Ngezahayo et al., 2000). Thus, in voltage-dependent protocols, and similarly in voltage-dependent models, "wiring together" is possible without "firing together"-indicating that the theoretical framework sketched above goes beyond a narrow view of Hebbian learning; cf. **Figure 1Aii**.

If we restrict the discussion of the postsynaptic variable to super-threshold spikes, then the Clopath model becomes identical to the triplet STDP model (Pfister et al., 2006) which is in turn closely related to other nonlinear STDP models (Senn et al., 2001; Froemke and Dan, 2002; Izhikevich and Desai, 2003) as well as to the BCM model discussed above (Pfister et al., 2006; Gjorjieva et al., 2011). Classic pair-based STDP models (Gerstner et al., 1996; Kempter et al., 1999; Song et al., 2000; van Rossum et al., 2000; Rubin et al., 2001) are further examples of the general theoretical framework of Equation (1) and so are some models of structural plasticity (Helias et al., 2008; Deger et al., 2012, 2018; Fauth et al., 2015). Hebbian models of synaptic consolidation have several hidden flag variables (Fusi et al., 2005; Barrett et al., 2009; Benna and Fusi, 2016) but can also be situated as examples within the general framework of Hebbian rules. Note that in most of the examples so far the measured synaptic weight is a linear function of the synaptic flag variable(s). However, this does not need to be the case. For example, in some voltage-based (Brader et al., 2007) or calcium-based models (Shouval et al., 2002; Rubin et al., 2005), the synaptic flag is transformed into a weight change only if e_{ij} is above or below some threshold, or only after some further filtering.

To summarize, in the theoretical literature the class of Hebbian models is a rather general framework encompassing all those models that are driven by a combination of presynaptic activity and the state of the postsynaptic neuron. In this view, Hebbian models depend on two factors related to the activity of the presynaptic and the state of the postsynaptic neuron. The correlations between the two factors can be extracted on different time scales using one or, if necessary, several flag variables. The flag variables trigger a change of the measured synaptic weight. In the following we build on Hebbian learning, but extend the theoretical framework to include a third factor.

2.2. Three-Factor Learning Rules

We are interested in a framework where a Hebbian co-activation of two neurons leaves one or several flags (eligibility trace) at the synapse connecting these neurons. The flag is not directly visible and does not automatically trigger a change of the synaptic weight. An actual weight change is implemented only if a third signal, e.g., a phasic increase of neuromodulator activity or an additional input (signaling the occurrence of a special event) is present at the same time or in the near future. Theoreticians refer to such a plasticity model as a three-factor learning rule (Xie and Seung, 2004; Legenstein et al., 2008; Vasilaki et al., 2009; Frémaux et al., 2013; Frémaux and Gerstner, 2016). Three-factor rules have also been called "neoHebbian" (Lisman et al., 2011; Lisman, 2017) or "heterosynaptic (modulatory-input dependent)" (Bailey et al., 2000) and can be traced back to the 1960s (Crow, 1968), if not earlier. To our knowledge the wording "three factors" was first used by (Barto, 1985). The terms eligibility and eligibility traces have been used in (Klopf, 1972; Sutton and Barto, 1981, 1998; Barto et al., 1983; Barto, 1985; Williams, 1992; Schultz, 1998) but in some of the early studies it remained unclear whether eligibility traces can be set by presynaptic activity alone (Klopf, 1972; Sutton and Barto, 1981) or only by Hebbian co-activation of pre- and postsynaptic neurons (Barto et al., 1983; Barto, 1985; Williams, 1992; Schultz, 1998; Sutton and Barto, 1998).

The basic idea of a modern eligibility trace is that a synaptic flag variable e_{ij} is set according to Equation (1) by coincidences between presynaptic activity x_j and a postsynaptic factor y_i. The update of the synaptic weight w_{ij}, as measured via the spine volume or the amplitude of the excitatory postsynaptic potential

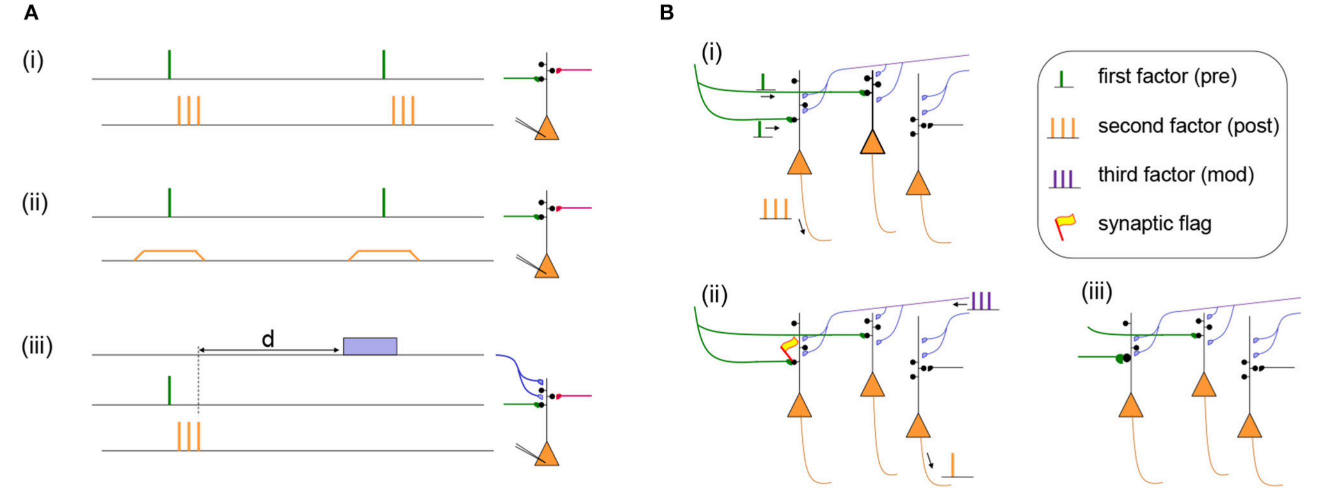

FIGURE 1 | (A) Two Hebbian protocols and one three-factor learning protocol. (i) Hebbian STDP protocol with presynaptic spikes (presynaptic factor) followed by a burst of postsynaptic spikes (postsynaptic factor). Synapses in the stimulated pathway (green) will typically show LTP while an unstimulated synapse (red) will not change its weight (Markram et al., 1997). (ii) Hebbian voltage pairing protocol of presynaptic spikes (presynaptic factor) with a depolarization of the postsynaptic neuron (postsynaptic factor). Depending on the amount of depolarization the stimulated pathway (green) will show LTP or LTD while an unstimulated synapse (red) does not change its weight (Artola and Singer, 1993; Ngezahayo et al., 2000). (iii) Results of a Hebbian induction protocol are influenced by a third factor (blue) even if it is given after a delay *d*. The third factor could be a neuromodulator such as dopamine, acetylcholine, noreprinephrine, or serotonin (Pawlak et al., 2010; Yagishita et al., 2014; Brzosko et al., 2015, 2017; He et al., 2015; Bittner et al., 2017). **(B)** Specificity of three-factor learning rules. (i) Presynaptic input spikes (green) arrive at two different neurons, but only one of these also shows postsynaptic activity (orange spikes). (ii) A synaptic flag is set only at the synapse with a Hebbian co-activation of pre- and postsynaptic factors; the synapse become then eligible to interact with the third factor (blue). Spontaneous spikes of other neurons do not interfere. (iii) The interaction of the synaptic flag with the third factor leads to a strengthening of the synapse (green).

(EPSP), is given by

$$\frac{d}{dt} w_{ij} = e_{ij} \, M^{3rd}(t) \tag{2}$$

where $M^{3rd}(t)$ refers to the global third factor (Izhikevich, 2007; Legenstein et al., 2008; Frémaux et al., 2013). Therefore, according to Equation 2, a non-zero third factor is needed to transform the eligibility trace into a weight change; cf. **Figure 1Aiii**. Note that the weight change is proportional to $M^{3rd}(t)$. Thus, the third factor influences the speed of learning. In the absence of the third factor ($M^{3rd}(t) = 0$), the synaptic weight is not changed. We emphasize that a positive value of the synaptic flag in combination with a negative value $M^{3rd} < 0$ leads to a decrease of the weight. Therefore, the third factor also influences the *direction* of change.

In the discussion so far, $M^{3rd}(t)$ in Equation (2) can take positive and negative values. Such a behavior is typical for a phasic signal which we may mathematically define as the deviation from a running average. We may, for example, think of the third factor as a phasic neuromodulatory signal. However, the third factor could also be biologically implemented by *positive* excursions of the activity using two different neuromodulators with very low baseline activity. The activity of the first modulator could indicate positive values of the third factor and that of the second modulator negative ones - similar to ON and OFF cells in the retina. Similarly, the framework of neoHebbian three-factor rules is general enough to enable biological implementations with separate eligibility traces for LTP and LTD as discussed above in the context of the Clopath model (Clopath et al., 2010).

What could be the source of such a third factor, be it a single neuromodulator or several different ones? The third factor could be triggered by attentional processes, surprising events, or reward. Phasic signals of neuromodulators such as dopamine, serotonin, acetylcholine, or noradrenaline are obvious candidates for a third factor, but potentially not the only ones. Note that axonal branches of most dopaminergic, serotonergic, cholinergic, or adrenergic neurons project broadly onto large regions of cortex so that a phasic neuromodulator signal arrives at many neurons and synapses in parallel (Schultz, 1998). Since neuromodulatory information is shared by many neurons, the variable $M^{3rd}(t)$ of the third factor has no neuron-specific index (neither i nor j) in our mathematical formulation. Because of its nonspecific nature, the theory literature sometimes refers to the third factor as a "global" broadcasting signal, even though in practice not every brain region and every synapse is reached by each neuromodulator. The learning rule with the global modulator, as formulated in Equation 2 will be called Type 1 for the remainder of this paper.

To account for some neuron-specific aspects, three-factor learning rules of Type 2,

$$\frac{d}{dt} w_{ij} = e_{ij} \, h_i(M^{3rd}(t)) \,, \tag{3}$$

contain a neuron-specific function h_i that determines how the global third factor M^{3rd} influences synaptic plasticity of the postsynaptic neuron i. Including an index i in the function $h_i(M^{3rd})$ keeps the theory flexible enough to set (for example) $h_i(M^{3rd}) \equiv 0$ for the subset of neurons i that are not reached by a given neuromodulator. In this case, the classification of a given learning rule as Type 1 and Type 2 is somewhat arbitrary as it depends on whether a population of neurons with a three-factor learning rule is embedded in a larger network with static synapses or not. But there are also existing models, where the implementation of a certain function requires the possibility that one subpopulation has plasticity rules under a third factor whereas another one does not (Brea et al., 2013; Rezende and Gerstner, 2014).

The framework of Equation 3 also includes networks where the distribution of neuromodulatory information to different neurons is done with fixed, but random feedback weights b_i (Lillicrap et al., 2016; Guerguiev et al., 2017); we simply need to set $h_i\left(M^{3rd}\right) = h\left(b_i\, M^{3rd}\right)$. It does not, however, include the general case of supervised learning with backpropagation of errors.

One of the important differences between supervised and reinforcement learning is that in most modern supervised learning tasks, such as auto-encoders, the target is, just like the input, a high-dimensional vector. For supervised learning of high-dimensional targets, we need to generalize Equation (3) further, to three-factor learning rules of Type 3,

$$\frac{d}{dt} w_{ij} = e_{ij}\, M_i^{3rd}(t)\,, \tag{4}$$

where M_i^{3rd} is now a neuron-specific (hence non-global) error feedback signal. For the case of standard backpropagation in layered networks, M_i^{3rd} is calculated by a weighted sum over the errors in the next layer closer to the output; a calculation which needs a well-tuned feedback circuit from the output back to previous layers (Roelfsema and van Ooyen, 2005; Lillicrap et al., 2016; Roelfsema and Holtmaat, 2018). Interestingly, learning similar, but not identical, to backpropagation is still possible with feedback circuits, where the direct feedback from the output is replaced with fixed random weights (Lillicrap et al., 2016; Guerguiev et al., 2017), or in networks with a single hidden layer and winner-take-all activity (one-hot coding) in the output layer (Roelfsema and van Ooyen, 2005; Rombouts et al., 2015). In the latter case, the neuron-specific third factor M_i^{3rd} further factorizes into a global modulator M^{3rd} and an attention signal A_i, which leads to a four-factor learning rule (Roelfsema and Holtmaat, 2018).

2.3. Examples and Theoretical Predictions

There are several known examples in the theoretical literature of neoHebbian three-factor rules. We briefly present four of these and formulate expectations derived from the theoretical framework which we would like to compare to experimental results in the next section.

2.3.1. Reward-Based Learning

As a first example of a Type 1 three-factor learning rule, we consider the relation of neoHebbian three-factor rules to reward-based learning. Temporal Difference (TD) algorithms such as SARSA(λ) or TD(λ) from the theory of reinforcement learning (Sutton and Barto, 1998) as well as learning rules derived from policy gradient theories (Williams, 1992) can be interpreted in neuronal networks in terms of neoHebbian three-factor learning rules. The resulting plasticity rules are applied to synapses connecting "state-neurons" (e.g., place cells coding for the current location of an animal) to "action neurons" e.g., cells initiating an action program such as "turn left") (Brown and Sharp, 1995; Suri and Schultz, 1999; Arleo and Gerstner, 2000; Foster et al., 2000; Xie and Seung, 2004; Loewenstein and Seung, 2006; Florian, 2007; Izhikevich, 2007; Legenstein et al., 2008; Vasilaki et al., 2009; Frémaux et al., 2013); for a review, see (Frémaux and Gerstner, 2016). The eligibility trace is increased during the joint activation of "state-neurons" and "action-neurons" and decays exponentially thereafter consistent with the framework of Equation (1). The third factor is defined as reward minus expected reward where the exact definition of expected reward depends on the implementation details. A long line of research by Wolfram Schultz and colleagues (Schultz et al., 1997; Schultz, 1998, 2002; Schultz and Dickinson, 2000) indicates that phasic increases of the neuromodulator dopamine have the necessary properties required for a third factor in the theoretical framework of reinforcement learning.

However, despite the rich literature on dopamine and reward-based learning accumulated during the last 25 years, there is scant data on the decay time constant τ_e of the eligibility trace e_{ij} in Equation (1) before 2015 (except the locusts study Cassenaer and Laurent, 2012). From the mathematical framework of neoHebbian three-factor rules it is clear that, in the context of action learning, the time constant of the eligibility trace (i.e., the duration of the synaptic flag) should roughly match the time span from the initiation of an action to the delivery of reward. As an illustration, let us imagine a baby that attempts to grasp her bottle of milk. The typical duration of one grasping movement is in the range of a second, but potentially only the third grasping attempt might be successful. Let us suppose that each grasping movement corresponds to the co-activation of some neurons in the brain. If the duration of the synaptic flag is much less than a second, the co-activation of pre- and postsynaptic neurons that sets the synaptic flag (eligibility trace) cannot be linked to the reward 1 s later and synapses do not change. If the duration of the synaptic flag is much longer than a second, then the two "wrong" grasping attempts are reinforced nearly as strongly as the third, successful one which mixes learning of "wrong" co-activations with the correct ones. Hence, *the existing theory of three-factor learning rules predicts that the synaptic flag (eligibility trace for action learning) should be in the range of a typical elementary action, about 200 ms to 2 s*; see, for example, p. 15 of Schultz (1998)[1], p.3 of Izhikevich (2007) [2], p.3 of Legenstein et al. (2008)[3],

[1] "Learning ... (with dopamine) on striatal synapses ... requires hypothetical traces of synaptic activity that last until reinforcement occurs and makes those synapses eligible for modification ..."

[2] "(The eligibility trace c) decays to $c = 0$ exponentially with the time constant $\tau_c = 1$ s."

[3] "The time scale of the eligibility trace is assumed in this article to be on the order of seconds."

p. 13327 of Frémaux et al. (2010)[4], or p. 13 of Frémaux et al. (2013)[5]. An eligibility trace of <100 ms or more than 10 s would be less useful for learning a typical elementary action or delayed reward task than an eligibility trace in the range of 200 ms to 2 s. The expected time scale of the synaptic eligibility trace should roughly match the maximal delay of reinforcers in conditioning experiments (Thorndike, 1911; Pavlov, 1927; Black et al., 1985), linking synaptic processes to behavior. For human behavior, delaying a reinforcer by 10 s during ongoing actions decreases learning compared to immediate reinforcement (Okouchi, 2009).

2.3.2. Surprise-Based Learning

As a second example of a Type 1 three-factor learning rule, we consider situations that go beyond standard reward-based learning. Even in the absence of reward, a surprising event might trigger a combination of neuromodulators such as noradrenaline, acetylcholine and dopamine that may act as third factor for synaptic plasticity. Imagine a small baby lying in the cradle with an attractive colorful object swinging above him. He spontaneously makes several arm movements until finally he succeeds, by chance, to grasp the object. There is no food reward for this action. However, the fact that he can now turn the object, look at it from different sides, or put it in his mouth is satisfying because it leads to many novel (and exciting!) stimuli. The basic idea is that, in such situations, novelty or surprise acts as a reinforcer even in complete absence of food rewards (Schmidhuber, 1991; Singh et al., 2004; Oudeyer et al., 2007). Theoreticians have studied these ideas in the context of curiosity (Schmidhuber, 2010), information gain during active exploration (Storck et al., 1995; Schmidhuber, 2006; Sun et al., 2011; Little and Sommer, 2013; Friston et al., 2016), and via formal definitions of surprise (Shannon, 1948; Storck et al., 1995; Itti and Baldi, 2009; Friston, 2010; Schmidhuber, 2010; Faraji et al., 2018). Note that surprise is not always linked to active exploration but can also occur in a passive situation, e.g., listening to tone beeps or viewing simple stimuli (Squires et al., 1976; Kolossa et al., 2013, 2015; Meyniel et al., 2016). Measurable physiological responses to surprise include pupil dilation (Hess and Polt, 1960) and the P300 component of the electroencephalogram (Squires et al., 1976).

If surprise can play a role similar to reward, then surprise-transmitting broadcast signals should speed-up plasticity. Indeed, theories of surprise as well as hierarchical Bayesian models predict a faster change of model parameters for surprising stimuli than for known ones (Yu and Dayan, 2005; Nassar et al., 2010; Mathys et al., 2011, 2014; Faraji et al., 2018) similar to, but more general than, the well-known Kalman filters (Kalman, 1960). Since the translation of these abstract models into networks of spiking neurons is still missing, precise predictions for surprise modulation of plasticity in the form of three-factor rules are not yet available. However, if we consider noradrenaline, acetylcholine, and/or dopamine as candidate neuromodulators signaling novelty and surprise, we expect that these neuromodulators should have a strong effect on plasticity so as to boost learning of surprising stimuli. The influence of tonic applications of various neuromodulators on synaptic plasticity has been shown in numerous studies (Gu, 2002; Reynolds and Wickens, 2002; Hasselmo, 2006; Pawlak et al., 2010). However, in the context of the above examples, we are interested in phasic neuromodulatory signals. Phasic signals conveying moments of surprise are most useful for learning if they are either synchronous with the stimulus to be learned (e.g., passive listening or viewing) or arise with a delay corresponding to one exploratory movement (e.g., grasping). Hence, we predict from these considerations a decay constant τ_e of the synaptic flag in the range of 1 s, but with a pronounced effect for synchronous or near-synchronous events.

2.3.3. Synaptic Tagging-and-Capture

As our third example of a Type 1 three-factor learning rule, we would like to comment on synaptic consolidation. The synaptic tagging-and-capture hypothesis (Frey and Morris, 1997; Reymann and Frey, 2007; Redondo and Morris, 2011) perfectly fits in the framework of three-factor learning rules: The joint pre- and postsynaptic activity sets the synaptic flag (called "tag" in the context of consolidation) which decays back to zero over the time of 1 h. To stabilize synaptic weights beyond 1 h an additional factor is needed to trigger protein synthesis required for long-term maintenance of synaptic weights (Reymann and Frey, 2007; Redondo and Morris, 2011). Neuromodulators such as dopamine have been identified as the necessary third factor for consolidation (Bailey et al., 2000; Reymann and Frey, 2007; Redondo and Morris, 2011; Lisman, 2017). Indeed, modern computational models of synaptic consolidation take into account the effect of neuromodulators (Clopath et al., 2008; Ziegler et al., 2015) in a framework reminiscent of the three-factor rule defined by Equations (1, 2) above. However, there are two noteworthy differences. First, in contrast to reward-based learning, the decay time τ_e of the synaptic tag e_{ij} is in the range of 1 h rather than 1 s, consistent with slice experiments (Frey and Morris, 1997) as well as with behavioral experiments (Moncada and Viola, 2007). Second, in slices, the measured synaptic weights w_{ij} are increased a few minutes after the end of the induction protocol and decay back with the time course of the synaptic tag whereas in the simplest implementation of the three-factor rule framework as formulated in Equations (1, 2) the visible weight is only updated at the moment when the third factor is present. However, slightly more involved models where the visible weight depends on both the synaptic tag variable and the long-term stable weight (Clopath et al., 2008; Ziegler et al., 2015) correctly account for the time course of the measured synaptic weights in consolidation experiments (Frey and Morris, 1997; Reymann and Frey, 2007; Redondo and Morris, 2011).

2.3.4. Supervised Learning With Segregated Dendrites

A recent study proposes a mechanism for implementing 3-factor rules of Type 3 in the context of supervised learning (Guerguiev et al., 2017). Instead of neuromodulators, they propose that top-down feedback signals from the network output to the apical

[4] "Candidate weight changes e_{ij} decay to zero with a time constant $\tau_e = 500$ms. The candidate weight changes e_{ij} are known as the eligibility trace in reinforcement learning."

[5] "The time scales of the eligibility traces we propose, (are) on the order of hundreds of milliseconds, .. Direct experimental evidence of eligibility traces still lacks, ..."

dendrites of pyramidal neurons serve as the 3rd factor in the 3-factor rule. If the output units have a stationary value y_k when driven by the feedforward network input and a stationary value $\hat{y}_k$ when shunted to the target value, then the changes of a weight w_{ij} from neuron j onto the soma (or basal dendrites) of neuron i is governed by Equation (4) with a third factor $M_i = \sum_k b_{ik}(\hat{y}_k - y_k)$ where b_{ik} are random feedback weights from the output neuron k to the apical dendrite of neuron i (Guerguiev et al., 2017). The authors assume relatively weak electrical coupling between the apical dendrite and the soma and suggest that bursts in the apical dendrite could transmit the value of the third factor to synapses onto the soma or basal dendrites (Guerguiev et al., 2017).

Similarly, the Urbanczik-Senn rule for supervised learning can be interpreted as a three-factor rule of Type 3 (Urbanczik and Senn, 2014). Target input to a neuron in the output layer is given at the soma and leads to a spike-train $S_i(t)$ while feedforward input from other neurons in the network arrives in a dendritic compartment where it generates a voltage y_i. The third factor $M_i^{3rd}(t) = S_i(t) - \phi(y_i(t))$ compares the actual spike train (including the somatic drive by the target) with the firing rate expected from dendritic input alone (Urbanczik and Senn, 2014). The authors assume relatively strong electrical coupling between the dendrite and the soma. Interestingly, the same learning rule can also be used in the absence of target information; in this case we prefer to interpret it as a Hebbian two-factor rule, as discussed in the next paragraph.

2.3.5. Summary

The Clopath rule discussed in the paragraph on Hebbian learning rules contains terms that combine a presynaptic factor with two postsynaptic factors, one for instantaneous superthreshold voltage and the other one for low-pass filtered voltage (Clopath et al., 2010). However, despite the fact that it is possible to write the Clopath rule as a multiplication of three factors, we do not classify it as a three-factor rule but rather as a two-factor rule with a nonlinear postsynaptic factor. The main difference to a true three-factor rule is that the third factor M^{3rd} should be related to a feedback signal conveying information on the performance of the network as a whole. As we have seen, this third factor can be a global scalar signal related to reward or surprise or a neuron-specific signal related to the error in the network output. With this nomenclature, the Urbanczik-Senn rule is, just like the Clopath rule (Clopath et al., 2010), a Hebbian two-factor rule if used in unsupervised learning (Urbanczik and Senn, 2014), but the same rule must be seen as a three-factor rule with a neuron-specific (non-global) third factor in the supervised setting.

In summary, the neoHebbian three-factor rule framework has a wide range of applicability. The framework is experimentally well-established in the context of synaptic consolidation where the duration of the flag ("synaptic tag") extracted from slice experiments (Frey and Morris, 1997) is in the range of 1 h, consistent with fear conditioning experiments (Moncada and Viola, 2007). This time scale is significantly longer than what is needed for behavioral learning of elementary actions or for memorizing surprising events. In the context of reward-based learning, theoreticians therefore hypothesized that a process analogous to setting a tag ("eligibility trace") must also exist on the time scale of 1 s. The next section discusses some recent experimental evidence supporting this theoretical prediction.

3. EXPERIMENTAL EVIDENCE FOR ELIGIBILITY TRACES

Recent experimental evidence for eligibility traces in striatum (Yagishita et al., 2014), cortex (He et al., 2015), and hippocampus (Brzosko et al., 2015, 2017; Bittner et al., 2017) is reviewed in the following three subsections.

3.1. Eligibility Traces in Dendritic Spines of Medium Spiny Striatal Neurons in Nucleus Accumbens

In their elegant imaging experiment of dendritic spines of nucleus accumbens neurons, Yagishita et al. (2014) mimicked presynaptic spike arrival by glutamate uncaging (presynaptic factor), paired it with three postsynaptic spikes immediately afterward (postsynaptic factor), repeated this STDP-like pre-before-post sequence ten times, and combined it with optogenetic stimulation of dopamine fibers (3rd factor) at various delays (Yagishita et al., 2014). The ten repetitions of the pre-before-post sequence at 10 Hz took about 1 s while stimulation of dopaminergic fibers (10 dopamine pulses at 30 Hz) projecting from the ventral tegmental area (VTA) to nucleus accumbens took about 0.3 s. In their paper, dopamine was counted as delayed by 1 s if the dopamine stimulation started immediately after the end of the 1 s-long induction period (delay = difference in switch-on time of STDP and dopamine), but for consistency with other data we define the delay d here as the time passed since the end of the STDP protocol. After 15 complete trials the spine volume, an indicator of synaptic strength (Matsuzaki et al., 2001), was measured and compared with the spine volume before the induction protocol. The authors found that dopamine promoted spine enlargement only if phasic dopamine was given in a narrow time window during or immediately after the 1 s-long STDP protocol; cf. **Figure 2A**.

The maximum enlargement of spines occurred if the dopamine signal started during the STDP protocol ($d = -0.4$ s), but even at a delay of $d = 1$ s LTP was still visible. Giving dopamine too early ($d = -2$ s) or too late ($d = +4$ s) had no effect. Spine enlargement corresponded to an increase in the amplitude of excitatory postsynaptic currents indicating that the synaptic weight was indeed strengthened after the protocol (Yagishita et al., 2014). Thus, we can summarize that we have in the *striatum a three-factor learning rule for the induction of LTP where the decay of the eligibility trace occurs on a time scale of 1 s*; cf. **Figure 2A**.

To arrive at these results, Yagishita et al. (2014) concentrated on medium spiny neurons in the nucleus accumbens core, a part of the ventral striatum of the basal ganglia. Functionally, striatum is a particularly interesting candidate for reinforcement learning (Brown and Sharp, 1995; Schultz, 1998; Arleo and Gerstner,

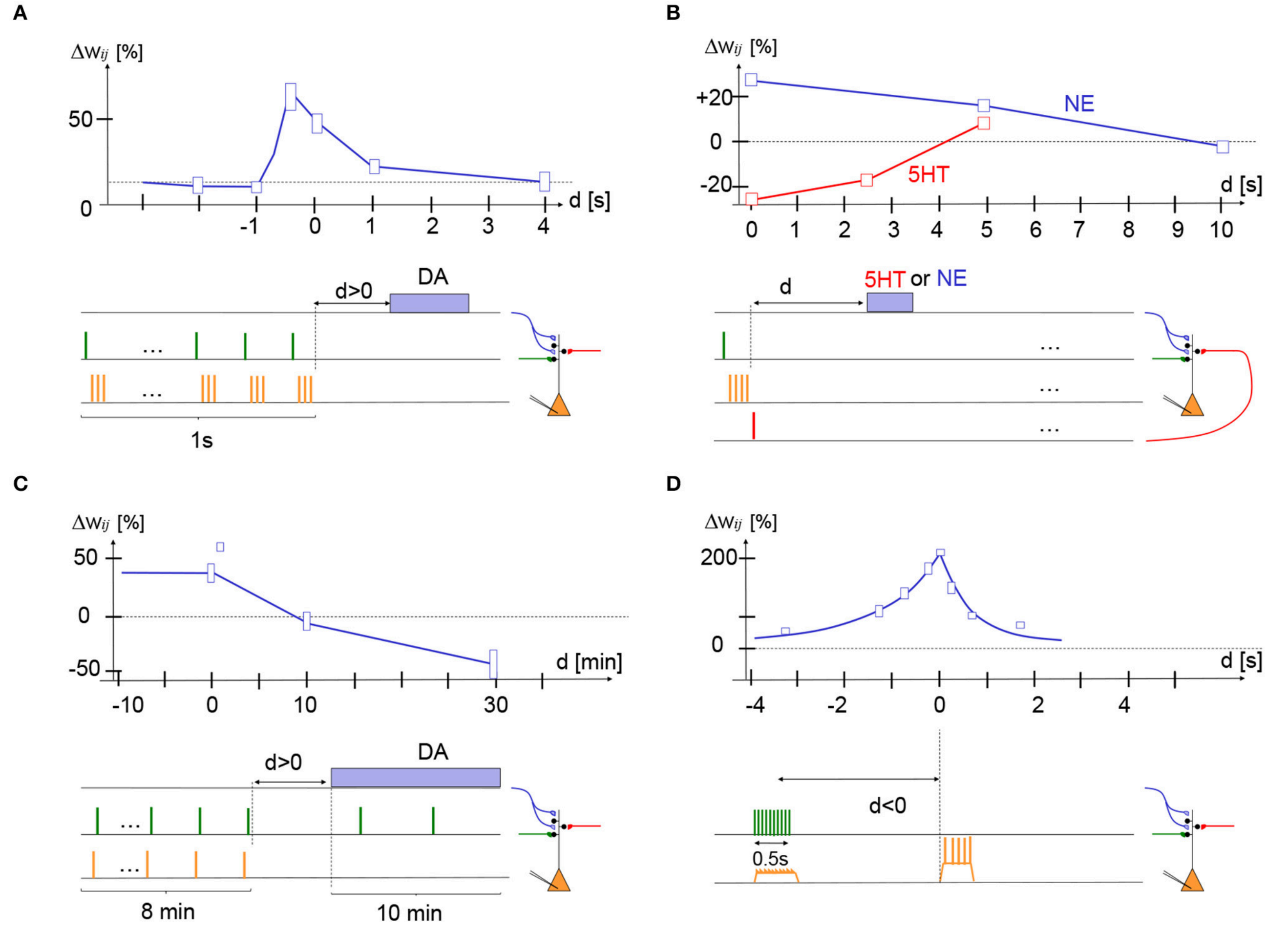

FIGURE 2 | Experimental support for synaptic eligibility traces. Fractional weight change (vertical axis) as a function of delay d of third factor (horizontal axis) for various protocols (schematically indicated at the bottom of each panel). **(A)** In striatum medium spiny cells, stimulation of presynaptic glutamatergic fibers (green) followed by three postsynaptic action potentials (STDP with pre-post-post-post at +10 ms) repeated 10 times at 10 Hz yields LTP if dopamine fibers are stimulated during the presentation ($d < 0$) or shortly afterward ($d = 0$ s or $d = 1$ s) but not if dopamine is given with a delay $d = 4$ s; redrawn after **Figure 1** of Yagishita et al. (2014), with delay d defined as time since end of STDP protocol. **(B)** In cortical layer 2/3 pyramidal cells, stimulation of two independent presynaptic pathways (green and red) from layer 4 to layer 2/3 by a single pulse combined with a burst of four postsynaptic spikes (orange). If the pre-before-post stimulation was combined with a pulse of norepinephrine (NE) receptor agonist isoproterenol with a delay of 0 or 5 s, the protocol gave LTP (blue trace). If the post-before-pre stimulation was combined with a pulse of serotonin (5-HT) of a delay of 0 or 2.5 s, the protocol gave LTD (red trace); redrawn after Figure 6 of He et al. (2015). **(C)** In hippocampus CA1, a post-before-pre (Δt = -20 ms) induction protocol yields LTP if dopamine is present during induction or given with a delay d of 0 or 1 min, but yields LTD if dopamine is absent or given with a delay of 30 min; redrawn after Figures 1F, 2B, and 3C (square data point at delay of 1 min) of Brzosko et al. (2015). **(D)** In hippocampus CA1, 10 extracellular stimuli of presynaptic fibers at 20 Hz cause depolarization of the postsynaptic potential. The timing of a complex spike (calcium plateau potential) triggered by current injection (during 300 ms) after a delay d, is crucial for the amount of LTP. If we interpret presynaptic spike arrival as the first, and postsynaptic depolarization as the second factor, the complex spike could be associated with a third factor; redrawn after Figure 3 of Bittner et al. (2017). Height of boxes gives a very rough estimate of standard deviation - see original papers and figures for details.

2000; Doya, 2000a; Daw et al., 2005) for several reasons. First, striatum receives highly processed sensory information from neocortex and hippocampus through glutamatergic synapses (Mink, 1996; Middleton and Strick, 2000; Haber et al., 2006). Second, striatum also receives dopamine input associated with reward processing (Schultz, 1998). Third, striatum is, together with frontal cortex, involved in the selection of motor action programs (Mink, 1996; Seo et al., 2012).

On the molecular level, the striatal three-factor plasticity depended on NMDA, CaMKII, protein synthesis, and dopamine D1 receptors (Yagishita et al., 2014; Shindou et al., 2018). CaMKII increases were found to be localized in the spine and to have roughly the same time course as the critical window for phasic dopamine suggesting that CaMKII could be involved in the "synaptic flag" triggered by the STDP-like induction protocol, while protein kinase A (PKA) was found to have a nonspecific cell-wide distribution suggesting an interpretation of PKA as a molecule linked to the dopamine-triggered third factor (Yagishita et al., 2014).

3.2. Two Distinct Eligibility Traces for LTP and LTD in Cortical Synapses

In a recent experiment of He et al. (2015), layer 2/3 pyramidal cells in slices from prefrontal or visual cortex were stimulated by an STDP protocol, either pre-before-post for LTP induction or post-before-pre for LTD induction. A neuromodulator was applied with a delay after a single STDP sequence before the whole protocol was repeated; cf. **Figure 2B**. Neuromodulators, either norepinephrine (NE), serotonin (5-HT), dopamine (DA), or acetylcholine (ACh) were ejected from a pipette for 10 s or from endogenous fibers (using optogenetics) for 1 s (He et al., 2015). It was found that NE was necessary for LTP whereas 5-HT was necessary for LTD. DA or ACh agonists had no effect in visual cortex but DA had a positive effect on LTP induction in frontal cortex (He et al., 2015).

For the STDP protocol, He et al. (2015) used extracellular stimulation of two presynaptic pathways from layer 4 to layer 2/3 (presynaptic factor) combined with a burst of 4 postsynaptic action potentials (postsynaptic factor), either pre-before-post or post-before-pre. In a first variant of the experiment, the STDP stimulation was repeated 200 times at 10 Hz corresponding to a total stimulation time of 20 s before the NE or 5-HT was given. In a second variant, instead of an STDP protocol, they paired presynaptic stimulation (first factor) with postsynaptic depolarization (second factor) to −10 mV to induce LTP, or to −40 mV to induce LTD. With both protocols it was found that LTP can be induced if the neuromodulator NE (third factor) arrived with a delay of 5 s or less after the LTP protocol, but not 10 s. LTD could be induced if 5-HT (third factor) arrived with a delay of 2.5 s or less after the LTD protocol, but not 5 s (He et al., 2015).

A third variant of the experiment involved optogenetic stimulation of the noradrenaline, dopamine, or serotonin pathway by repeated light pulses during 1 s applied immediately, or a few seconds, after a minimal STDP protocol consisting of a single presynaptic and four postsynaptic pulses (either pre-before-post or post-before-pre), a protocol that is physiologically more plausible. The minimal sequence of STDP pairing and neuromodulation was repeated 40 times at intervals of 20 s. Results with optogenetic stimulation were consistent with those mentioned above and showed in addition that application of NE or 5-HT immediately before the STDP stimulus did not induce LTP or LTD. *Overall these results indicate that in visual and frontal cortex, pre-before-post pairing leaves an eligibility trace that decays over 5–10 s and that can be converted into LTP by the neuromodulator noradrenaline. Similarly, post-before-pre pairing leaves a shorter eligibility trace that decays over 3 s and can be converted into LTD by the neuromodulator serotonin*; cf. **Figure 2B**.

Functionally, a theoretical model in the same paper (He et al., 2015) showed that the measured three-factor learning rules with two separate eligibility traces stabilized and prolonged network activity so as to allow "event prediction." The authors hypothesized that these three-factor rules were related to reward-based learning in cortex such as perceptual learning in monkeys (Schoups et al., 2001) or mice (Poort et al., 2015) or reward prediction (Shuler and Bear, 2006). The relation to surprise was not discussed but might be a direction for further explorations.

Molecularly, the transformation of the Hebbian pre-before-post eligibility trace into LTP involves beta adrenergic receptors and intracellular cyclic adenosine monophosphate (cAMP) whereas the transformation of the post-pre eligibility trace into LTD involves the 5-HT$_{2c}$ receptor (He et al., 2015). Both receptors are anchored at the postsynaptic density consistent with a role in the transformation of an eligibility trace into actual weight changes (He et al., 2015).

3.3. Eligibility Traces in Hippocampus

Two experimental groups studied eligibility traces in CA1 hippocampal neurons using complementary approaches. In the studies of Brzosko et al. (2015, 2017), CA1 neurons in hippocampal slices were stimulated during about 8 min in an STDP protocol involving 100 repetitions (at 0.2 Hz) of pairs of one extracellularly delivered presynaptic stimulation pulse (presynaptic factor) and one postsynaptic action potential (postsynaptic factor) (Brzosko et al., 2015). Repeated pre-before-post with a relative timing +10 ms gave LTP (in the presence of natural endogenous dopamine) whereas post-before-pre (−20 ms) gave LTD. However, with additional dopamine (third factor) in the bathing solution, post-before-pre at −20 ms gave LTP (Zhang et al., 2009). Similarly, an STDP protocol with post-before-pre at −10 ms resulted in LTP when endogenous dopamine was present, but in LTD when dopamine was blocked (Brzosko et al., 2015). Thus dopamine broadens the STDP window for LTP into the post-before-pre regime (Zhang et al., 2009; Pawlak et al., 2010). Moreover, in the presence of ACh during the STDP stimulation protocol, pre-before-post at +10ms also gave LTD (Brzosko et al., 2017). Thus ACh broadens the LTD window.

The crucial experiment of Brzosko et al. (2015) involved a delay in the dopamine (Brzosko et al., 2015). Brzosko et al. started to perfuse dopamine either immediately after the end of the post-before-pre (-20ms) induction protocol or with a delay. Since the dopamine was given for about 10 min, it cannot be considered as a phasic signal – but at least the *start* of the dopamine perfusion was delayed. Brzosko et al. found that the stimulus that would normally have given LTD turned into LTP if the delay of dopamine was in the range of 1 min or less, but not if dopamine started 10 min after the end of the STDP protocol (Brzosko et al., 2015). Note that for the conversion of LTD into LTP, it was important that the synapses were weakly stimulated at low rate while dopamine was present. Similarly, a prolonged pre-before-post protocol at +10 ms in the presence of ACh gave rise to LTD, but with dopamine given with a delay of <1 min the same protocol gave LTP (Brzosko et al., 2017). To summarize, *in the hippocampus a prolonged post-before-pre protocol (or a pre-before-post protocol in the presence of ACh) yields visible LTD, but also sets an invisible synaptic flag for LTP. If dopamine is applied with a delay of <1 min, the synaptic flag is converted into a positive* *weight change under continued weak presynaptic stimulation*; cf. **Figure 2C**.

Molecularly, the conversion of LTD into LTP after repeated stimulation of post-before-pre pulse pairings depended on NMDA receptors and on the cAMP - PKA signaling cascade (Brzosko et al., 2015). The source of dopamine could be in the Locus Coeruleus which would make a link to arousal and novelty (Takeuchi et al., 2016) or from other dopamine nuclei linked to reward (Schultz, 1998). Since the time scale of the synaptic flag reported in Brzosko et al. (2015, 2017) was in the range of minutes, the process studied by Brzosko et al. could be related to synaptic consolidation (Frey and Morris, 1997; Reymann and Frey, 2007; Redondo and Morris, 2011; Lisman, 2017) rather than eligibility traces in reinforcement learning where shorter time constants are needed (Izhikevich, 2007; Legenstein et al., 2008; Frémaux et al., 2010, 2013). The computational study in Brzosko et al. (2015) used an eligibility trace with a time constant of 2 s and showed that dopamine as a reward signal induced learning of reward location while ACh during exploration enabled a fast relearning after a shift of the reward location (Brzosko et al., 2017).

The second study combined *in vivo* with *in vitro* data (Bittner et al., 2017). From *in vivo* studies it has been known that CA1 neurons in mouse hippocampus can develop a novel, reliable, and rather broadly tuned, place field in a single trial under the influence of a "calcium plateau potential" (Bittner et al., 2015), visible as a complex spike at the soma. Moreover, an artificially induced complex spike was sufficient to induce such a novel place field *in vivo* (Bittner et al., 2015, 2017).

In additional slice experiments, several input fibers from CA3 to CA1 neurons were stimulated by 10 pulses from an extracellular electrode during 1 s. The resulting nearly synchronous inputs at, probably, multiple synapses caused a total EPSP that was about 10 mV above baseline at the soma, and potentially somewhat larger in the dendrite, but did not cause somatic spiking of the CA1 neuron. The stimulated synapses showed LTP if the presynaptic stimulation was paired with a calcium plateau potential (complex spike) in the postsynaptic neuron. LTP occurred, even if the presynaptic stimulation stopped 1 or 2 s before the start of the plateau potential or if the plateau potential started before the presynaptic stimulation (Bittner et al., 2017). The protocol has a remarkable efficiency since potentiation was around 200% after only 5 pairings. Thus, *the joint activation of many synapses sets a flag at the activated synapses which is translated into LTP if a calcium plateau potential (complex spike) occurs a few seconds before or after the synaptic activation*; cf. **Figure 2D**. Molecularly, the plasticity processes implied NMDA receptors and calcium channels (Bittner et al., 2017).

Functionally, synaptic plasticity in hippocampus is particularly important because of the role of hippocampus in spatial memory (O'Keefe and Nadel, 1978). CA1 neurons get input from CA3 neurons which have a narrow place field. The emergence of a broad place field in CA1 has therefore been interpreted as linking several CA3 neurons (that cover for example the 50 cm of the spatial trajectory traversed by the rat before the current location) to a single CA1 cell that codes for the current location (Bittner et al., 2017). Note that at the typical running speed of rodents, 50 cm correspond to several seconds of running. The broad activity of CA1 cells has therefore been interpreted as a predictive representation of upcoming events or places (Bittner et al., 2017). What could such an upcoming event be? For a rodent exploring a T-maze it might for example be important to develop a more precise spatial representation at the T-junction than inside one of the long corridors. With a broad CA1 place field located at the T-junction, information about the upcoming bifurcation could become available several seconds before the animal reaches the junction.

Bittner et al. interpreted their findings as the signature of an unusual form of STDP with a particularly long coincidence window on the behavioral time scale (Bittner et al., 2017). Given that the time span of several seconds between presynaptic stimulation and postsynaptic complex spike is outside the range of a potential causal relation between input and output, they classified the plasticity rule as non-Hebbian because the presynaptic neurons do not participate in firing the postsynaptic one (Bittner et al., 2017). As an alternative view, we propose to classify the findings of Bittner et al. as the signature of an eligibility trace that was left by the joint occurrence of a presynaptic spike arriving from CA3 (presynaptic factor) and a subthreshold depolarization at the location of the synapse in the postsynaptic CA1 neuron (postsynaptic factor); cf. **Figure 2D**. In this view, the setting of the synaptic flag is caused by a "Hebbian"-type induction, except that on the postsynaptic side there are no spikes but just depolarization, consistent with the role of depolarization as a postsynaptic factor (Artola and Singer, 1993; Ngezahayo et al., 2000; Sjöström et al., 2001; Clopath et al., 2010). In this view, the findings of Bittner et al. suggest that the synaptic flag set by the induction protocol leaves an eligibility trace which decays over 2 s. If a plateau potential (related to the third factor) is generated during these 2 s, the eligibility trace caused by the induction protocol is transformed into a measurable change of the synaptic weight. The third factor $M^{3rd}(t)$ in Equation (2) could correspond to the complex spike, filtered with a time constant of about 1 s. Importantly, plateau potentials can be considered as neuron-wide signals (Bittner et al., 2015) triggered by surprising, novel or rewarding events (Bittner et al., 2017). In this view, the results of Bittner et al. are consistent with the framework of neoHebbian three-factor learning rules. If the plateau potentials are indeed linked to surprising events, the three-factor rule framework predicts that *in vivo* many neurons in CA1 receive such a third input as a broadcast-like signal. However, only those neurons that also get, at the same time, sufficiently strong input from CA3 might develop the visible plateau potential (Bittner et al., 2015).

The main difference between the two alternative views is that, in the model discussed in Bittner et al. (2017), *each activated synapse* is marked by an eligibility trace (which is independent of the state of the postsynaptic neuron) whereas in the view of the three-factor rule, the eligibility trace is set only if the presynaptic activation coincides with a strong depolarization of the postsynaptic membrane. Thus, in the model of Bittner et al. the eligibility trace is set by the presynaptic factor alone whereas in the three-factor rule description it is set by the combination of pre- and postsynaptic factors. The two models can be distinguished in future experiments where either the postsynaptic voltage is controlled during presynaptic stimulation or where the number of simultaneously stimulated input fibers is minimized. The prediction of the three-factor rule is that spike arrival at a single synapse, or spike arrival in conjunction with a very small depolarization of <2 mV above rest, is not sufficient to set an eligibility trace. Therefore, LTP will not occur in these cases even if a calcium plateau potential occurs 1 s later.

4. DISCUSSION AND CONCLUSION

4.1. Policy Gradient vs. TD-learning

Algorithmic models of TD-learning with discrete states and in discrete time do not need eligibility traces that extend beyond one time step (Sutton and Barto, 1998). In a scenario where the only reward is given in a target state that is several action steps away from the initial state, reward information shifts, over multiple trials, from the target state backwards, even if the one-step eligibility trace connects only one state to the next (Sutton and Barto, 1998). Nevertheless, extended eligibility traces across multiple time steps are considered convenient heuristic tools to speed up learning in temporal difference algorithms such as $TD(\lambda)$ or SARSA(λ) (Singh and Sutton, 1996; Sutton and Barto, 1998).

In policy gradient methods (Williams, 1992) as well as in continuous space-time TD-learning (Doya, 2000b; Frémaux et al., 2013) eligibility traces appear naturally in the formulation of the problem of reward maximization. Importantly, a large class of TD-learning and policy gradient methods can be formulated as three-factor rules for spiking neurons where the third factor is defined as reward minus expected reward (Frémaux and Gerstner, 2016). In policy gradient methods and related three-factor rules, expected reward is calculated as a running average of the reward (Frémaux et al., 2010) or fixed to zero by choice of reward schedule (Florian, 2007; Legenstein et al., 2008). In TD-learning the expected reward in a given time step is defined as the difference of the value of the current state and that of the next state (Sutton and Barto, 1998). In the most recent large-scale applications of reinforcement learning the expected immediate reward in policy gradient is calculated by a TD-algorithm for state-dependent value estimation (Greensmith et al., 2004; Mnih et al., 2016). An excellent modern summary of Reinforcement Learning Algorithms and their historical predecessors can be found in (Sutton and Barto, 2018).

4.2. Supervised Learning vs. Reinforcement Learning

The experiments in Bittner et al. (2015, 2017) provide convincing evidence that plateau potentials are relevant for the described plasticity events and could be related to the third factor in three-factor rules. But in view of the difference between Equations (2) and (4) the question arises whether the third factor in the Bittner et al. experiments should be considered as a global or as a neuron-specific factor. Obviously, a plateau potential is neuron-specific. The more precise reformulation of this question therefore is whether this specificity is covered by a Type 2 factor written as $h_i(M^{3rd})$ (see Equation 3) or whether it needs the more general Type 3 formulation with M_i^{3rd} (see Equation 4). We see two potential interpretations.

(i) A surprise- or novelty-related global (scalar) neuromodulator M^{3rd} is capable of pushing all CA1 neurons into a state ready to generate a plateau potential, but only a fraction of the neurons actually receive this message and stochastically generate a plateau potential. The term $h_i(M^{3rd})$ expresses the heterogeneity of this process. However, amongst the subset of neurons with $h_i(M^{3rd}) > 0$ only those neurons that have a nonzero eligibility trace will implement synaptic plasticity. Thus the third factor is initially global, but triggers in the end very specific plasticity events limited to a few neurons and synapses only.

(ii) A (potentially high-dimensional) mismatch-related error signal is randomly projected onto different neurons, including those in CA1. The effect is a neuron-specific third factor M_i^{3rd} with index i. This second possibility is particularly intriguing because it relates to theories of attention-gated learning (Roelfsema and Holtmaat, 2018) and learning with segregated synapses (Guerguiev et al., 2017) as instantiations of approximate backpropagation of errors. The high-dimensional signal could be related to the mismatch between what the animal expects to see and what it actually sees in the next instant. In this second interpretation, we leave the field of generalized reinforcement learning and the experiments in Bittner et al. (2015, 2017) can be seen as a manifestation of supervised learning.

4.3. Specificity

If phasic scalar neuromodulator signals are broadcasted over large areas of the brain, the question arises whether synaptic plasticity can still be selective. In the framework of three-factor rules, specificity is inherited from the synaptic flags which are set by the combination of presynaptic spike arrival *and* an elevated postsynaptic voltage at the location of the synapses. The requirement is met only for a small subset of synapses, because presynaptic activity alone or postsynaptic activity alone are not sufficient; cf. **Figure 1B**. Furthermore, among all the flagged synapses only those that show, over many trials, a correlation with the reward signal will be consistently reinforced (Loewenstein and Seung, 2006; Legenstein et al., 2008).

Specificity can further be enhanced by an attentional feedback mechanism (Roelfsema and van Ooyen, 2005; Roelfsema et al., 2010) that restricts the number of eligible synapses to the "interesting" ones, likely to be involved in the task. Such an attentional gating signal acts as an additional factor and turns the three-factor into a four-factor learning rule (Rombouts et al., 2015; Roelfsema and Holtmaat, 2018). Additional specificity can also arise from the fact that not all neurons react in the same way to a modulator as implemented by the notation $h_i(M^{3rd})$ of Type 2 rules (Brea et al., 2013; Rezende and Gerstner, 2014); as well as from additional factors that indicate whether a specific neuron in a population spikes in agreement with the majority of that population (Urbanczik and Senn, 2009). Maximal specificity is achieved with a neuron-specific third factor M_i^{3rd} (Type 3 rules) as in modern implementations of supervised learning (Lillicrap et al., 2016; Guerguiev et al., 2017).

4.4. Mapping to Neuromodulators

A global third factor is likely to be related to neuromodulators, but from the perspective of a theoretician there is no need to assign one neuromodulator to surprise and another one to reward. Indeed, the theoretical framework also works if each neuromodulator codes for a different combination of variables such as surprise, novelty or reward, just as we can use different coordinate systems to describe the same physical system (Frémaux and Gerstner, 2016). Thus, whether dopamine is purely reward related or also novelty related (Ljunberg et al., 1992; Schultz, 1998; Redgrave and Gurney, 2006) is not critical for the development of three-factor learning rules as long as dimensions relating to novelty, surprise, and reward are all covered by the set of neuromodulators.

Complexity in biology is increased by the fact that dopamine neurons projecting from the VTA to the striatum can have separate circuits and functions changing from reward in ventral striatum to novelty in the the tail of striatum (Menegas et al., 2017). Similarly, dopaminergic fibers starting in the VTA can have a different function than those starting in Locus Coeruleus (Takeuchi et al., 2016). Furthermore, findings over the last decade indicate that midbrain dopamine neurons generally show a high diversity of responses and input-output mappings (Fiorillo et al., 2013; Roeper, 2013). Finally, the time scale of eligibility traces could vary from one brain area to the next, in line with the general idea that higher cortical areas show more persistent activity than primary sensory areas (Wang and Kennedy, 2016). If the time scale of eligibility traces is slower in higher areas, we speculate that temporal links between more abstract, slowly evolving concepts could be picked up by plasticity rules. The framework of three-factor rules is general enough to allow for these, and many other, variations.

4.5. Alternatives to Eligibility Traces for Bridging the Gap Between the Behavioral and Neuronal Time Scales

From a theoretical point of view, there is nothing—apart from conceptual elegance—to favor eligibility traces over alternative neuronal mechanisms to associate events that are separated by a second or more. For example, memory traces hidden in the rich firing activity patterns of a recurrent network (Maass et al., 2002; Jaeger and Haas, 2004; Buonomano and Maass, 2009; Susillo and Abbott, 2009) or short-term synaptic plasticity in recurrent networks (Mongillo et al., 2008) could be involved in learning behavioral tasks with delayed feedback. In some models, neuronal, rather than synaptic, activity traces have been involved in learning a delayed paired-associate task (Brea et al., 2016) and a combination of synaptic eligibility traces with prolonged single-neuron activity has been used for learning on behavioral time scales (Rombouts et al., 2015). The empirical studies reviewed here support the idea that the brain makes use of the elegant solution with synaptic eligibility traces and three-factor learning rules, but do not exclude that other mechanisms work in parallel.

4.6. The Paradoxical Nature of Predictions in Computational Neuroscience

If a neuroscientist thinks of a theoretical model, he often imagines a couple of assumptions at the beginning, a set of results derived from simulations or mathematical analysis, and ideally a few novel predictions–but is this the way modeling works? There are at least two types of predictions in computational neuroscience, detailed predictions and conceptual predictions. Well-known examples of detailed predictions have been generated from variants of multi-channel biophysical Hodgkin-Huxley type (Hodgkin and Huxley, 1952) models such as: "if channel X is blocked then we predict that ... " where X is a channel with known dynamics and predictions include depolarization, hyperpolarization, action potential firing, action potential backpropagation or failure thereof. All of these are useful predictions readily translated to and tested in experiments.

Conceptual predictions derived from abstract conceptual models are potentially more interesting, but more difficult to formulate. Conceptual models develop ideas and form our thinking of how a specific neuronal system could work to solve a behavioral task such as working memory (Mongillo et al., 2008), action selection and decision making (Sutton and Barto, 1998), long-term stability of memories (Crick, 1984; Lisman, 1985; Fusi et al., 2005), memory formation and memory recall (Willshaw et al., 1969; Hopfield, 1982). Paradoxically these models often make no detailed predictions in the sense indicated above. Rather, in these and other conceptual theories, the most relevant model features are formulated as *assumptions* which may be considered, in a loose sense, as playing the role of *conceptual predictions*. To formulate it as a short slogan: Assumptions are predictions. Let us return to the conceptual framework of three-factor rules: the purification of rough ideas into the role of three factors is the important conceptual work - and part of the assumptions. Moreover, the specific choice of time constant in the range of 1 s for the eligibility trace has been formulated by theoreticians as one of the model assumptions, rather than as a prediction; cf. the footnotes in section "Examples and theoretical predictions." Why is this the case?

Most theoreticians shy away from calling their conceptual modeling work a "prediction," because there is no logical necessity that the brain must work the way they assume in their model–the brain could have found a less elegant, different, but nevertheless functional solution to the problem under consideration; see the examples in the previous subsection. What a good conceptual model in computational neuroscience shows is that there *exists* a (nice) solution that should ideally not be in obvious contradiction with too many known facts. Importantly, conceptual models necessarily rely on assumptions which in many cases have not (yet) been shown to be true. The response of referees to modeling work in experimental journals therefore often is: "but this has never been shown." Indeed, some assumptions may look far-fetched or even in contradiction with known facts: for example, to come back to eligibility traces, experiments on synaptic tagging-and-capture have shown in the 1990s that the time scale of a synaptic flag is in the range of one *hour* (Frey and Morris, 1997; Reymann and Frey, 2007; Redondo and Morris, 2011; Lisman, 2017), whereas the theory of eligibility traces for action learning needs a synaptic flag on the time scale of one *second*. Did synaptic tagging results imply that three-factor rules for action learning were wrong, because they used the wrong time scale? Or, on the contrary, did these experimental results rather imply that a biological machinery for three-factor rules was indeed in place which could therefore, for other neuron types and brain areas, be used and re-tuned to a different time scale (Frémaux et al., 2013)?

As mentioned earlier, the concepts of eligibility traces and three-factor rules can be traced back to the 1960s, from models formulated in words (Crow, 1968), to firing rate models formulated in discrete time and discrete states (Klopf, 1972; Sutton and Barto, 1981, 1998; Barto et al., 1983; Barto, 1985; Williams, 1992; Schultz, 1998; Bartlett and Baxter, 1999), to models with spikes in a continuous state space and an explicit time scale for eligibility traces (Xie and Seung, 2004; Loewenstein and Seung, 2006; Florian, 2007; Izhikevich, 2007; Legenstein et al., 2008; Vasilaki et al., 2009; Frémaux et al., 2013). Despite the mismatch with the known time scale of synaptic tagging in hippocampus (and lack of experimental support in other brain areas), theoreticians persisted, polished their theories, talked at conferences about these models, until eventually the experimental techniques and the scientific interests of experimentalists were aligned to directly test the assumptions of these theories. In view of the long history of three-factor learning rules, the recent elegant experiments (Yagishita et al., 2014; Brzosko et al., 2015, 2017; He et al., 2015; Bittner et al., 2017) provide an instructive example of how conceptual theories can influence experimental neuroscience.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

This project has been funded by the European Research Council (grant agreement no. 268 689, MultiRules), by the European Union Horizon 2020 Framework Program under grant agreements no. 720270 and no. 785907 (Human Brain Project SGA1 and SGA2), and by the Swiss National Science Foundation (no. 200020 165538).

REFERENCES

Arleo, A., and Gerstner, W. (2000). Spatial cognition and neuro-mimetic navigation: a model of hippocampal place cell activity. *Biol. Cybern.* 83, 287–299. doi: 10.1007/s004220000171

Artola, A., and Singer, W. (1993). Long-term depression of excitatory synaptic transmission and its relationship to long-term potentiation. *Trends Neurosci.* 16, 480–487. doi: 10.1016/0166-2236(93)90081-V

Bailey, C. H., Giustetto, M., Huang, Y.-Y., Hawkins, R. D., and Kandel, E. R. (2000). Is heterosynaptic modulation essential for stabilizing hebbian plasiticity and memory. *Nat. Rev. Neurosci.* 1, 11–20. doi: 10.1038/35036191

Barrett, A. B., Billings, G. O., Morris, R. G., and van Rossum, M. C. (2009). State based model of long-term potentiation and synaptic tagging and capture. *PLoS Comput. Biol.* 5:e1000259. doi: 10.1371/journal.pcbi.1000259

Bartlett, P. L., and Baxter, J. (1999). *Hebbian Synaptic Modification in Spiking Neurons That Learn*. Technical report, Australian National University.

Barto, A. (1985). Learning by statistical cooperation of self-interested neuron-like computing elements. *Hum. Neurobiol.* 4, 229–256.

Barto, A., Sutton, R., and Anderson, C. (1983). Neuronlike adaptive elements that can solve difficult learning and control problems. *IEEE Trans. Syst. Man Cybern.* 13, 835–846.

Benna, M. K., and Fusi, S. (2016). Computational principles of synaptic memory consolidation. *Nat. Neurosci.* 19, 1697–1706. doi: 10.1038/nn.4401

Bienenstock, E. L., Cooper, L. N., and Munroe, P. W. (1982). Theory of the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex. *J. Neurosci.* 2, 32–48. doi: 10.1523/JNEUROSCI.02-01-00032.1982

Bittner, K. C., Grienberger, C., Vaidya, S. P., Milstein, A. D., Macklin, J. J., Suh, J., et al. (2015). Conjunctive input processing drives feature selectivity in hippocampal CA1 neurons. *Nat. Neurosci.* 357, 1133–1142. doi: 10.1038/nn.4062

Bittner, K. C., Milstein, A. D., Grienberger, C., Romani, S., and Magee, J. C. (2017). Behavioral time scale synaptic plasticity underlies CA1 place fields. *Nature* 357, 1033–1036. doi: 10.1126/science.aan3846

Bliss, T. V., and Lømo, T (1973). Long-lasting potentation of synaptic transmission in the dendate area of anaesthetized rabbit following stimulation of the perforant path. *J. Physiol.* 232, 351–356.

Black, J., Belluzzi, J. D., and Stein, L. (1985). Reinforcement delay of one second severely impairs acquisition of brain self-stimulation. *Brain Res.* 359, 113–119. doi: 10.1016/0006-8993(85)91418-0

Bliss, T. V., and Collingridge, G. L. (1993). A synaptic model of memory: long-term potentiation in the hippocampus. *Nature* 361, 31–39. doi: 10.1038/361031a0

Bosch, M., Castro, J., Saneyoshi, T., Matsuno, H., Sur, M., and Hayashi, Y. (2014). Structural and molecular remodeling of dendritic spine substructures during long-term potentation. *Neuron* 82, 444–459. doi: 10.1016/j.neuron.2014.03.021

Brader, J. M., Senn, W., and Fusi, S. (2007). Learning real-world stimuli in a neural network with spike-driven synaptic dynamics. *Neural Comput.* 19, 2881–2912. doi: 10.1162/neco.2007.19.11.2881

Brea, J., Gaál, A. T., Urbancik, R., and Senn, W. (2016). Prospective coding by spiking neurons. *PLoS Comput. Biol.* 12:e1005003. doi: 10.1371/journal.pcbi.1005003

Brea, J., Senn, W., and Pfister, J.-P. (2013). Matching recall and storage in sequence learning with spiking neural networks. *J. Neurosci.* 33, 9565–9575. doi: 10.1523/JNEUROSCI.4098-12.2013

Brown, M. A., and Sharp, P. E. (1995). Simulation of spatial learning in the Morris water maze by a neural network model of the hippocampal-formation and nucleus accumbens. *Hippocampus* 5, 171–188. doi: 10.1002/hipo.450050304

Brzosko, Z., Schultz, W., and Paulsen, O. (2015). Retroactive modulation of spike timing-dependent plasticity by dopamine. *eLife* 4:e09685. doi: 10.7554/eLife.09685

Brzosko, Z., Zannone, S., Schultz, W., Clopath, C., and Paulsen, O. (2017). Sequential neuromodulation of hebbian plasticity offers mechanism for effective reward-based navigation. *eLife* 6:e27756. doi: 10.7554/eLife.27756

Buonomano, D. V., and Maass, W. (2009). State-dependent computations: spatiotemporal processing in cortical networks. *Nat. Rev. Neurosci.* 10, 113–125. doi: 10.1038/nrn2558

Cassenaer, S., and Laurent, G. (2012). Conditional modulation of spike-timing-dependent plasticity for olfactory learning. *Nature* 482, 47–52. doi: 10.1038/nature10776

Clopath, C., Büsing, L., Vasilaki, E., and Gerstner, W. (2010). Connectivity reflects coding: a model of voltage-based spike-timing-dependent-plasticity with homeostasis. *Nat. Neurosci.* 13, 344–352. doi: 10.1038/nn.2479

Clopath, C., Ziegler, L., Vasilaki, E., Büsing, L., and Gerstner, W. (2008). Tag-trigger-consolidation: a model of early and late long-term-potentiation and depression. *PLoS Comput. Biol.* 4:e1000248. doi: 10.1371/journal.pcbi.1000248

Crick, F. (1984). Neurobiology-memory and molecular turnover. *Nature* 312:101. doi: 10.1038/312101a0

Crow, T. (1968). Cortical synapses and reinforcement: a hypothesis. *Nature* 219, 736–737. doi: 10.1038/219736a0

Daw, N. D., Niv, Y., and Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nat. Neurosci.* 8, 1704–1711. doi: 10.1038/nn1560

Deger, M., Helias, M., Rotter, S., and Diesmann M., (2012). Spike-timing dependence of structural plasticity explains cooperative synapse formation in the neocortex. *PLoS Comput. Biol.* 8:e1002689. doi: 10.1371/journal.pcbi.1002689

Deger, M., Seeholzer, A., and Gerstner, W. (2018). Multicontact co-operativity in spike-timing-dependent structural plasticity stabilizes networks. *Cereb. Cortex* 28, 1396–1415. doi: 10.1093/cercor/bhx339

Doya, K. (2000a). Complementary roles of basal ganglia and cerebellum in learning and motor control. *Curr. Opin. Neurobiol.* 10, 732–739. doi: 10.1016/S0959-4388(00)00153-7

Doya, K. (2000b). Temporal difference learning in continuous time and space. *Neural Comput.* 12, 219–245.

Faraji, M., Preuschoff, K., and Gerstner, W. (2018). Balancing new against old information: the role of puzzlement surprise in learning. *Neural Comput.* 30, 34–83. doi: 10.1162/neco_a_01025

Fauth, M., Wörgötter, F., and Tetzlaff, C. (2015). The formation of multi-synaptic connections by the interaction of synaptic and structural plasticity and their functional consequences. *PLoS Comput. Biol.* 11:e1004031. doi: 10.1371/journal.pcbi.1004031

Fiorillo, C. D., Yun, S. R., and Song, M. R. (2013). Diversity and homogeneity in responses of midbrain dopamine neurons. *J. Neurosci.* 33, 4693–4709. doi: 10.1523/JNEUROSCI.3886-12.2013

Florian, R. V. (2007). Reinforcement learning through modulation of spike-timing-dependent synaptic plasticity. *Neural Comput.* 19, 1468–1502. doi: 10.1162/neco.2007.19.6.1468

Foster, D. J., Morris, R. G., and Dayan, P. (2000). Models of hippocampally dependent navigation using the temporal difference learning rule. *Hippocampus* 10, 1–16. doi: 10.1002/(SICI)1098-1063(2000)10:1<1::AID-HIPO1>3.0.CO;2-1

Frémaux, N., and Gerstner, W. (2016). Neuromodulated spike-timing dependent plasticity and theory of three-factor learning rules. *Front. Neural Circ.* 9:85. doi: 10.3389/fncir.2015.00085

Frémaux, N., Sprekeler, H., and Gerstner, W. (2010). Functional requirements for reward-modulated spike-timing-dependent plasticity,. *J. Neurosci.* 40, 13326–13337. doi: 10.1523/JNEUROSCI.6249-09.2010

Frémaux, N., Sprekeler, H., and Gerstner, W. (2013). Reinforcement learning using continuous time actor-critic framework with spiking neurons. *PLoS Comput. Biol.* 9:e1003024. doi: 10.1371/journal.pcbi.1003024

Frey, U., and Morris, R. (1997). Synaptic tagging and long-term potentiation. *Nature* 385, 533–536. doi: 10.1038/385533a0

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci.* 11, 127–138. doi: 10.1038/nrn2787

Friston, K., Fitzgerald, T., Rigoli, F., Schwartenbeck, P., O'Doherty, J., and Pezzulo, G. (2016). Active inference and learning. *Neurosci. and Behav. Rev.* 68, 862–879. doi: 10.1016/j.neubiorev.2016.06.022

Froemke, R. C., and Dan, Y. (2002). Spike-timing dependent plasticity induced by natural spike trains. *Nature* 416, 433–438. doi: 10.1038/416433a

Fusi, S., Drew, P. J., and Abbott, L. F. (2005). Cascade models of synaptically stored memories. *Neuron* 45, 599–611. doi: 10.1016/j.neuron.2005.02.001

Gerstner, W., Kempter, R., van Hemmen, J. L., and Wagner, H. (1996). A neuronal learning rule for sub-millisecond temporal coding. *Nature* 383, 76–78. doi: 10.1038/383076a0

Gjorjieva, J., Clopath, C., Audet, J., and Pfister, J. P. (2011). A triplet spike-timing dependent plasticity model generalizes the Bienenstock-Cooper-Munro rule to higher-order spatiotemporal correlations. *Proc. Natl. Sci. Acad. U.S.A.* 108, 19383–19388. doi: 10.1073/pnas.1105933108

Graupner, M., and Brunel, N. (2007). STDP in a bistable synapse model based on CaMKII and associate signaling pathways. *PLoS Comput. Biol.* 3:e221. doi: 10.1371/journal.pcbi.0030221

Greensmith, E., Bartlett, P., and Baxter, J. (2004). Variance reduction techniques for gradient estimates in reinforcement learning. *J. Machine Learn. Res/* 5, 1471–1530.

Gu, Q. (2002). Neuromodulatory transmitter systems in the cortex and their role in cortical plasticity. *Neuroscience* 111, 815–835. doi: 10.1016/S0306-4522(02)00026-X

Guerguiev, J., Lillicrap, T. P., and Richards, B. A. (2017). Towards deep learning with segregated dendrites. *elife* 6:e22901. doi: 10.7554/eLife.22901

Haber, S. N., Kim, K.-S., Mailly, P., and Calzavara, R. (2006). Reward-related cortical inputs define a large striatal region in primates that interface with associative cortical connections, providing a substrate for incentive-based learning. *J. Neurosci.* 26, 8368–8376. doi: 10.1523/JNEUROSCI.0271-06.2006

Hasselmo, M. (2006). The role of acetylcholine in learning and memory. *Curr. Opin. Neurobiol.* 16, 710–715. doi: 10.1016/j.conb.2006.09.002

He, K., Huertas, M., Hong, S. Z., Tie, X., Hell, J.W., Shouval, H., et al. (2015). Distinct eligibility traces for LTP and LTD in cortical synapses. *Neuron* 88, 528–538. doi: 10.1016/j.neuron.2015.09.037

Hebb, D. O. (1949). *The Organization of Behavior*. New York, NY: Wiley.

Helias, M., Rotter, S., Gewaltig, M.-O., and Diesmann, M. (2008). Structural plasticity controlled by calcium based correlation detection. *Front. Comput. Neurosci.* 2:7. doi: 10.3389/neuro.10.007.2008

Hess, E. H., and Polt, J. M. (1960). Pupil size as related to interest value of visual stimuli. *Science* 132, 349–350. doi: 10.1126/science.132.3423.349

Hodgkin, A. L., and Huxley, A. F. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *J Physiol* 117, 500–544. doi: 10.1113/jphysiol.1952.sp004764

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci. U.S.A.* 79, 2554–2558. doi: 10.1073/pnas.79.8.2554

Huganir, R. L., and Nicoll, R. A. (2013). AMPARs and synaptic plasticity: the last 25 years. *Neuron* 80, 704–717. doi: 10.1016/j.neuron.2013.10.025

Itti, L., and Baldi, P. (2009). Bayesian surprise attracts human attention. *Vis. Res.* 49, 1295–1306. doi: 10.1016/j.visres.2008.09.007

Izhikevich, E. (2007). Solving the distal reward problem through linkage of STDP and dopamine signaling. *Cereb. Cortex* 17, 2443–2452. doi: 10.1093/cercor/bhl152

Izhikevich, E. M., and Desai, N. S. (2003). Relating STDP to BCM. *Neural Comput.* 15, 1511–1523. doi: 10.1162/089976603321891783

Jaeger, H., and Haas, H. (2004). Harnessing nonlinearity: predicting chaotic systems and saving energy in wireless communication. *Science* 304, 78–80. doi: 10.1126/science.1091277

Kalman, R. (1960). A new approach to linear filtering and prediction problems. *J. Basic Eng.* 82, 35–45. doi: 10.1115/1.3662552

Kempter, R., Gerstner, W., and van Hemmen, J. L. (1999). Hebbian learning and spiking neurons. *Phys. Rev. E* 59, 4498–4514. doi: 10.1103/PhysRevE.59.4498

Klopf, A. (1972). *Brain Function and Adaptive Systems-a Heterostatic Theory*. Air Force Cambridge research laboratories, special Reports Vol. 133, 1–70.

Kolossa, A., Fingscheidt, T., Wessel, K., and Kopp, B. (2013). A model-based approach to trial-by-trial p300 amplitude fluctuations. *Front. Hum. Neurosci.* 6:359. doi: 10.3389/fnhum.2012.00359

Kolossa, A., Kopp, B., and Finscheidt, T. (2015). A computational analysis of the neural bases of bayesian inference. *NeuroImage* 106, 222–237. doi: 10.1016/j.neuroimage.2014.11.007

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

Levy, W. B., and Stewart, O. (1983). Temporal contiguity requirements for long-term associative potentiation/depression in hippocampus. *Neurosci,* 8, 791–797. doi: 10.1016/0306-4522(83)90010-6

Lillicrap, T. P., Cownden, D. B., Tweed, D., and Akerman, C. J. (2016). Random synaptic feedback weights support error backpropagation for deep learning. *Nat. Commun.*, 7:13276. doi: 10.1038/ncomms13276

Lisman, J. (1985). A mechanism for memory storage insensitive to molecular turnover: a bistable autophosphorylating kinase. *Proc. Natl. Acad. Sci. U.S.A.* 82, 3055–3057. doi: 10.1073/pnas.82.9.3055

Lisman, J. (1989). A mechanism for Hebb and anti-Hebb processes underlying learning and memory. *Proc. Natl. Acad. Sci. U.S.A.* 86, 9574–9578. doi: 10.1073/pnas.86.23.9574

Lisman, J. (2003). Long-term potentiation: outstanding questions and attempted synthesis. *Phil. Trans. R. Soc. Lond B Biol. Sci.* 358, 829–842. doi: 10.1098/rstb.2002.1242

Lisman, J. (2017). Glutamatergic synapses are structurally and biochemically complex because of multiple plasticity processes: long-term potentiation, long-term depression, short-term potentiation and scaling. *Phil. Trans. Roy. Soc. B* 372:20160260. doi: 10.1098/rstb.2016.0260

Lisman, J., Grace, A. A., and Duzel, E. (2011). A neoHebbian framework for episodic memory; role of dopamine-dependent late LTP. *Trends Neurosci.* 34, 536–547. doi: 10.1016/j.tins.2011.07.006

Little, D. J., and Sommer, F. T. (2013). Learning and exploration in action-perception loops. *Front. Neural Circ.* 7:37. doi: 10.3389/fncir.2013.00037

Ljunberg, T., Apicella, P., and Schultz, W. (1992). Responses of monkey dopamine neurons during learning of behavioral interactions. *J. Neurophysiol.* 67, 145–163. doi: 10.1152/jn.1992.67.1.145

Loewenstein, Y., and Seung, H. (2006). Operant matching is a generic outcome of synaptic plasticity based on the covariance between reward and neural activity. *Proc. Natl. Acad. Sci. U.S.A.* 103, 15224–15229. doi: 10.1073/pnas.0505220103

Löwel, S., and Singer, W. (1992). Selection of intrinsic horizontal connections in the visual cortex by correlated neuronal activity. *Science* 255, 209–212. doi: 10.1126/science.1372754

Maass, W., Natschläger, T., and Markram, H. (2002). Real-time computing without stable states: a new framework for neural computation based on perturbations. *Neural Comput.* 14, 2531–2560. doi: 10.1162/089976602760407955

Markram, H., Lübke, J., Frotscher, M., and Sakmann, B. (1997). Regulation of synaptic efficacy by coincidence of postysnaptic AP and EPSP. *Science* 275, 213–215. doi: 10.1126/science.275.5297.213

Martin, S. J., Grimwood, P. D., and Morris, R. G. (2000). Synaptic plasticity and memory: an evaluation of the hypothesis. *Ann. Rev. Neurosci.* 23, 649–711. doi: 10.1146/annurev.neuro.23.1.649

Mathys, C., Daunizeau, J., Friston, K. J., and Stephan, K. E. (2011). A bayesian foundation for individual learning under uncertainty. *Front. Hum. Neurosci.* 5:39. doi: 10.3389/fnhum.2011.00039

Mathys, C. D., Lomakina, E. I., Daunizeau, K. H., Iglesias, S., Brodersen, K. J., Friston, K., et al. (2014). Uncertainty in perception and the hierarchical gaussian filter. *Front. Hum. Neurosci.* 8:825. doi: 10.3389/fnhum.2014.00825

Matsuzaki, M., Ellis-Davies, G. C., Nemoto, T., Iione, M., Miyashita, Y., and Kasai, H. (2001). Dendritic spine geometry is critical for AMPA receptor expression in hippocampal CA1 pyramidal neurons. *Nat. Neurosci.* 4, 1086–1092. doi: 10.1038/nn736

Menegas, W., Babayan, B. M., Uchida, N., and Watabe-Uchida, M. (2017). Opposite initialization to novel cues in dopamine signaling in ventral and posterior striatum in mice. *eLife* 6:e21886. doi: 10.7554/eLife.21886

Meyniel, F., Maheu, M., and Dehaene, S. (2016). Human inferences about sequences: a minimal transition probability model. *PLoS Comput. Biol.* 12:e1005260. doi: 10.1371/journal.pcbi.1005260

Middleton, F. A., and Strick, P. (2000). Basal ganglia and cerebellar loops: motor and cognitive circuits. *Brain Res. Rev.* 31, 236–250. doi: 10.1016/S0165-0173(99)00040-5

Miller, K. D., and MacKay, D. J. C. (1994). The role of constraints in hebbian learning. *Neural Comput.* 6, 100–126. doi: 10.1162/neco.1994.6.1.100

Mink, J. (1996). The basal ganglia: focused selection and inhibition of competing motor programs. *Progr. Neurobiol.* 50, 381–425. doi: 10.1016/S0301-0082(96)00042-1

Mnih, V., Badia, A., Mirza, M., Graves, A., Harley, T., Lillicrap, T., et al. (2016). "Asynchronous methods for deep reinforcement learning," in *Proceedings of the 33rd International Conference on Machine Learning*, eds M. F. Balcan and K. Q. Weinberger (New York, NY), 1928–1937.

Moncada, D., and Viola, H. (2007). Induction of long-term memory by exposure to novelty requires protein synthesis: Evidence for a behavioral tagging. *J. Neurosci.* 27, 7476–7481. doi: 10.1523/JNEUROSCI.1083-07.2007

Mongillo, G., Barak, O., and Tsodyks, M. (2008). Synaptic theory of working memory. *Science* 319, 1543–1546. doi: 10.1126/science.1150769

Nassar, M. R, Wilson, R. C., Heasly, B., and Gold, J. I. (2010). An approximately bayesian delta-rule model explains the dynamics of belief updating in a changing environment. *J. Neurosci.* 30, 12366–12378. doi: 10.1523/JNEUROSCI.0822-10.2010

Ngezahayo, A., Schachner, M., and Artola, A. (2000). Synaptic activation modulates the induction of bidirectional synaptic changes in adult mouse hippocampus. *J. Neurosci.* 20, 2451–2458. doi: 10.1523/JNEUROSCI.20-07-02451.2000

Oja, E. (1982). A simplified neuron model as a principal component analyzer. *J. Math. Biol.* 15, 267–273. doi: 10.1007/BF00275687

O'Keefe, J., and Nadel, L. (1978). *The Hippocampus as a Cognitive Map*, Vol. 3. Oxford: Clarendon Press Oxford.

Okouchi, H. (2009). Response acquisition by humans with delayed reinforcement. *J. Exp. Anal. Behav.* 91, 377–390. doi: 10.1901/jeab.2009.91-377

Oudeyer, P., Kaplan, F., and Hafner, V. (2007). Intrinsic motivation systems for autonomous mental development. *IEEE Trans. Evol. Comput.* 11, 265–286. doi: 10.1109/TEVC.2006.890271

Pavlov, I. P. (1927). *Conditioned Reflexes: An Investigation of the Physiological Activity of the Cerebral Cortex.* Oxford, UK: Oxford University Press; Humpfrey Milford.

Pawlak, V., Wickens, J., Kirkwood, A., and Kerr, J. (2010). Timing is not everything: neuromodulation opens the STDP gate. *Front. Synaptic Neurosci.* 2:146. doi: 10.3389/fnsyn.2010.00146

Pfister, J.-P., Toyoizumi, T., Barber, D., and Gerstner, W. (2006). Optimal spike-timing-dependent plasticity for precise action potential firing in supervised learning. *Neural Comput.* 18, 1318–1348. doi: 10.1162/neco.2006.18.6.1318

Poort, J., Khan, A. G., Pachitariu, M., Nemri, A., Orsolic, I., Krupic, J., et al. (2015). Learning enhances sensory and multiple non-sensory representations in primary visual cortex. *Neuron* 86, 1478–1490. doi: 10.1016/j.neuron.2015.05.037

Redgrave, P., and Gurney, K. (2006). The short-latency dopamine signal: a role in discovering novel actions? *Nat. Rev. Neurosci.* 7, 967–975. doi: 10.1038/nrn2022

Redondo, R. L., and Morris, R. G. (2011). Making memories last: the synaptic tagging and capture hypothesis. *Nat. Rev. Neurosci.* 12, 17–30. doi: 10.1038/nrn2963

Reymann, K. G., and Frey, J. U. (2007). The late maintenance of hippocampal LTP: requirements, phases,synaptic tagging, late associativity and implications. *Neuropharmacology* 52, 24–40. doi: 10.1016/j.neuropharm.2006.07.026

Reynolds, J. N., and Wickens, J. R. (2002). Dopamine-dependent plasticity of corticostriatal synapses. *Neural Netw.* 15, 507–521. doi: 10.1016/S0893-6080(02)00045-X

Rezende, D., and Gerstner, W. (2014). Stochastic variational learning in recurrent spiking networks. *Front. Comput. Neurosci.* 8:38. doi: 10.3389/fncom.2014.00038

Roelfsema, P. R., and Holtmaat, A. (2018). Control of synaptic plasticity in deep cortical networks. *Nat. Rev. Neurosci.* 19, 166–180. doi: 10.1038/nrn.2018.6

Roelfsema, P. R., and van Ooyen, A. (2005). Attention-gated reinforcement learning of internal representations for classification. *Neural Comput.* 17, 2176–2214. doi: 10.1162/0899766054615699

Roelfsema, P. R., van Ooyen, A., and Watanabe, T. (2010). Perceptual learning rules based on reinforcers and attention. *Trends Cogn. Sci.* 14, 64–71. doi: 10.1016/j.tics.2009.11.005

Roeper, J. (2013). Dissecting the diversity of midbrain dopamine neurons. *Trends Neurosci.* 36, 336–342. doi: 10.1016/j.tins.2013.03.003

Rombouts, J. O., Bothe, S. M., and Roelfsema, P. R. (2015). How attention can create synaptic tags for the learning of working memories in sequential tasks. *PLoS Comput. Biol.* 11:e1004060. doi: 10.1371/journal.pcbi.1004060

Rubin, J. E., Gerkin, R. C., Bi, G.-Q., and Chow, C. C. (2005). Calcium time course as a signal for spike-timing-dependent plasticity. *J. Neurophysiol.* 93, 2600–2613. doi: 10.1152/jn.00803.2004

Rubin, J., Lee, D. D., and Sompolinsky, H. (2001). Equilibrium properties of temporally asymmetric Hebbian plasticity. *Phys. Rev. Lett.* 86, 364–367. doi: 10.1103/PhysRevLett.86.364

Schmidhuber, J. (1991). "Curious model-building control systems," in *Proceedings of the International Joint Conference on Neural Networks*, Vol. 2, (Singapore: IEEE press), 1458–1463.

Schmidhuber, J. (2006). Developmental robotics, optimal artificial curiosity, creativity, music, and the fine arts. *Connect. Sci.* 18, 173–187. doi: 10.1080/09540090600768658

Schmidhuber, J. (2010). Formal theory of creativity, fun, and intrinsic motivation (1990–2010). *IEEE Trans. Auton. Mental Dev.* 2, 230–247. doi: 10.1109/TAMD.2010.2056368

Schoups, A., Vogels, R., Quian, N., and Orban, G. (2001). Practising orientation identification improves orientation coding in V1 neurons. *Nature* 412, 549–553. doi: 10.1038/35087601

Schultz, W. (1998). Predictive reward signal of dopamine neurons. *J. Neurophysiol.* 80, 1–27. doi: 10.1152/jn.1998.80.1.1

Schultz, W. (2002). Getting formal with dopamine and reward. *Neuron* 36, 241–263. doi: 10.1016/S0896-6273(02)00967-4

Schultz, W., Dayan, P., and Montague, P. R. (1997). A neural substrate for prediction and reward. *Science* 275, 1593–1599. doi: 10.1126/science.275.5306.1593

Schultz, W., and Dickinson, A. (2000). Neuronal coding of prediction errors. *Ann. Rev. Neurosci.* 23, 472–500. doi: 10.1146/annurev.neuro.23.1.473

Senn, W., Tsodyks, M., and Markram, H. (2001). An algorithm for modifying neurotransmitter release probability based on pre- and postsynaptic spike timing. *Neural Comput.* 13, 35–67. doi: 10.1162/089976601300014628

Seo, M., Lee, E., and Averbeck, B. (2012). Action selection and action value in frontal-striatal circuits. *Neuron* 74, 947–960. doi: 10.1016/j.neuron.2012.03.037

Shannon, C. (1948). A mathematical theory of communication. *Bell Syst. Techn. J.* 27, 37–423. doi: 10.1002/j.1538-7305.1948.tb01338.x

Shatz, C. (1992). The developing brain. *Sci. Am.* 267, 60–67. doi: 10.1038/scientificamerican0992-60

Shindou, T., Shindou, M., Watanabe, S., and Wickens, J. (2018). A silent eligibility trace enables dopamine-dependent synaptic plasticity for reinforcement learning in the mouse striatum. *Eur. J. Neurosci.* doi: 10.1111/ejn.13921. [Epub ahead of print].

Shouval, H. Z., Bear, M. F., and Cooper, L. N. (2002). A unified model of NMDA receptor dependent bidirectional synaptic plasticity. *Proc. Natl. Acad. Sci. U.S.A.* 99, 10831–10836. doi: 10.1073/pnas.152343099

Shuler, M., and Bear, M. (2006). Reward timing in the primary visual cortex. *Science* 311, 1606–1609. doi: 10.1126/science.1123513

Singh, S., Barto, A., and Chentanez, N. (2004). Intrinsically motivated reinforcement learning. *Adv. Neural Inform. Proc. Syst.* 17, 1281–1288.

Singh, S., and Sutton, R. (1996). Reinforcement learning with replacing eligibility traces. *Mach. Learn.* 22, 123–158. doi: 10.1007/BF00114726

Sjöström, P. J., Turrigiano, G. G., and Nelson, S. B. (2001). Rate, timing, and cooperativity jointly determine cortical synaptic plasticity. *Neuron* 32, 1149–1164. doi: 10.1016/S0896-6273(01)00542-6

Song, S., Miller, K., and Abbott, L. (2000). Competitive Hebbian learning through spike-time-dependent synaptic plasticity. *Nat. Neurosci.* 3, 919–926. doi: 10.1038/78829

Squires, K. C., Wickens, C., Squires, N. K., and Donchin, E. (1976). The effect of stimulus sequence on the waveform of the cortical event-related potential. *Science* 193, 1141–1146. doi: 10.1126/science.959831

Storck, J., Hochreiter, S., and Schmidhuber, J. (1995). "Reinforcement-driven information acquisition in non-deterministic environments," in *Proceedings of ICANN'95,* Vol.2 (Paris: EC2-CIE), 159–164.

Sun, Y., Gomez, F., and Schmidhuber, J. (2011). "Planning to be surprised: optimal Bayesian exploration in dynamic environments," in *Artificial General Intelligence*, Lecture Notes in Computer Science, Vol. 6830, eds J. Schmidhuber, K. R. Thorisson, and M. Looks (Springer), 41–51.

Suri, R. E., and Schultz, W. (1999). A neural network with dopamine-like reinforcement signal that learns a spatial delayed response task. *Neuroscience* 91, 871–890. doi: 10.1016/S0306-4522(98)00697-6

Susillo, D., and Abbott, L. F. (2009). Generating coherent patterns of activity from chaotic neural networks. *Neuron* 63, 544–557. doi: 10.1016/j.neuron.2009.07.018

Sutton, R., and Barto, A. (1998). *Reinforcement Learning*. Cambridge: MIT Press.

Sutton, R., and Barto, A. (2018). *Reinforcement Learning: an introduction* (2nd Edn.) Cambridge, MA: MIT Press.

Sutton, R. S., and Barto, A. G. (1981). Towards a modern theory of adaptive networks: expectation and prediction. *Psychol. Rev.* 88, 135–171. doi: 10.1037/0033-295X.88.2.135

Takeuchi, T., Duszkiewicz, A. J., Sonneborn, A., Spooner, P. A., Yamasaki, M., Watanabe, M., et al. (2016). Locus coeruleus and dopaminergic consolidation of everyday memory. *Nature* 537, 357–362. doi: 10.1038/nature19325

Thorndike, E. (1911). *Animal Intelligence*. Darien, CT: Hafner.

Urbanczik, R., and Senn, W. (2009). Reinforcement learning in populations of spiking neurons. *Nat. Neurosci.* 12, 250–252. doi: 10.1038/nn.2264

Urbanczik, R., and Senn, W. (2014). Learning by the dendritic prediction of somatic spiking. *Neuron* 81, 521–528. doi: 10.1016/j.neuron.2013.11.030

van Rossum, M. C. W., Bi, G. Q., and Turrigiano, G. G. (2000). Stable Hebbian learning from spike timing-dependent plasticity. *J. Neurosci.* 20, 8812–8821. doi: 10.1523/JNEUROSCI.20-23-08812.2000

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

Wang, X.-J., and Kennedy, H. (2016). Brain structure and dynamics across scales: in search of rules. *Curr. Opin. Neurobiol.* 37, 92–98. doi: 10.1016/j.conb.2015.12.010

Williams, R. (1992). Simple statistical gradient-following methods for connectionist reinforcement learning. *Mach. Learn.* 8, 229–256. doi: 10.1007/BF00992696

Willshaw, D. J., Bunemann, O. P., and Longuet-Higgins, H. C. (1969). Non-holographic associative memory. *Nature* 222, 960–962. doi: 10.1038/222960a0

Xie, X., and Seung, H. S. (2004). Learning in neural networks by reinforcement of irregular spiking. *Phys. Rev. E* 69:41909. doi: 10.1103/PhysRevE.69.041909

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. doi: 10.1126/science.1255514

Yu, A. J., and Dayan, P. (2005). Uncertainty, neuromodulation, and attention. *Neuron* 46, 681–692. doi: 10.1016/j.neuron.2005.04.026

Zhang, J. C., Lau, P. M., and Bi,G. Q. (2009). Gain in sensitivity and loss in temporal contrast of STDP by dopaminergic modulation at hippocampal synapses. *Proc. Natl. Acad. Sci. U.S.A.* 106, 13028–13033. doi: 10.1073/pnas.0900546106

Zhang, L. I., Tao, H. W., Holt, C. E., Harris, W. A., and Poo, M. (1998). A critical window for cooperation and competition among developing retinotectal synapses. *Nature* 395, 37–44. doi: 10.1038/25665

Ziegler, L., Zenke, F., Kastner, D. B., and Gerstner, W. (2015). Synaptic consolidation: from synapses to behavioral modeling. *J. Neurosci.* 35, 1319–1334. doi: 10.1523/JNEUROSCI.3989-14.2015

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.