

Teaching signal synchronization in deep neural networks with prospective neurons

Nicolas Zucchet^{1,*} Qianqian Feng^{2,*} Axel Laborieux³
 Friedemann Zenke^{3,4} Walter Senn^{5,†} João Sacramento^{6,†}

¹Computer Science Department, ETH Zürich

²Gatsby Computational Neuroscience Unit, University College London

³Friedrich Miescher Institute for Biomedical Research

⁴Faculty of Science, University of Basel

⁵Department of Physiology, University of Bern

⁶Google, Paradigms of Intelligence Team

Abstract

Working memory requires the brain to maintain information from the recent past to guide ongoing behavior. Neurons can contribute to this capacity by slowly integrating their inputs over time, creating persistent activity that outlasts the original stimulus. However, when these slowly integrating neurons are organized hierarchically, they introduce cumulative delays that create a fundamental challenge for learning: teaching signals that indicate whether behavior was correct or incorrect arrive out-of-sync with the neural activity they are meant to instruct. Here, we demonstrate that neurons enhanced with an adaptive current can compensate for these delays by responding to external stimuli prospectively – effectively predicting future inputs to synchronize with them. First, we show that such prospective neurons enable teaching signal synchronization across a range of learning algorithms that propagate error signals through hierarchical networks. Second, we demonstrate that this successfully guides learning in slowly integrating neurons, enabling the formation and retrieval of memories over extended timescales. We support our findings with a mathematical analysis of the prospective coding mechanism and learning experiments on motor control tasks. Together, our results reveal how neural adaptation could solve a critical timing problem and enable efficient learning in dynamic environments.

*Shared first authorship. † Shared senior authorship. Correspondance to nzucchet@ethz.ch.

Understanding how the brain learns and adapts to new situations is a central question in neuroscience. Effective learning requires associating current neural activity with the delayed consequences of behavior, whether from inherent factors such as delayed rewards or processing delays within the brain. Such processing delays are accentuated by the brain’s hierarchical organization, where different areas specialize in various aspects of information processing [Kandel et al., 2000]. Indeed, working memory requires neurons to slowly integrate their inputs over time to maintain information from the recent past, but these slow dynamics introduce processing delays that accumulate across hierarchical levels. This creates a fundamental problem for learning: when teaching signals – whether external rewards or internal error signals – arrive at output-related areas, the neural activity in earlier layers is already out of sync due to cumulative processing delays, creating a temporal mismatch between the activity that should be instructed and the teaching signals guiding learning.

Error backpropagation [Werbos, 1974, Rumelhart et al., 1986] has emerged as the canonical algorithm for solving the spatial credit assignment problem in deep artificial neural networks, with numerous recent propositions exploring its potential biological implementations [e.g., Lee et al., 2015, Lillicrap et al., 2016, Scellier and Bengio, 2017, Whittington and Bogacz, 2017, Richards et al., 2019, Meulemans et al., 2022]. However, these models typically assume instantaneous transmission, failing to account for the slow dynamics of the neurons responsible for working memory. To address this limitation, researchers have proposed solutions including forward models [Miall and Wolpert, 1996, Wolpert et al., 1998, Gilra and Gerstner, 2017], eligibility traces [Izhikevich, 2007, Gerstner et al., 2018, Zenke and Ganguli, 2018, Bellec et al., 2020], and prospective neurons [Mi et al., 2014, Haider et al., 2021, Senn et al., 2024]. In this work, we focus on the latter, examining how prospective neurons can mitigate delays, both theoretically and practically, and investigating how adaptive neurons can serve as practical implementations of such prospective mechanisms.

We structure the paper as follows. First, we focus on the theoretical foundations of prospective neurons by formalizing delay as a deviation from an ideal trajectory that needs to be tracked – specifically, what neural activity would look like if no processing delays existed. Through this lens, we demonstrate that prospective neurons can effectively circumvent delays after an initial warm-up phase, whereas standard slowly integrating neurons prove insufficient. Next, we explore adaptive neurons as physically realistic implementations of such prospective neurons, analyzing how this physical implementation impacts both tracking and learning performance. Finally, we validate our theoretical insights on a range of behaviorally relevant motor control learning problems, including scenarios with delayed rewards that require tuning the memory storage process. Our findings reveal that the tracking ability of prospective neurons enables effective teaching signal synchronization, thus facilitating online learning in neural networks and providing insights into how biological systems might overcome internal signal propagation delays.

1 Theoretical properties of prospective neurons

Biological neural networks comprise dynamical neurons that integrate information over time, with leaky integrators [Abbott, 1999] standing as the archetypal model for such neurons. The temporal integration of neural signals presents a critical challenge, particularly for neurons responsible for propagating errors, as it results in a misalignment between neural activities and their corresponding learning signals. In this section, we first formalize such delays as deviations from a reference trajectory obtained under an idealized, instantaneous system. We then analyze how the network structure, neuronal characteristics, and external stimuli properties interplay to influence these deviations. Finally, we investigate prospective dynamics as a potential solution to mitigate temporal delays.

1.1 Teaching signal synchronization as a tracking problem

While our ideas apply to a broad class of plasticity rules, we focus our exposition on gradient-based learning and the backpropagation algorithm, as it is the canonical algorithm for learning deep neural networks. In this context, teaching signals are error signals derived from an objective function.

Let us consider a feedforward neural network whose neurons' voltages u^l at layer l are determined by

$$u^{l+1} = W^{l+1} \rho(u^l) \quad \text{and} \quad u^0 = x \quad (1)$$

where ρ is a nonlinear activation function, W^{l+1} the synaptic strength matrix, and x the input of the network, which we consider fixed for now. The voltage of the last layer, u^L , is taken to be the output of the network and is compared with a target y , through the loss function $\ell(u^L, y)$. Error signals δ are backpropagated through the network hierarchy with

$$\delta^l = \rho'(u^l) W^{l+1 \top} \delta^{l+1} \quad \text{and} \quad \delta^L = \nabla \ell(u^L, y). \quad (2)$$

Those equations follow from defining the gradient of the loss with respect to the post-synaptic voltages, $\delta^l := \nabla_{u^l} \ell$, and recursively applying the chain rule to this quantity. When using the mean squared error loss, the error signal for the output units becomes $\delta^L = u^L - y$. Error signals can then be combined with voltages to obtain the gradient following update $\Delta W^{l+1} \propto -\nabla_{W^{l+1}} \ell(u^L, y) = -\delta^{l+1} \rho(u^l)^\top$. While these equations provide a theoretically grounded learning rule, they ignore temporal dynamics: voltages and errors are assumed to propagate instantaneously through the network hierarchy. It can be justified by a separation of timescales argument: the input x and target output y evolve at a slower rate than the neurons, so that they can be considered constant. Through the instantaneous nature of signal processing in this model, error signals always arrive on time.

However, realistic models of learning should take into account scenarios in which external stimuli evolve faster than the response time of the considered neural network. The

separation of timescales argument mentioned above thus no longer holds. We must take into account the temporal dynamics of neural activity. Here, we study a standard leaky integrator model, in which (u_t, δ_t) satisfies the differential equation

$$\begin{aligned}\tau \dot{u}_t &= -u_t + W\rho(u_t) \quad \text{and} \quad u_t^0 = x_t \\ \tau \dot{\delta}_t &= -\delta_t + \rho'(u_t)W^\top \delta_t \quad \text{and} \quad \delta_t^L = \nabla \ell(u_t^L, y_t)\end{aligned}\tag{3}$$

where \dot{u} (resp. $\dot{\delta}$) stands for the temporal derivative of u (resp. δ), and τ is the neuronal integration time constant, which we consider to be the same for all u and δ for simplicity. For conciseness, we have concatenated the voltages u^l (resp. δ^l) of all layers into a single vector u (resp. δ). The matrix W contains all the W^l matrices in its lower block diagonal. Equilibrium states of the dynamics (3) satisfy the backpropagation equations (1) and (2), but they are never reached when (x_t, y_t) evolves rapidly over time. As a corollary, error signals always arrive with some delay, which increases with the depth of the network and the characteristic time constants of the neurons involved.

We formalize the intuition presented above by considering the trajectory of neural activities obtained by instantaneously processing external stimuli as an ideal target that the true neural activity should follow. Stated as such, neurons must solve a tracking problem, and tracking error becomes a clear metric to quantify delay. In the perfect tracking limit, the network behaves like its instantaneous counterpart. Formally, let s_t be a vector summarizing the state of the system variables at time t (the voltages u_t and error signals δ_t in our previous example). The input s receives information from both external stimuli and the rest of the network is $f_\theta(s_t, t)$. It depends on the parameters θ of the network (e.g., the synaptic weights W), as well as on x_t and y_t through t . We define a *target trajectory* s^* through the self-consistency equation

$$s_t^* = f_\theta(s_t^*, t)\tag{4}$$

for all t . The instantaneous backpropagation equations (1) and (2), as well as the ones underlying a large variety of more biologically plausible learning algorithms, fit under this framework, as we shall see in Section 3.1. It is important to note that the target trajectory s^* is implicitly defined, in contrast to more traditional control problems where targets are directly provided to the system. Additionally, when f_θ encodes cyclic dependencies through, e.g., recurrent connections, multiple such trajectories can exist.

1.2 Leaky integrators cannot track target trajectories

We demonstrate that leaky integrators satisfying the differential equation $\tau \dot{s}_t = -s_t + f_\theta(s_t, t)$ cannot track the reference trajectory without a strict separation of timescales. Specifically, Theorem 1 states that, under mild regularity assumptions on the structure of f_θ , the target trajectory s_t^* can only be approached up to some error. This error decreases when the network gets faster or external inputs, such as x_t , slow down.

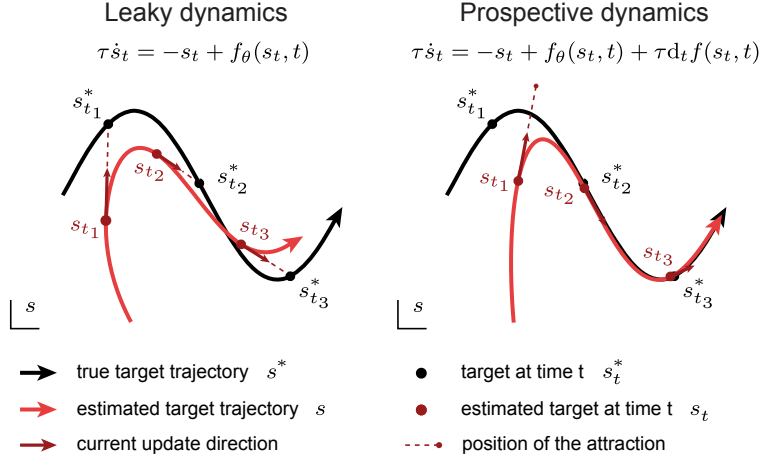


Figure 1: **Prospective dynamics can track the target trajectory whereas leaky dynamics cannot.** Under the leaky integrator dynamics, the current state s_t is attracted towards the current target s_t^* . By the time it arrives there, the target will have already moved away, so s is always lagging behind s^* . Adding a prospective input to the leaky dynamics enables perfect tracking, as the position of s^* in the near future is indirectly estimated and will serve as a target for the s dynamics.

Theorem 1. Let f_θ be such that the largest eigenvalue of the symmetric part of the Jacobian $\partial_s f_\theta(s, t)$ is always smaller than a constant $1 - \mu$ for $\mu > 0$. Let s_t^* be a target trajectory that satisfies $s_t^* = f_\theta(s_t^*, t)$ for all t and $\|\dot{s}_t^*\| \leq \gamma$. Then, the trajectory of states s_t obtained by integrating the leaky dynamics $\tau \dot{s}_t = -s_t + f_\theta(s_t, t)$ satisfies

$$\limsup_{t \rightarrow \infty} \|s_t - s_t^*\| \leq \frac{\gamma\tau}{\mu}.$$

Furthermore, this bound is tight; that is, we can find a f for which the upper bound is reached.

This result formalizes the intuition that the trajectory s obtained through leaky integration lags behind the target s^* . More precisely, s asymptotically lies in a tube centered on s^* . Intuitively, this occurs as s is attracted to the current target, which will have already moved away when reached, cf. Figure 1. The ratio τ/μ reflects the effective time constant of the system: τ corresponds to the time constant of individual neurons, while $1/\mu$ is linked to network synaptic weights. For instance, as the depth of the network increases, so does $1/\mu$. Under constant external inputs, this constant also represents the worst exponential convergence rate that the system can exhibit. The γ constant reflects both network geometry and the rate of change in external inputs. We provide more details on these considerations and a proof of Theorem 1 in Appendix A.

This result provides a tight upper bound for the tracking error $\|s_t - s_t^*\|$. We now verify whether the characterization it provides holds in practice. To this end, we consider

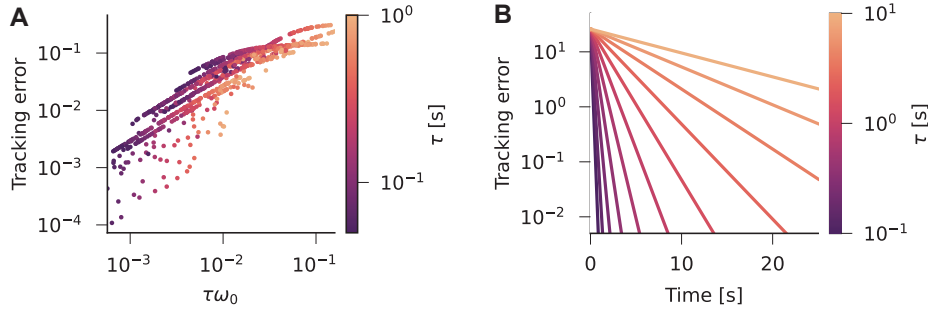


Figure 2: **Theoretical predictions on tracking properties of leaky and prospective neurons hold in simulation.** **A.** Empirical tracking error for the leaky dynamics (3) as a function of the neurons time constant τ times the characteristic angular velocity of the inputs ω_0 . This quantity is directly linked to the theoretical bound of Theorem 1. Each dot corresponds to random τ and ω_0 values. As predicted by Theorem 1, the tracking error scales linearly with $\tau\omega_0$ in the worst case, which measures how slow the neurons are compared to the input. Consequently, teaching signals arrive out-of-sync. **B.** Under the prospective dynamics, the instantaneous empirical tracking error converges exponentially fast to 0, with the neurons time constant τ modulating the convergence speed, consistently with Theorem 2.

a fully connected feedforward neural network with two hidden layers, whose neural activity and corresponding error signals follow the dynamics of (3). We provide random linear combinations of sine waves as input to the network and compare its outputs with those of an instantaneous feedforward teacher network receiving the same input. As the quantities involved in the theorem can be costly to estimate directly, we use proxies in practice. We consider the maximal value of $\| -s_t + f_\theta(s_t, t) \|$ after an initial convergence phase as a proxy for the tracking error of Theorem 1. This is a reasonable choice, as these quantities are linearly correlated under the theorem’s assumptions. We refer to this metric as the *empirical tracking error* and use it to measure how out-of-sync error signals arrive.

We vary two parameters: the time constant τ of the neurons and the typical angular velocity ω_0 at which the sinusoidal inputs vary (used as a proxy for γ ; see Methods). All other factors, including neural network weights, remain fixed. Figure 2A illustrates how the empirical tracking error evolves as a function of τ and ω_0 . Our empirical results confirm the theory: better tracking is only reliably achievable through timescale separation, that is, as $\tau\omega_0$ approaches 0.

1.3 Prospective neurons can track target trajectories

As argued in the previous section, vanilla leaky integrator dynamics cannot achieve such tracking as the state s_t is attracted to the current target and is thus always lagging behind the target s_t^* . Instead, if we manage to estimate what the target state will be

in the future and use it as an implicit target for neural activity, the gap between s_t and s_t^* will progressively vanish, as illustrated in Figure 1. In the following, we show that slightly modifying the leaky dynamics by adding a prospective input to the inputs that leaky neurons receive will asymptotically (on a time scale of τ) lead to perfect target tracking. As a consequence, prospective neurons behave like instantaneous neurons when given enough adaptation time.

More concretely, we introduce prospective dynamics that estimate what the input $f_\theta(s_t, t)$ each neuron receives will be in τ seconds through a first-order approximation and add it to the input. That is, we consider the state dynamics

$$\tau \dot{s}_t = -s_t + f_\theta(s_t, t) + \tau d_t f_\theta(s_t, t). \quad (5)$$

This extra input enables leaky neurons to perfectly track equilibrium, as demonstrated in Theorem 2; see Appendix A for a proof. Figure 2B shows that the Theorem holds in practice, using the same experimental setup as in the previous section.

Theorem 2. *Let s follow the prospective dynamics (5). Then, assuming that $\partial_s f_\theta(s_t, t)$ is always invertible and f_θ is Lipschitz continuous in t on that trajectory, we have*

$$\|s_t - f_\theta(s_t, t)\| = \|s_0 - f_\theta(s_0, 0)\| \exp\left(-\frac{t}{\tau}\right)$$

and

$$\limsup_{t \rightarrow \infty} \|s_t - s_t^*\| = 0,$$

i.e., after an initial exponential convergence phase, s_t and s_t^ coincide.*

We conclude this section by noting that prospective dynamics precede our work. Most closely related are Mi et al. [2014], Haider et al. [2021], and Senn et al. [2024]. The discussion section details connections with existing work from theoretical neuroscience as well as from other fields.

2 Bio-physical implementation of prospective neurons

So far, we have shown that prospective inputs enable leaky neurons to solve the tracking problem inherent to teaching signal synchronization, but we have ignored all details regarding their physical implementation. We now delve into two of these crucial details: We first show how a mechanism appearing in adaptive neuron models [Brette and Gerstner, 2005, Gerstner et al., 2014] can implement prospective dynamics plausibly by subtracting from the instantaneous input current a low-pass filtered version of itself, provided by an adaptation current. Then, we study the robustness of the proposed dynamics to mismatches between the neurons' and prospective inputs' time constants, since we cannot expect them to perfectly match in biological neurons.

2.1 Adaptive neurons can be prospective

Prospective dynamics are a theoretical ideal that enables the efficient tracking of the instantaneous neural activity trajectory and thus remove delays, but they require the physical estimation of time derivatives. In analog filter design [Agarwal and Lang, 2005], it is well established that true differentiators are not physically realizable, as they require an infinite amount of energy. One simple, realizable differentiator is the high-pass filter. More precisely, if a is a low-pass filter of $f_\theta(u_t, t)$ following

$$\tau_a \dot{a}_t = -a_t + f_\theta(u_t, t), \quad (6)$$

then $\tau_a^{-1}(f_\theta(u_t, t) - a_t)$ converges to $d_t f_\theta(u_t, t)$ when τ_a converges to 0. While it introduces some approximation in the estimation of the $d_t f_\theta(u_t, t)$, it is more robust to noise. Plugging this estimate into the prospective dynamics (5) gives

$$\tau \dot{u}_t = -u_t + \left(1 + \frac{\tau}{\tau_a}\right) f_\theta(u_t, t) - \frac{\tau}{\tau_a} a_t \quad (7)$$

Such neural dynamics are reminiscent of adaptive neuron models, in which a can be interpreted as an adaptive current. There exist subtle differences with the adaptive neuron model: the adaptive current here integrates the input current f_θ and not the voltage, the leak term is linear, and the input and adaptive currents have specific weighting factors. We discuss potential physiological implementations of this mechanism in the discussion.

Next, we wonder how effective such adaptive neurons are at tracking the target trajectory and, in particular, whether the adaptive current improves the tracking ability of leaky neurons. To that extent, we study the linearization of the adaptive dynamics around $\tau_a = 0$ up to the first order in τ_a : we show in Appendix B that

$$\tau \dot{u}_t = -u_t + f_\theta(u_t, t) + \tau d_t f_\theta(u_t, t) + \tau \tau_a d_t^2 f_\theta(u_t, t) + O(\tau \tau_a^2) \quad (8)$$

Intuitively, up to a first-order approximation, the tracking error of adaptive dynamics scales with $\tau_a \tau$ when it scales with τ for the vanilla leaky dynamics (Theorem 1). For small enough τ_a , the former are thus better trackers than the latter. Additionally, biological neurons may be implementing more elaborate differentiating circuits, which would make the gap even larger.

We now measure the impact of adding an adaptive current and that of the integration time on the tracking error effect, as shown in Figure 3C, using the teacher-student setting from the previous section. We vary the neuronal time constant τ and consider different τ_a/τ ratios. We make two observations: First, for all τ values considered, the adaptive dynamics are better trackers than the leaky dynamics, even when τ_a is greater than τ . Second, the tracking error also approaches 0 more quickly as τ converges to 0. Those results confirm that the addition of adaptive currents to leaky neurons improves their tracking abilities.

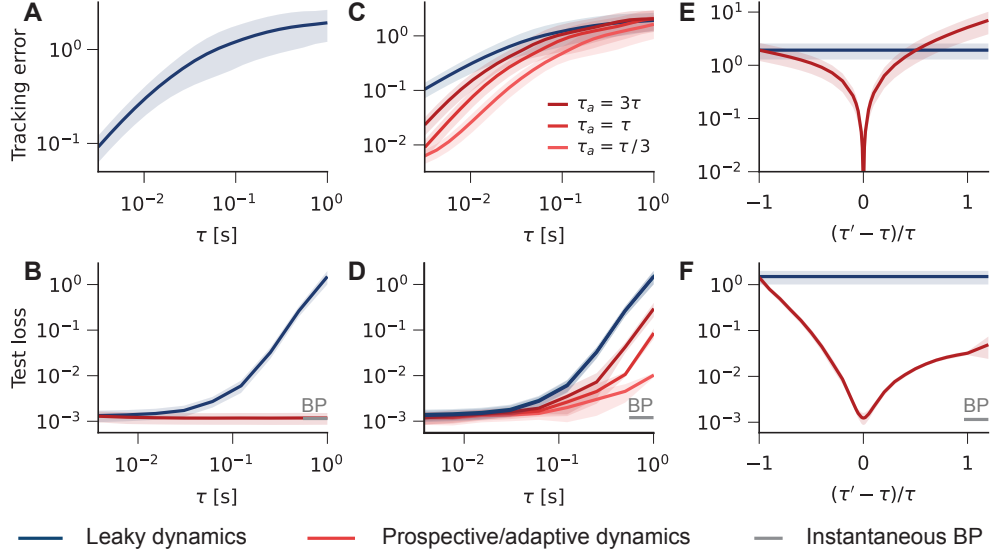


Figure 3: **Better target trajectory tracking translates to greater learning performance.** Top row measures the empirical tracking error in the setup of Sections 1 and 2, bottom row the test loss in the teacher-student task of Section 3.1. Note that the two setups are similar but have some differences, cf. Methods. **A, B.** Leaky dynamics (3) require $\tau \rightarrow 0$ to perfectly track the reference trajectory and thus effectively learn (blue). Prospective dynamics (5) can do so for any τ (red), matching the performance of instantaneous backpropagation (grey). **C, D** Adaptive dynamics (7) track and learn better than leaky dynamics and the effect amplifies as τ_a gets smaller, relatively to τ . **E, F.** Time constant mismatches as in (10) hinder tracking and thus learning in the prospective dynamics. Yet, for a wide range of mismatches, the learning loss remains significantly lower than the one of leaky dynamics.

2.2 Time constant mismatches' impact on prospectiveness is minimal

Besides the requirement for a differentiator circuit, which we discussed in the previous section, our theoretical analysis also relies on the assumption that all neurons have the same time constant τ and that it matches that of their prospective inputs. The first one is straightforward to relax, as our theory equally applies to this regime. In the following, we study the impact of relaxing the second assumption.

Let us consider a simple prospective network in which

$$\tau \dot{u}_t = -u_t + W\rho(u_t) + \tau d_t[W\rho(u_t)]. \quad (9)$$

We ignore teaching signals here for the sake of simplicity, but the following analysis naturally extends to that case. In the previous equation, the prospective input can be computed pre-synaptically by leveraging the fact that $W\rho(u_t) + d_t[W\rho(u_t)] = W(\rho(u_t) + \tau d_t \rho(u_t))$, or post-synaptically, by receiving the input $W\rho(u_t)$ and then

computing its prospective value. We argue that the former is more sensitive to time constant mismatches than the latter. Indeed, if each neuron has a different prospective time constant, the different firing rates received by a neuron will have different underlying time constants. However, only two time constants are involved when prospective inputs are computed post-synaptically. We thus consider that prospective inputs are computed post-synaptically in the following manner and note that this is consistent with adaptive input currents.

Let us now analyze the impact of a time constant mismatch on the tracking properties of the prospective dynamics. Such a mismatch can occur in the adaptive dynamics if the constants in front of f_θ and a_t do not match those prescribed by our theory. To keep the analysis simple, we assume that all neurons have a time constant equal to τ and that the time constant within the prospective input is τ' . The network dynamics introduced above becomes

$$\tau \dot{u}_t = -u_t + W\rho(u_t) + \tau' d_t[W\rho(u_t)]. \quad (10)$$

The deviation from the target trajectory can be expressed through a first-order approximation as

$$u_t = u_t^* + (\tau' - \tau) d_\tau u_t + O((\tau' - \tau)^2) \quad (11)$$

and we are interested in finding an expression for $d_\tau u_t$ to quantify the effect of the mismatch. By injecting Eq. 11 into Eq. 10, if ρ is close to being linear, we show in Appendix B that the deviation $d_\tau u_t$ satisfies the differential equation

$$\tau(d_\tau \dot{u}_t) = -d_\tau u_t + \dot{u}_t^*. \quad (12)$$

It follows that the first-order deviation from the target in $\tau - \tau'$ grows as the target moves faster and that $\|u_t - u_t^*\| \leq |\tau - \tau'| \gamma + O((\tau - \tau')^2)$ at all time t , with γ as in Theorem 1. A consequence of this result is that faster inputs increase the sensitivity of the tracking to time constant mismatches, as they lead to faster u^* and thus larger γ . We also note that this is a local result around $\tau' = \tau$. Larger deviations can qualitatively change the picture. For example, the dynamics can be contractive for $\tau' = \tau$, but explosive for some $\tau' > \tau$, meaning that the distance between u_t^* and u_t can no longer be uniformly bounded over time. Appendix B provides a detailed example.

We simulate a time constant mismatch in Figure 3E and observe that prospective dynamics lead to better tracking abilities than leaky dynamics for a wide range of prospective time constants τ' . Our theoretical analysis suggests that as τ gets smaller, the tracking error grows more slowly as a function of $(\tau' - \tau)/\tau$, so the negative peak we observe around 0 becomes flatter. Fast neurons will thus be relatively less sensitive to a time constant mismatch. Additionally, the prospective and leaky dynamics match for $\tau' = 0$, so we can reasonably expect prospective dynamics to exhibit better tracking for all $\tau' < \tau$ in general, similarly to what we observe in Figure 3.

3 Learning with prospective neurons

Previous sections established the appealing theoretical properties of prospective neurons in terms of tracking, as well as their robustness to physical implementations. In this section, we demonstrate that these benefits translate to learning. We begin with the simple teacher-student setting introduced in Section 1, using prospective neurons to enable teaching signal synchronization across a variety of learning rules. This controlled environment allows us to study how non-ideal prospective neurons affect teaching signal synchronization and learning performance. We then investigate how prospective neurons support reward-based learning in a control task and finally combine them with leaky neurons to successfully solve a reaching task that integrates both working memory and motor control.

3.1 Prospective neurons support teaching signal synchronization in a large variety of learning rules

Our first series of experiments investigates how imprecise tracking impacts online learning performance in feedforward networks and applies prospective dynamics to more biologically plausible learning rules. We employ a teacher-student setup similar to our previous theoretical analysis, with one key difference: synaptic plasticity is now activated. We continue to use the continuous-time backpropagation of errors from (3) as our default learning rule, but with W evolving over time according to the following dynamics:

$$\tau_W \dot{W}_t = -\delta_t \rho(u_t)^\top, \quad (13)$$

and using approximate prospective dynamics (adaptive neurons or $\tau' \neq \tau$) whenever needed. The network therefore learns online, without buffering any updates. As discussed in Section 1.1, these parameter updates follow the gradient when the network is at equilibrium. To isolate the impact of individual components of the learning rule from sampling noise and reduce result variance, we use several samples in parallel (batch size of 50). While this approach is not fully online, we relax this constraint in subsequent sections to train neural networks in a purely online manner.

Better tracking leads to better online learning performance. Throughout this paper, we have argued that learning performance and tracking the target trajectory of neural activities are intimately related through teaching signal synchronization. We now verify this relationship quantitatively.

We find that prospective dynamics achieve learning performance equivalent to the instantaneous backpropagation of errors algorithm, where activities and error signals are computed instantaneously at each time step, following Equations 1 and 2. This finding holds across a wide range of τ values, as shown in Figure 3B. Notably, even in challenging conditions where the sequence length is 5 times the value of τ – prevent-

Connectivity	Method	Train loss	Test loss
Feedforward	Instantaneous BP	3.94×10^{-2}	1.15×10^{-3}
Feedforward	Prospective BP	4.16×10^{-2}	1.13×10^{-3}
Feedforward	Prospective DFA	5.97×10^{-1}	1.32×10^{-2}
Feedforward	Prospective FA	8.08×10^{-1}	1.75×10^{-2}
Feedforward	Leaky BP	7.08×10^1	1.32×10^1
Recurrent	Prospective RBP	1.85×10^{-3}	8.34×10^{-4}
Recurrent	Prospective hEP	1.53×10^{-2}	1.10×10^{-2}
Recurrent	Leaky RBP	1.94×10^{-2}	1.30×10^{-2}
Recurrent	Leaky hEP	1.36×10^1	3.39×10^1

Table 1: **Comparison of different learning algorithms on the teacher-student learning task of Section 3.1.** The first group of learning rules are used to train feedforward neural networks and the second one neural networks with recurrent connections relaxed to equilibrium. Prospective dynamics greatly improve learning performance, to the point of matching the performance of instantaneous backpropagation, which is explained by better tracking abilities. They additionally apply to a large variety of learning rules, such as (direct) feedback alignment or holomorphic equilibrium propagation. The former requires explicit error representation while the latter leverages oscillations to implicitly compute them. (R)BP stands for (recurrent) backpropagation, (D)FA for (direct) feedback alignment, and hEP for holomorphic equilibrium propagation.

ing the dynamics from reaching perfect tracking – prospective dynamics still achieve competitive performance.

Our results demonstrate that tracking precision strongly correlates with online learning performance. By examining the effects of leaky dynamics, adaptive dynamics, and time constant mismatch in prospective dynamics (Figure 3B, D, and F), we observe that in all cases, learning performance approaches that of instantaneous backpropagation as the tracking error converges to 0. This confirms that the dependencies on data and network properties highlighted by our theory are qualitatively the same regarding learning performance.

Prospective dynamics applied to different learning rules. To demonstrate the generality of the notion of teaching signal synchronization as tracking, we show that prospective dynamics enable effective online learning across various learning rules beyond the backpropagation method we have considered so far.

Backpropagation suffers from the weight transport problem [Grossberg, 1987]: its feed-

back pathway is symmetric to the forward one, which is biologically implausible. Algorithms such as feedback alignment [Lillicrap et al., 2016] and direct feedback alignment [Nøkland, 2016] were developed to address this limitation by using distinct random feedback pathways to backpropagate errors. We reformulate these rules using continuous-time dynamics similar to (3) and present their performance in Table 1. While they do not match the performance of backpropagation, they achieve significant results and outperform backpropagation with leaky dynamics, which uses symmetric weights.

The learning rules discussed so far require two key elements: neural activities and error signals. Several studies suggest that apical dendrites within biological neurons may encode these error terms [Sacramento et al., 2018, Richards and Lillicrap, 2019, Mikulasch et al., 2023], although conclusive empirical evidence remains elusive. An alternative class of models uses variations in neural activity from external feedback to implicitly evaluate error signals [Ackley et al., 1985, Baldi and Pineda, 1991, Movellan, 1991, Scellier and Bengio, 2017], typically requiring two distinct phases: one with a teaching signal and one without. To overcome this constraint, early work by Baldi and Pineda [1991] proposed that teaching signal oscillations could enable single-phase learning without directly representing errors. In this approach, feedback remains continuously active with ongoing variations, eliminating the need for discrete phases. However, this introduces a new timescale for the oscillating signal, which must be sufficiently slow to allow neurons to track the target trajectory but faster than the input sequences to provide meaningful signals. Prospective dynamics solve this challenge by enabling neurons to effectively track feedback oscillations, regardless of their intrinsic timescales, potentially providing a key mechanism for contrastive learning in the brain. Using the holomorphic equilibrium propagation framework [Laborieux and Zenke, 2022, hEP], we demonstrate that prospective dynamics indeed support contrastive learning, with oscillation-based methods performing comparably to instantaneous backpropagation (see Table 1), without requiring explicit error signal representation. Note that the model without a teaching signal still uses feedback connections and is therefore not a feedforward neural network. This kind of model is known as a deep equilibrium model in the machine learning literature [Bai et al., 2019] and needs to be trained with recurrent backpropagation [Almeida, 1989, Pineda, 1989, RBP]. Compared to (3), the learning and neural dynamics hardly change, as shown in Appendix A.

Finally, we highlight that numerous additional learning rules requiring neural activity to be at equilibrium can benefit from prospective dynamics. These include predictive coding [Whittington and Bogacz, 2017] (which yields the neuronal least-action principle of Senn et al. [2024]) and control-based learning [Meulemans et al., 2022].

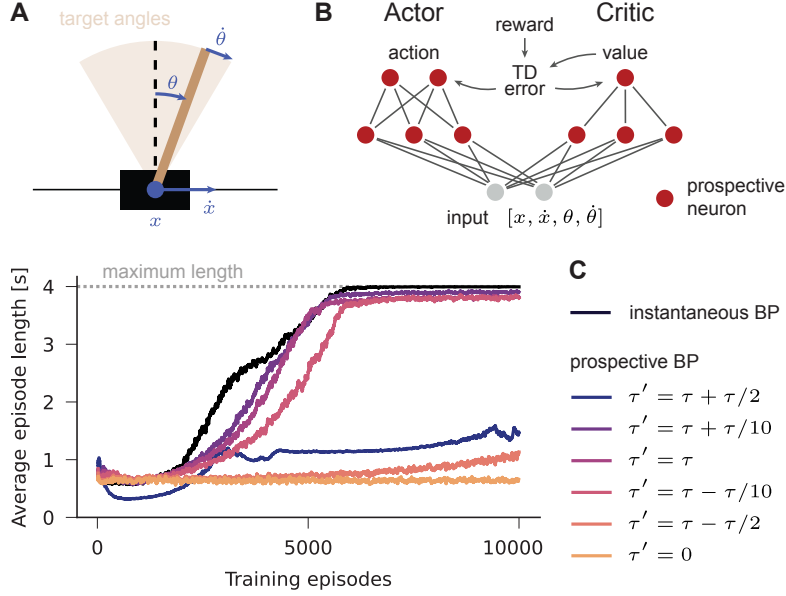


Figure 4: **Prospective dynamics support learning in a control task.** **A.** Visual depiction of the inverted-pendulum task, in which the goal is to balance the pole in an upright position. Reward rate is 1 as long as the pole remains in the “target angles” region and that cart is not too far from the origin. **B.** The current state, comprising the position x and velocity \dot{x} of the cart, and the angle θ and angular velocity $\dot{\theta}$ of the pole, is fed to an actor-critic network. This network outputs an action (left or right) and an estimated value representing the expected discounted reward. At each time step, the estimated value is combined with the reward to compute the temporal difference (TD) error, which then serves as a teaching signal for both the actor and the critic. **C.** The prospective backpropagation algorithm solves the task – maintaining the pole in the target region for 4s – as efficiently as its instantaneous counterpart. Gradually reducing the prospective component of dynamics ($\tau' \rightarrow 0$ with τ' as in Section 2.2) diminishes learning capability, eventually preventing learning altogether. Increasing τ' has similar effects. Small deviations around $\tau' = \tau$ do not significantly affect performance, as seen with the $\tau' = \tau - \tau/10$ and $\tau' = \tau + \tau/10$ curves. We use $\tau = 100\text{ms}$ and report the rolling average of episode length using a 30-episode window, averaged across 5 seeds. Standard deviations (approximately 500ms) are omitted to avoid visual clutter, with no significant differences observed between methods. See Methods section for implementation details.

3.2 Prospective neurons support reward-based learning in control tasks

Having demonstrated the benefits of prospective neurons in simplified settings, we now examine their performance in a behaviorally relevant motor control task: the inverted-pendulum (also known as **Cartpole**, Figure 4A). In this task, the agent must balance a

pole in an upright position for as long as possible, given information about the current cart position, cart velocity, pole angle, and pole angular velocity. The agent receives a reward only when the pole remains within a designated target region.

While this task is typically straightforward for reinforcement learning algorithms in its standard form, we investigate a more challenging variant with three key modifications: First, we operate in the near continuous-time regime (simulation timestep $\Delta t = 1\text{ms}$), where estimating the expected discounted reward becomes notably difficult [Tallec et al., 2019]. Second, we implement purely online learning without leveraging any offline replay or trajectory batching. Third, since the agent’s actions influence future states, delays become even more detrimental than in our previous experiments. Given the Markovian nature of the environment, we implemented the agent as a memory-less feedforward neural network trained with an online continuous version of the advantage actor-critic algorithm [Doya, 2000]. The algorithm and the choice of parameters are discussed in Appendix D.

Figure 4 illustrates the advantage of prospective dynamics (over leaky dynamics) in this task. Most notably, adaptive dynamics enable rapid error signal delivery: the prospective backpropagation algorithm ($\tau = 100\text{ms}$) performs comparably to its instantaneous counterpart. We further investigate how the time-constant mismatch of the prospective component affects learning performance ($\tau' \neq \tau$). For this analysis, we simulate the dynamics described in Section 2.2, varying τ' between 0τ and 1.5τ . The learning trajectories for various τ' values in Figure 4C demonstrate a clear pattern: as dynamics approach those of a leaky integrator ($\tau' = 0\tau$), performance deteriorates progressively until the agent completely fails to learn the task (occurring around $\tau' \approx \tau/2$). This confirms that prospective dynamics are essential for successful task completion. We also note that the agent is still able to learn to solve the task almost perfectly under a relatively small time-constant mismatch on both sides ($\tau' = \tau + \tau/10$, $\tau' = \tau - \tau/10$). This is consistent with the teacher-student tracking example in Fig 3.

3.3 Prospective neurons support the learning of memory-storing non-prospective neurons

Previous sections demonstrated that prospective neurons enable faster signal propagation through neural networks, resulting in improved learning performance. However, these neurons are inherently memory-less in the prospective limit and thus insufficient for solving complex spatio-temporal learning tasks. To address this limitation, we combine them with non-prospective leaky neurons that maintain a memory of past inputs, training the resulting hybrid network with a continuous time version of the online real-time recurrent learning algorithm [Williams and Zipser, 1989, RTRL].

Specifically, we implement a network architecture consisting of one layer of complex-valued non-prospective linear leaky neurons followed by two feedforward layers of prospective neurons, as illustrated in Figure 5A. The non-prospective leaky neurons

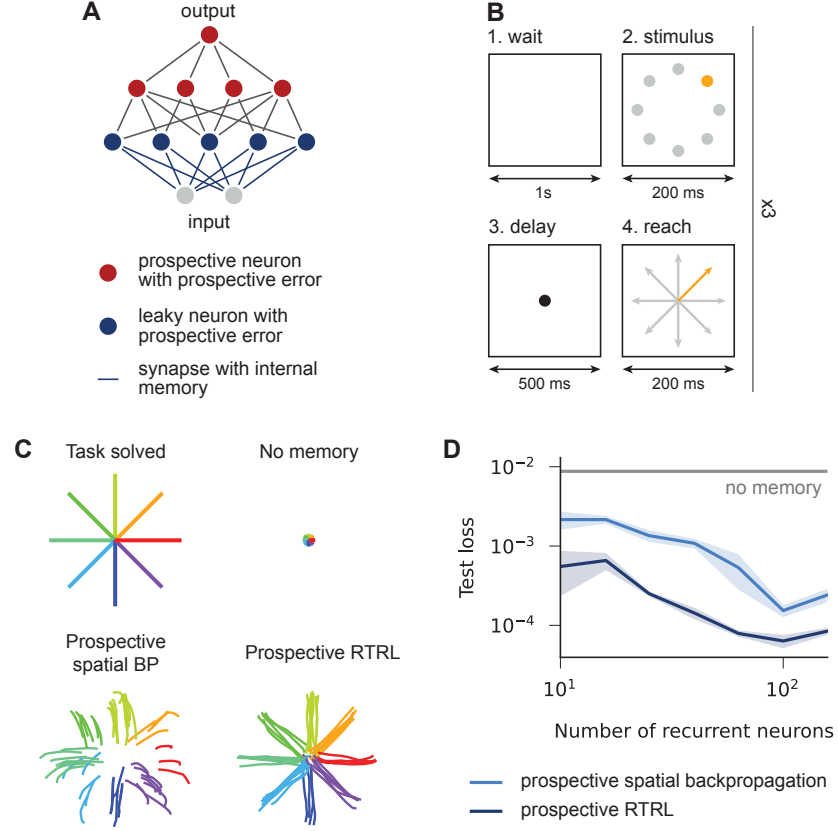


Figure 5: Prospective neurons support learning of memory-storing non-prospective leaky neurons. **A.** We consider a multi-layer neural network combining complex-valued leaky neurons that integrate information over time (first layer) with prospective neurons (in the subsequent two layers). All neurons have prospective errors, following the prospective version of the δ dynamics in (3). The first layer’s role is to store memories of past inputs, while the subsequent layers provide non-linear and instantaneous processing of these memories. **B.** We train the network on a delayed reaching task requiring production of context-dependent target trajectories. **C.** The network successfully solves the task when trained online with prospective real-time recurrent learning. When memory-specific parameters are frozen (prospective spatial backpropagation), the network can only adjust existing movements but cannot refine them to match the desired behavior. When the leaky neurons are replaced by prospective ones, the network loses its memory capacity and produces no movement as its best response. **D.** Learning memory-specific parameters becomes increasingly critical when the number of non-prospective leaky neurons is limited. As this number approaches infinity, the pressure decreases since the feedforward network in the last two layers has a wider range of movements to select from to solve the task.

operate without recurrent connections between them, making exact online gradient calculation tractable [Zucchet et al., 2023]. Using complex-valued neurons (instead of real-valued) ensures that the network remains expressive as a dense linear recurrent layer Orvieto et al. [2023]. Online gradient calculation requires maintaining one internal state per incoming synapse of the leaky neurons – conceptually similar to an eligibility trace [Zenke and Ganguli, 2018, Bellec et al., 2020] – and one per such neuron. We refer to these states as parameter sensitivities s_t^θ , which mathematically correspond to the derivative $s_t^\theta := d_\theta h_t$, where h_t represents the state of the leaky neurons. These sensitivities can then be combined with error signals δ (similar to those in instantaneous spatial backpropagation of errors through the network hierarchy) to estimate the gradient. For example, the update for the time constants τ of the leaky neurons is given by:

$$\tau_\tau \dot{\tau}_t = s_t^\tau \odot \delta_t \quad (14)$$

where \odot denotes the element-wise product. This weight update follows the gradient accurately when error signals δ_t arrive in sync. To ensure this synchronization, the use of prospective neurons and errors for the final two layers is critical, as well as having prospective errors associated with each leaky neuron. Further architectural details and learning rule derivations are provided in the Methods section.

Despite its relative simplicity, this network architecture can learn non-trivial tasks in a purely online manner. We demonstrate this capability using a delayed reaching task (Figure 5B) where the network receives a stimulus indicating a reaching direction, must wait for a “Go” cue, and then move a virtual arm in the corresponding direction. Figure 5C illustrates how different methods perform on this task after processing 500 sequences. With the learning rule described above, the network achieves near-perfect task performance. When memory is removed (e.g., by replacing leaky neurons with prospective ones), the network produces no movement after learning, as it cannot retain the directional information needed for the delayed response. Alternatively, when leaky neurons are present but their time constants remain fixed (prospective spatial BP; similar to reservoir computing [Lukoševičius and Jaeger, 2009]), the network learns to select and modify existing movement patterns but cannot reshape them to produce precisely targeted movements. This limitation becomes particularly pronounced in networks with fewer non-prospective neurons (Figure 5D), where the available variety of movement is more restricted. Notably, with our chosen time constant ($\tau = 100\text{ms}$), we found that networks using standard leaky dynamics (not shown in the figure) instead of prospective dynamics do not outperform the memory-less baseline – highlighting that teaching signal synchronization is essential for effective spatio-temporal learning.

We conclude by noting that this kind of architecture, while exhibiting rudimentary recurrence patterns, has recently demonstrated remarkable power for sequence learning when stacked hierarchically [Gu et al., 2022, Orvieto et al., 2023]. The online learning rule we explore naturally extends to these deeper networks [Zucchet et al., 2023], albeit with minor approximations to the gradient. We thus expect it to be able to tackle more challenging tasks.

4 Discussion

Connections to prospective dynamics. Dynamics similar to the prospective dynamics studied in this paper have been examined in various fields, such as neuroscience, optimization theory, and statistical estimation. Physiological evidence [Ulrich, 2002] has highlighted the ability of biological neurons to phase-advance their output. Drawing from these findings, Mi et al. [2014] studied spike frequency adaptation as a potential mechanism for tracking and phase-advancing traveling waves in attractor networks. The neuronal least action [Senn et al., 2024] and latent equilibrium [Haider et al., 2021] theories propose that this prospective ability may enable neural activity to respond instantaneously to external stimuli. Ellenberger et al. [2024] explore how time constant mismatches of the type we consider in Section 2.2 could support temporal credit assignment. The dynamics we study are a generalization of those of the neuronal least action theory beyond the energy-based systems it considers. Our dynamics are also a generalization of the prediction-correction algorithm [Zhao and Swamy, 1998, Simonetto and Dall’Anese, 2017] in time-varying optimization [Simonetto et al., 2020]. This problem arises in many fields, such as adaptive control [Landau et al., 2011] and online learning [Zinkevich, 2003], and it consists of finding a trajectory of parameters that minimizes a time-varying cost function. The prediction-correction algorithm consists of two parts: the prediction part of the dynamics, which aims to keep the tracking error constant, and a correction part that pushes the current parameters towards the current minimizer. The mathematical details of these two connections are provided in the Methods section. Finally, our prospective dynamics have interesting connections to control and estimation theory. When considering the input received by a neuron as an error term, the prospective input is akin to a proportional-derivative controller (PD) [Minorsky, 1922] that steers neural activity. The prediction-correction view is also reminiscent of Kalman filtering [Kalman, 1960], in which the current state estimate is updated using a model of the true dynamics of the system and corrected using the current observation. Instead of being explicitly defined through a differential equation, as in the Kalman filter, the target is implicitly defined as an equilibrium point here.

Mechanisms to deal with the delays inherent to neural computation. Prospective neurons add to the list of mechanisms that can reduce delays in biological neural networks. At the population level, Knight [1972] and van Vreeswijk and Sompolinsky [1996, 1998] have shown that groups of neurons can respond to external inputs faster than the characteristic time constant of individual neurons. When it comes to learning, this implies that teaching signals could arrive in sync at the population level, which is not precise enough to support the learning of hierarchical networks. At the neuronal level, eligibility traces [Gerstner et al., 2018] have been argued to support temporal credit assignment [Zenke and Ganguli, 2018, Bellec et al., 2020]. Intuitively, the role of these traces is to apply a similar delay to the pre-synaptic term in the learning rule as that of the post-synaptic term, which is the error signal in our case. This can sometimes recover the true gradient signal for neurons that are not recurrently connected

[e.g., Mozer, 1989], assuming the downstream processing is instantaneous. However, the deeper the downstream network, the harder it becomes to mimic the delay it induces on the error with a simple fading memory. As a consequence, there will be a temporal mismatch between the pre- and post-synaptic terms of the learning process, which will be particularly pronounced for remote layers, and learning will be significantly affected. Prospective neurons constitute a unique mechanism in their ability to enable the precise teaching signals required for learning deep networks.

Neural implementation of adaptive dynamics. We have thoroughly discussed the theoretical properties of prospective neurons in Section 1 and have shown that adaptive dynamics constitute a more realistic implementation while maintaining most of the important prospective properties. Critical to the theory of Section 2 and its experimental counterpart is the hypothesis that the adaptation time constant τ_a is faster than the membrane time constant τ . While this is outside the regime in which adaptive neurons traditionally lie [Brette and Gerstner, 2005], we argue that mechanisms that could support fast adaptation exist. For instance, it might be caused by the fast inactivation of sodium currents involved in generating action potentials. In this case, the estimate of the input temporal derivative is interpreted as a sodium current. It has an ultra-fast activation in the range of 1 ms and a fast inactivation in the range of 10 ms (respectively, the gating variables m and h in the Hodgkin-Huxley model [Hodgkin and Huxley, 1952, Gerstner et al., 2014]). Although the sodium current involves a product of these voltage-dependent gating variables, their overall effect can be approximated as the sum of an ultra-fast voltage-dependent excitation (f) and fast inhibition (a) that, together, cause a prospective voltage drive [Brandt et al., 2024].

Methods

Link with neuronal least action (NLA) principle [Senn et al., 2024] latent equilibrium (LE) [Haider et al., 2021] theories. These two theories both study how prospective neurons theoretically enable instantaneous information propagation in energy-based models that rely on prediction errors. The NLA derives the dynamics from a least action principle, while the LE minimizes energy. The difference between these theories relies on which quantities are prospective: input currents/firing rates in the NLA and voltages in the LE. Our theory is a generalization of the NLA to more general dynamical systems; we recover the NLA by having $f_\theta(s, t) = s - \nabla_s E_\theta(s, t)$ with E an energy function. This has important consequences: First, on a conceptual level, it helps decouple the role of prospective dynamics from other modeling choices. Second, it simplifies the mathematical analysis of the dynamics, especially in non-ideal settings. Third, and perhaps more importantly, it enables the decoupling of neural activity from error signals. In particular, activity can be non-prospective when errors are prospective. This is crucial for combining prospective neurons with the RTRL rule we use in Section 3.3.

Link with the prediction-correction algorithm. The prediction-correction algorithms aim to solve a time-varying optimization problem $\min_s E(s, t)$ for all t , using the solutions obtained from previous timesteps. Assuming $f(s, t)$ to be $s - \nabla_s E(s, t)$, the prediction-correction algorithm [Simonetto and Dall’Anese, 2017] minimizes the cost E through

$$\tau \dot{s}_t = \left(\text{Id} - \frac{\partial f}{\partial s}(s_t, t) \right)^{-1} \left(-s_t + f(s_t, t) + \tau \frac{\partial f}{\partial t}(s_t, t) \right) \quad (15)$$

In this setting, s does not represent any neural activity but rather the parameters of the system that are to-be-optimized. While different from the prospective dynamics at first glance, both dynamics are the same. To make this link, we can rewrite the prospective dynamics of Eq. 5 as $d_t [s_t - f_\theta(s_t, t)] = -[s_t - f_\theta(s_t, t)]$. Then, we apply the chain rule to the left-hand side term of the previous equation and invert the obtained Jacobian to get the desired result. The correction part of (15),

$$\tau \dot{s} = (\text{Id} - \partial_s f(s_t, t))^{-1} (-s_t + f(s_t, t)), \quad (16)$$

is the continuous-time version of Newton’s algorithm for finding a solution of $s - f(s, t)$ as if t is fixed. On the other side, the prediction part, $\tau \dot{s}_t = \tau (\text{Id} - \partial_s f(s_t, t))^{-1} \partial_t f(s_t, t)$, aims to keep the error $s_t - f(s_t, t)$ constant and arises from the implicit function theorem [Dontchev and Rockafellar, 2009]. The prospective dynamics we study can thus also be understood from this perspective.

Simulation of prospective dynamics. Several challenges arise when simulating prospective dynamics on digital computers. On one side, numerical integration of the prospective dynamics (5) requires schemes suited for implicit differential equations, as both left- and right-hand sides depend on \dot{s} . Such schemes do not benefit from the extensive theoretical understanding that ordinary differential equations do. On the other side, simulating the equivalent prediction-correction dynamics (15) requires inverting the Jacobian matrix $(\text{Id} - \partial_s f(s_t, t))$ at every time step, which is notoriously expensive. We found that the following Euler-like implicit integration scheme works well for our purposes: we update s by approximating \dot{s}_t as $\Delta t^{-1}[s_{t+\Delta t} - s_t]$ and approximate $d_t[f_\theta(s_t, t)]$ as $\Delta t^{-1}[f_\theta(s_t, t) - f_\theta(s_{t-\Delta t}, t - \Delta t)]$, with Δt being the step size of the integration scheme. That is,

$$s_{t+\Delta t} = s_t + \frac{\Delta t}{\tau} (-s_t + f_\theta(s_t, t)) + f_\theta(s_t, t) - f_\theta(s_{t-\Delta t}, t - \Delta t). \quad (17)$$

While there exists a flourishing literature studying the theoretical properties of numerical schemes for differential algebraic equations [e.g. März, 1992], of which the prospective dynamics is an example, we are not aware of any theoretical guaranties for this specific integration scheme. Yet, we found that it performed similarly to the Euler integration of the prediction-correction dynamics, while ours has a much lower computational footprint.

Experimental setup for Sections 1 and 2. We use a feedforward neural network with 2 hidden layers of size $5 - 20 - 20 - 1$ and a sigmoid non-linearity. The weights of the network are drawn from a normal distribution with variance $1/n_{\text{out}}$ and n_{out} representing the number of output neurons of a given layer. The teacher network has the same size, but its weights are drawn from a distribution with a standard deviation 3 times larger. Each of the 5 features of the input x to the network is a random linear combination of 1000 sine waves with phases φ_i uniformly sampled from $[0, 2\pi]$ and angular velocities ω_i from $[\omega_0, 2\omega_0]$. More precisely, we have

$$(x_t)_i = \sum_{j=1}^{1000} M_{ij}(s_t)_j \quad \text{with } (s_t)_j = \sin(\omega_j t + \varphi_j) \text{ and } M_{ij} \sim \mathcal{N}\left(0, \frac{1}{\sqrt{d_{\text{input}}}}\right) \quad (18)$$

The target output y_t used in the dynamics (3) is the instantaneous processing of x_t by the teacher network.

In Figure 2A, we use the Euler integration scheme with a step size of $\Delta t = 10\text{ms}$ for 1000 sequences of length $T = 100s$. The different weights are fixed throughout this experiment; we only vary τ (its logarithm is drawn uniformly from $[\log 0.05, \log 1]$) and ω_0 (its logarithm is drawn uniformly from $[\log 0.01, \log 0.2]$). We additionally resample all the sine waves for each sequence. The metric we report is the $\max_{t \geq 25s} \|s_t - f_\theta(s_t, t)\|$.

In Figure 2B, we use the integration scheme detailed above with a step size of 0.05 ms . We take ω_0 to be $2\pi/10$. The metric we report is $\|s_t - f_\theta(s_t, t)\|$.

Experimental setup for Section 3.1. The experimental setup is almost the same as the one for the tracking measurements described above. There are a few additional changes: we use a student of size $10 - 100 - 100 - 5$ and a teacher of size $10 - 20 - 20 - 5$, both with ReLU non-linearities. The input combines 100 different sine waves with a characteristic angular velocity of $\omega_0 = 2\pi/10$, and the step size Δt is equal to 1 ms . By default, τ is taken to be equal to $1s$, and each sequence is of length $5s$. Note that we use shorter time sequences than in the tracking experiments here to reduce temporal correlations in the data and thus make gradient-based learning easier. We train the networks over 100 epochs, using the Adam optimizer with a learning rate $\Delta t/\tau_W$ and a cosine scheduler. We test the learned network on a held-out set of 5000 sequences at the end of learning, keeping the same neural dynamics as those during learning. The network weights are updated after each time step. The learning rate is tuned independently for each method using a grid search over $[0.001, 0.003, 0.01, 0.03, 0.1, 0.3, 1]$ and 2 seeds. For the plots with varying parameters (Figure 3B, D and E), we pick the learning rate yielding the best average test loss.

For the recurrent BP and holomorphic EP learning experiments (Figure 1), we used the same setup as above, with the following differences: The architecture of the student network is a continuous Hopfield network with a shifted sigmoid $z \mapsto 1/(1 + e^{-4z+2})$ instead of ReLU. The teacher network is still a feed-forward network using the sigmoid

activation instead of ReLU for faster target computation. Both the teacher and student networks have the same dimensions as in the instantaneous BP experiment. For holomorphic EP, the complex teaching oscillations $\beta = |\beta|e^{i\omega_\beta t}$ have an amplitude of $|\beta| = 0.2$, and a pulsation of $\omega_\beta = 190\pi$ rad/s. The error is decoded from the on-line neuronal oscillations using an exponential moving average filter on a timescale of $\tau_{\text{ema}} = 31.5\text{ms}$. The best parameters for $|\beta|$, ω_β , and τ_{ema} are found through grid search. The training lasts for 200 epochs. For more details on the theory of holomorphic EP, see Appendix C.

Experimental setup for Section 3.2. The simulation of the inverted pendulum is done with Euler integration ($\Delta t = 0.001\text{s}$). At each timestep, the agent can take one of two actions that exert a force either to the left or to the right. An episode ends if either the maximum timestep is reached ($T = 4\text{s}$) or if the pole falls outside an admissible range ($-2.4 \leq x \leq 2.4, -12^\circ \leq \theta \leq 12^\circ$). We note that since prospective tracking takes time to react to discontinuity, we restrict the agent to receive a reward of 1 only when the pole is balancing within a smaller range ($-2 \leq x \leq 2, -8.5^\circ \leq \theta \leq 8.5^\circ$) than the true admissible range and receives a reward of 0 otherwise. This change was critical for the value function estimator and its gradient to converge and learn the end of the episode properly. We parameterize the actor and critic with one layer neural networks consisting of 64 hidden neurons and ReLU activation. We train the two networks for 10000 episodes with a continuous advantage actor-critic (A2C) loss [Doya, 2000, Mnih et al., 2016]. The networks are optimized with stochastic gradient descent with early stopping, where the gradient is calculated with prospective backpropagation dynamics. The learning rate, $\Delta t/\tau_W$, is tuned for prospective BP for each τ 's in the range $[10^{-7}, 10^{-6}, 10^{-5}, 10^{-4}, 10^{-3}]$. We note that within the hyperparameter setups where training is effective, the learning rate did not affect the speed of learning in prospective BP with different τ . A detailed description of the full online prospective A2C training algorithm is provided in the Appendix D.

Learning rule for the network of Section 3.3. In this series of experiments, we study a network comprising one layer of complex-valued leaky neurons, followed by a one hidden-layer feedforward neural network. The latter is exactly the same as what we use in the rest of the paper, including how error signals evolve, and it follows the dynamics of the prospective backpropagation algorithm. We will thus assume that it is always tracking the target trajectory. The complex-valued leaky neurons followed the dynamics:

$$\tau \dot{u}_t = -(1 + i\omega)u_t + Wx_t. \quad (19)$$

Note that here u and W are complex-valued terms. One can think of u as a linearized version of a complex neuron, e.g. [Izhikevich, 2001, Schaffer et al., 2013]. We do not enter further biophysical implementation details and focus on the learning dynamics instead.

We can now compute the loss of the gradient with respect to some of the parameters θ of the leaky neuron

$$\frac{dL}{d\theta} = \int \frac{dL_t}{d\theta} dt = \int \frac{\partial L_t}{\partial u_t} \frac{du_t}{d\theta} dt \quad (20)$$

where we used that u_t only directly affects L_t in the second equation. The term $\partial_{u_t} L_t$ is exactly the δ_t , up to a transpose, of (spatial) backpropagation, so it can be made accessible by having a prospective error “neuron” associated with each leaky neuron. We are left with estimating $d_\theta u_t$. As those leaky neurons are independent of each other, the parameters of one do not affect the state of the others, so we can ignore all the $d_{\theta_i} u_{t,j}$ for $i \neq j$; with a slight abuse of notation, we get rid of the identically zero terms in $d_\theta u_t$ in the following. Importantly, it implies that the non-zero terms of $d_\theta u_t$ are exactly the size of θ , so we can store this extra state in the neuron for τ and ω , and in the synapse for W . We have

$$\tau \frac{d\dot{u}_t}{d\tau} = (-\mathbf{1} + i\omega) \frac{du_t}{d\tau} - \tau^{-1} ((-\mathbf{1} + i\omega)u_t + Wx_t) \quad (21)$$

$$\tau \frac{d\dot{u}_t}{d\omega} = (-\mathbf{1} + i\omega) \frac{du_t}{d\omega} + iu_t \quad (22)$$

$$\tau \frac{d\dot{u}_t}{dW} = \text{diag}[-1 + i\omega] \odot \frac{du_t}{dW} + x_t \mathbf{1}^\top \quad (23)$$

In the last equation, we used $\mathbf{1}$ to denote a vector of ones of appropriate size. Finally, updating those parameters according to

$$\tau_\tau \dot{\tau}_t = -\frac{du_t}{d\tau} \odot \delta_t \quad (24)$$

$$\tau_\omega \dot{\omega}_t = -\frac{du_t}{d\omega} \odot \delta_t \quad (25)$$

$$\tau_W \dot{W}_{t,ij} = -\frac{du_t}{dW_{ij}} \delta_{t,j} \quad (26)$$

will follow the instantaneous gradient. Note that if parameter updates are buffered and not applied, the estimated gradient for all parameters will exactly match the gradient, that is typically computed with backpropagation-through-time. In short, we show that if each neuron and synapse has an additional hidden state that evolves with the equations above, exact online gradient calculation is possible for the kind of neural networks we study here. In our experiments, it will not be entirely true as plasticity is always on and as we initialize neural activity at a default state that is not yet on the target trajectory.

Experimental setup for Section 3.3 Each sequence of the delayed reach task consists of 3 subsequences of the form “Wait”, “Stimulus”, “Delay” and “Reach” which last respectively 1s, 200ms, 500ms and 200ms. A sequence is thus 5.7s. We use $\Delta t = 1\text{ms}$. The input given to the network has 9 dimensions, 8 for the one-hot encoded

version of each stimulus, and 1 for the “Go” cue in the “Reach” phase. The target movement, which is 2-dimensional, is a line from the center to a location encoded by the stimuli. We use the mean squared distance to the target movement as the loss to train the network. The network consists of one layer of leaky neurons and two layers of prospective neurons. We fix the hidden layer of the multi-layer perceptron to have 100 neurons and vary the number of leaky neurons from 10 to 158 in Figure 5. The time constants of the prospective neurons (u and δ) are fixed at 100ms. The logs of the time constants of the leaky neurons are randomly initialized uniformly from $[\log 100\text{ms}, \log 10\text{s}]$. In the prospective RTRL algorithm, the multilayer perceptron is trained using the prospective backpropagation algorithm, and the leaky neurons are trained with the learning rule described above. We train all networks on 500 sequences without any replay, which more or less corresponds to all possible stimulus combinations, using the Adam optimizer and a cosine learning rate scheduler with an initial learning rate of 0.01, obtained after a grid search on both methods. Note that it corresponds to having a characteristic time constant of 10s for all learned parameters. We then measure the test loss on 500 sequences that may have been seen during training. The goal of this loss is not to test generalization but rather the optimization abilities of our learning rule.

Acknowledgments

We thank Jean-Pascal Pfister, Frederico Benitez, and Alexander Meulemans for early discussions on the project, as well as Seijin Kobayashi and Asier Mujika for their advice on reinforcement learning methods. This research was supported by the Swiss National Science Foundation (grant numbers PZ00P3_186027, PCEFP3_202981 and TMPFP3_210282), an ETH Research Grant (ETH-23 21-1), EU’s Horizon Europe Research and Innovation Program (CONVOLVE, grant agreement number 101070374) funded through SERI (ref 1131-52302), and the Novartis Research Foundation.

References

- Eric R Kandel, James H Schwartz, Thomas M Jessell, Steven Siegelbaum, A James Hudspeth, Sarah Mack, and others. *Principles of neural science*, volume 4. McGraw-hill New York, 2000.
- Paul Werbos. *Beyond regression: new tools for prediction and analysis in the behavioral sciences*. Ph.D. thesis, Harvard University, 1974.
- David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams. Learning representations by back-propagating errors. *Nature*, 323(6088), 1986.
- Dong-Hyun Lee, Saizheng Zhang, Asja Fischer, and Yoshua Bengio. Difference target propagation. *Machine Learning and Knowledge Discovery in Database*, 2015.

- Timothy P. Lillicrap, Daniel Cownden, Douglas B. Tweed, and Colin J. Akerman. Random synaptic feedback weights support error backpropagation for deep learning. *Nature Communications*, 7(1), 2016.
- Benjamin Scellier and Yoshua Bengio. Equilibrium propagation: bridging the gap between energy-based models and backpropagation. *Frontiers in Computational Neuroscience*, 11, 2017.
- James C. R. Whittington and Rafal Bogacz. An approximation of the error backpropagation algorithm in a predictive coding network with local hebbian synaptic plasticity. *Neural Computation*, 29(5), 2017.
- Blake A. Richards, Timothy P. Lillicrap, Philippe Beaudoin, Yoshua Bengio, Rafal Bogacz, Amelia Christensen, Claudia Clopath, Rui Ponte Costa, Archy de Berker, Surya Ganguli, Colleen J. Gillon, Danijar Hafner, Adam Kepecs, Nikolaus Kriegeskorte, Peter Latham, Grace W. Lindsay, Kenneth D. Miller, Richard Naud, Christopher C. Pack, Panayiota Poirazi, Pieter Roelfsema, João Sacramento, Andrew Saxe, Benjamin Scellier, Anna C. Schapiro, Walter Senn, Greg Wayne, Daniel Yamins, Friedemann Zenke, Joel Zylberberg, Denis Therien, and Konrad P. Kording. A deep learning framework for neuroscience. *Nature Neuroscience*, 22(11), 2019.
- Alexander Meulemans, Nicolas Zucchet, and Seijin Kobayashi. The least-control principle for local learning at equilibrium. In *Advances in Neural Information Processing Systems*, 2022.
- R.C. Miall and D.M. Wolpert. Forward models for physiological motor control. *Neural Networks*, 9(8), 1996.
- Daniel M Wolpert, R Chris Miall, and Mitsuo Kawato. Internal models in the cerebellum. *Trends in cognitive sciences*, 2(9), 1998.
- Aditya Gilra and Wulfram Gerstner. Predicting non-linear dynamics by stable local learning in a recurrent spiking neural network. *eLife*, 6, 2017.
- Eugene M. Izhikevich. Solving the distal reward problem through linkage of STDP and dopamine signaling. *Cerebral Cortex*, 17(10), 2007.
- Wulfram Gerstner, Marco Lehmann, Vasiliki Liakoni, Dane Corneil, and Johanni Brea. Eligibility traces and plasticity on behavioral time scales: experimental support of neohebbian three-factor learning rules. *Frontiers in neural circuits*, 12, 2018.
- Friedemann Zenke and Surya Ganguli. Superspike: supervised learning in multilayer spiking neural networks. *Neural Computation*, 30(6), 2018.
- Guillaume Bellec, Franz Scherr, Anand Subramoney, Elias Hajek, Darjan Salaj, Robert Legenstein, and Wolfgang Maass. A solution to the learning dilemma for recurrent networks of spiking neurons. *Nature Communications*, 11(1):3625, 2020.

- Yuanyuan Mi, C. C. Alan Fung, K. Y. Michael Wong, and Si Wu. Spike frequency adaptation implements anticipative tracking in continuous attractor neural networks. In *Advances in Neural Information Processing Systems*, 2014.
- Paul Haider, Benjamin Ellenberger, Laura Kriener, Jakob Jordan, Walter Senn, and Mihai A. Petrovici. Latent Equilibrium: A unified learning theory for arbitrarily fast computation with arbitrarily slow neurons. In *Advances in Neural Information Processing*, 2021.
- Walter Senn, Dominik Dold, Akos F. Kungl, Benjamin Ellenberger, Jakob Jordan, Yoshua Bengio, João Sacramento, and Mihai A. Petrovici. A neuronal least-action principle for real-time learning in cortical circuits. *eLife*, 2024.
- Larry F Abbott. Lapicque’s introduction of the integrate-and-fire model neuron (1907). *Brain Research Bulletin*, 50(5-6), 1999.
- Romain Brette and Wulfram Gerstner. Adaptive exponential integrate-and-fire model as an effective description of neuronal activity. *Journal of Neurophysiology*, 94(5), 2005.
- Wulfram Gerstner, Werner M Kistler, Richard Naud, and Liam Paninski. *Neuronal dynamics: From single neurons to networks and models of cognition*. Cambridge University Press, 2014.
- Anant Agarwal and Jeffrey Lang. *Foundations of analog and digital electronic circuits*. Elsevier, 2005.
- Stephen Grossberg. Competitive learning: From interactive activation to adaptive resonance. *Cognitive Science*, 11(1), 1987.
- Arild Nøkland. Direct feedback alignment provides learning in deep neural networks. In *Advances in Neural Information Processing Systems*, 2016.
- João Sacramento, Rui P. Costa, Yoshua Bengio, and Walter Senn. Dendritic cortical microcircuits approximate the backpropagation algorithm. In *Advances in Neural Information Processing Systems*, 2018.
- Blake A. Richards and Timothy P. Lillicrap. Dendritic solutions to the credit assignment problem. *Current Opinion in Neurobiology*, 54, 2019.
- Fabian A. Mikulasch, Lucas Rudelt, Michael Wibral, and Viola Priesemann. Where is the error? Hierarchical predictive coding through dendritic error computation. *Trends in Neurosciences*, 46(1), 2023.
- David H Ackley, Geoffrey E Hinton, and Terrence J Sejnowski. A learning algorithm for Boltzmann machines. *Cognitive science*, 9(1), 1985.

- Pierre Baldi and Fernando Pineda. Contrastive learning and neural oscillations. *Neural Computation*, 3(4), 1991.
- Javier R. Movellan. Contrastive Hebbian learning in the continuous Hopfield model. In *Connectionist Models*. 1991.
- Axel Laborieux and Friedemann Zenke. Holomorphic equilibrium propagation computes exact gradients through finite size oscillations. In *Advances in Neural Information Processing Systems*, 2022.
- Shaojie Bai, J Zico Kolter, and Vladlen Koltun. Deep equilibrium models. In *Advances in Neural Information Processing Systems*, 2019.
- Luís B. Almeida. Backpropagation in perceptrons with feedback. In Rolf Eckmiller and Christoph v.d. Malsburg, editors, *Neural Computers*. Springer Berlin Heidelberg, 1989.
- Fernando J. Pineda. Recurrent backpropagation and the dynamical approach to adaptive neural computation. *Neural Computation*, 1(2), 1989.
- Corentin Tallec, Léonard Blier, and Yann Ollivier. Making deep Q-learning methods robust to time discretization. In *International Conference on Machine Learning*, 2019.
- Kenji Doya. Reinforcement learning in continuous time and space. *Neural Computation*, 12(1), 2000.
- Ronald J. Williams and David Zipser. A learning algorithm for continually running fully recurrent neural networks. *Neural Computation*, 1(2):270–280, 1989.
- Nicolas Zucchet, Robert Meier, Simon Schug, Asier Mujika, and João Sacramento. Online learning of long-range dependencies. In *Advances in Neural Information Processing Systems*, 2023.
- Antonio Orvieto, Samuel L. Smith, Albert Gu, Anushan Fernando, Caglar Gulcehre, Razvan Pascanu, and Soham De. Resurrecting recurrent neural networks for long sequences. In *International Conference on Machine Learning*, 2023.
- Mantas Lukoševičius and Herbert Jaeger. Reservoir computing approaches to recurrent neural network training. *Computer Science Review*, 3(3), 2009.
- Albert Gu, Karan Goel, and Christopher Ré. Efficiently modeling long sequences with structured state spaces. In *International Conference on Learning Representations*, 2022.
- Daniel Ulrich. Dendritic resonance in rat neocortical pyramidal cells. *Journal of Neurophysiology*, 87(6), 2002.

- Benjamin Ellenberger, Paul Haider, Jakob Jordan, Kevin Max, Ismael Jaras, Laura Kriener, Federico Benitez, and Mihai A. Petrovici. Backpropagation through space, time, and the brain. *arXiv arXiv:2403.16933*, 2024.
- Y. Zhao and M.N.S. Swamy. A novel technique for tracking time-varying minimum and its applications. In *IEEE Canadian Conference on Electrical and Computer Engineering*, volume 2, 1998.
- Andrea Simonetto and Emiliano Dall’Anese. Prediction-correction algorithms for time-varying constrained optimization. *IEEE Transactions on Signal Processing*, 65(20), 2017.
- Andrea Simonetto, Emiliano Dall’Anese, Santiago Paternain, Geert Leus, and Georgios B. Giannakis. Time-varying convex optimization: Time-structured algorithms and applications. *Proceedings of the IEEE*, 108(11), 2020.
- Ioan Doré Landau, Rogelio Lozano, Mohammed M’Saad, and Alireza Karimi. *Adaptive control: Algorithms, analysis and applications*. Communications and Control Engineering. Springer London, 2011.
- Martin Zinkevich. Online convex programming and generalized infinitesimal gradient ascent. In *International Conference on Machine Learning*, 2003.
- Nicolas Minorsky. Directional stability of automatically steered bodies. *Journal of the American Society for Naval Engineers*, 34(2), 1922.
- Rudolph Emil Kalman. A new approach to linear filtering and prediction problems. *Transactions of the ASME-Journal of Basic Engineering*, 82(Series D), 1960.
- Bruce W. Knight. Dynamics of encoding in a population of neurons. *The Journal of General Physiology*, 59(6), 1972.
- C. van Vreeswijk and H. Sompolinsky. Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science*, 274(5293), 1996.
- C. van Vreeswijk and H. Sompolinsky. Chaotic balanced state in a model of cortical circuits. *Neural Computation*, 10(6), 1998.
- Michael C Mozer. A focused backpropagation algorithm for temporal pattern recognition. *Complex Systems*, 3, 1989.
- Alan L Hodgkin and Andrew F Huxley. A quantitative description of membrane current and its application to conduction and excitation in nerve. *The Journal of Physiology*, 117(4), 1952.
- Simon Brandt, Mihai Alexandru Petrovici, Walter Senn, Katharina Anna Wilmes, and Federico Benitez. Prospective and retrospective coding in cortical neurons. *arXiv preprint arXiv:2405.14810*, 2024.

- Asen L Dontchev and R Tyrrell Rockafellar. *Implicit functions and solution mappings*. 2009.
- Roswitha März. Numerical methods for differential algebraic equations. *Acta Numerica*, 1, 1992.
- Volodymyr Mnih, Adria Puigdomenech Badia, Mehdi Mirza, Alex Graves, Timothy Lillicrap, Tim Harley, David Silver, and Koray Kavukcuoglu. Asynchronous methods for deep reinforcement learning. In *International Conference on Machine Learning*, 2016.
- Eugene M Izhikevich. Resonate-and-fire neurons. *Neural Networks*, 14(6-7), 2001.
- Evan S. Schaffer, Srdjan Ostojic, and L. F. Abbott. A complex-valued firing-rate model that approximates the dynamics of spiking networks. *PLoS Computational Biology*, 9(10), 2013.
- Yann LeCun. A theoretical framework for back-propagation. In *Proceedings of the 1998 Connectionist Models Summer School*, 1988.
- Nicolas Zucchet and João Sacramento. Beyond backpropagation: implicit gradients for bilevel optimization. *Neural Computation*, 34(12), 2022.
- Boris T Polyak. Introduction to optimization. *New York, Optimization Software*, 1987.
- Alexey Popkov. Gradient methods for nonstationary unconstrained optimization problems. *Automation and Remote Control*, 66, 2005.
- Axel Laborieux and Friedemann Zenke. Improving equilibrium propagation without weight symmetry through Jacobian homeostasis. In *International Conference on Learning Representations*, 2024.
- Vidyesh Rao Aniseti, A. Kandala, B. Scellier, and J. M. Schwarz. Frequency propagation: Multi-mechanism learning in nonlinear physical networks. *Neural Computation*, 36(4), 2024.
- Robert Tjarko Lange. gymnax: A JAX-based Reinforcement Learning Environment Library, 2022. URL <http://github.com/RobertTLange/gymnax>.

A Theoretical derivations

A.1 Derivation of the error backpropagation dynamics in continuous time

In Section 1.1, we claimed that the dynamics (3) compute the error signal of backpropagation. We prove this statement by leveraging the Lagrangian derivation of LeCun [1988]. To this extent, we introduce the Lagrangian

$$\mathcal{L}_\theta(u, \delta, t) = \ell(u, t) + \delta^\top (f_\theta(u, t) - u) \quad (27)$$

and want to compute the derivative with respect to θ of $\ell(u_t^*, t)$ s.t. $u_t^* = f_\theta(u_t^*, t)$. Note that we write the dependency on time t to be consistent with Section 1.1, but we consider it fixed for now. The Lagrange multiplier method provides a way to compute the derivative of the loss: it is equivalent to computing $d_\theta \mathcal{L}_\theta(u_t^*, \delta_t^*, t)$ with u_t^* and δ_t^* such that $\partial_u \mathcal{L}_\theta(u_t^*, \delta_t^*, t) = 0$ and $\partial_\delta \mathcal{L}_\theta(u_t^*, \delta_t^*, t) = 0$ as

$$\frac{d}{d\theta} \mathcal{L}_\theta(u_t^*, \delta_t^*, t) = \frac{\partial \mathcal{L}_\theta}{\partial \theta}(u_t^*, \delta_t^*, t) + \frac{\partial \mathcal{L}_\theta}{\partial u}(u_t^*, \delta_t^*, t) \frac{du_t^*}{d\theta} + \frac{\partial \mathcal{L}_\theta}{\partial \delta}(u_t^*, \delta_t^*, t) \frac{d\delta_t^*}{d\theta} \quad (28)$$

$$= \frac{\partial \mathcal{L}_\theta}{\partial \theta}(u_t^*, \delta_t^*, t) + 0 + 0 \quad (29)$$

$$= \frac{\partial f_\theta}{\partial \theta}(u_t^*, t)^\top \delta_t^*. \quad (30)$$

We therefore have a simple way to estimate the gradient, given that u_t^* and δ_t^* are accessible. In the case of $f_\theta(u, t) = W\rho(u)$ with $u_t^0 = x_t$, as in Section 1.1, we have

$$\frac{d}{d\theta} \mathcal{L}_\theta(u_t^*, \delta_t^*, t) = \rho(u_t^*) \delta_t^{*\top}, \quad (31)$$

which is the usual backpropagation update. Note that there is a slight abuse of notation in the previous equation, as this update only applies to the entries of the weight matrix that are not identically zero.

We have justified the θ -update and are now left with computing u_t^* and δ_t^* . One way to reach this equilibrium is to run the coupled dynamical system

$$\begin{cases} \tau \dot{u}_t = \frac{\partial \mathcal{L}_\theta}{\partial \delta}^\top(u_t, \delta_t, t) \\ \tau \dot{\delta}_t = \frac{\partial \mathcal{L}_\theta}{\partial u}^\top(u_t, \delta_t, t) \end{cases} \quad (32)$$

that is

$$\begin{cases} \tau \dot{u}_t = -u_t + W\rho(u_t) \text{ and } u_t^* = x_t \\ \tau \dot{\delta}_t = -\delta_t + \rho'(u_t)W^\top \delta_t \text{ and } \delta_t^L = \nabla \ell(u_t^L, y_t) \end{cases} \quad (33)$$

for our f_θ of interest. If the time dependency in f_θ and L can be ignored, the equilibrium points of these equations satisfy the stationary conditions $\partial_u \mathcal{L}_\theta(u_t^*, \delta_t^*, t) = 0$ and $\partial_\delta \mathcal{L}_\theta(u_t^*, \delta_t^*, t) = 0$, and we can use them to compute the θ -update.

However, in this paper, we are interested in the setting in which the assumption that f_θ and l are independent of time breaks. Finding neural activity u and error δ trajectories that satisfy the stationary equations of the Lagrangian becomes an equilibrium tracking problem. Following the leaky error backpropagation dynamics of Eq. 33 will inevitably lead to delays in neural activity and poor error signal synchronization, as we analyzed in Theorem 1. Instead, we apply the prospective dynamics to always track equilibrium. Finally, we update θ online, using the neural activity u_t and error signal δ_t currently available. This yields our *Prospective BP* algorithm:

$$\begin{cases} \tau \dot{u}_t = -u_t + W_t \left(\rho(u_t) + \tau \rho(\dot{u}_t) \right) \\ \tau \dot{\delta}_t = -\delta_t + \rho'(u_t) W_t^\top \delta_t + \tau \frac{d}{dt} [\rho'(u_t) W_t^\top \delta_t] \\ \tau_W \dot{W}_t = -\delta_t \rho(u_t)^\top \end{cases} \quad (34)$$

In our simulations of Section 3, we use these dynamics with $f_\theta(u, t) = W\rho(u)$ and a feedforward architecture (W lower diagonal), and measure the loss on the last layer of the network.

As a side note, we highlight the links and differences between the algorithm derived above and (recurrent) backpropagation (RBP) equations. In short, it runs the two phases of (R)BP at the same time. Let us explain why. In (R)BP, the first phase consists in finding an equilibrium satisfying

$$u_t^* = f_\theta(u_t^*, t). \quad (35)$$

The resulting equilibrium can be expressed explicitly, as in the feedforward example we used in Section 1, or, in the general case, requires to run dynamics akin to

$$\tau_u \dot{u}_t = -u_t + f_\theta(u_t, \theta). \quad (36)$$

The second phase requires computing the error signal:

$$\delta_t^* = \left(\text{Id} - \frac{\partial f_\theta}{\partial u}(u_t^*, t) \right)^{-\top} \nabla_u \ell(u_t^*, t). \quad (37)$$

Interestingly, this equality can be both derived from the Lagrangian perspective ($\partial_u \mathcal{L}_\theta(u_t^*, \delta_t^*, t) = 0$), or by leveraging the implicit function theorem. It holds both in the feedforward case (acyclic computational graph) or in the recurrent case (cyclic computational graph). When the underlying network is feedforward, it can easily be computed going backward in the network hierarchy [Rumelhart et al., 1986]. When the network is recurrent, this is more tedious and this error term is usually computed by solving the linear system:

$$\left(\text{Id} - \frac{\partial f_\theta}{\partial u}(u_t^*, t) \right)^\top \delta_t^* = \nabla_u \ell(u_t^*, t), \quad (38)$$

see, e.g., [Almeida, 1989, Pineda, 1989, Zucchet and Sacramento, 2022]. One way to solve it, assuming f_θ and L to be independent of time, is to run the dynamics

$$\tau_\delta \dot{\delta}_t = -\delta_t + \frac{\partial f_\theta}{\partial u}(u_t^*, t)^\top \delta_t + \nabla_u \ell(u_t^*, t) \quad (39)$$

One can remark that the δ dynamics can be run simultaneously to the u one, as δ does not influence its evolution. This corresponds to the leaky error backpropagation algorithm of (33).

To summarize, the continuous-time backpropagation algorithm can be obtained by framing (recurrent) backpropagation as an equilibrium-based algorithm and remarking that the two phases can be run at the same time. This way, we can get rid of one of the shortcomings of backpropagation, specifically the need for signals specifying in which phase we are and the update locking issue. A crucial step needed to derive this algorithm is to remark that the instantaneous rate-based models used in computational neuroscience and machine learning are by essence equilibrium-based.

A.2 Proof of Theorem 1

We here restate Theorem 1 and demonstrate it.

Theorem 1. *Let f_θ be such that the biggest eigenvalue of the symmetric part of the Jacobian $\partial_s f(s, t)$ is always smaller than a constant $1 - \mu$ for $\mu > 0$. Let s_t^* satisfy $s_t^* = f_\theta(s_t^*, t)$ for all t and $\|\dot{s}_t^*\| \leq \gamma$. Then, the trajectory of states s_t obtained by integrating the leaky dynamics $\tau \dot{s}_t = -s_t + f_\theta(s_t, t)$ verifies*

$$\limsup_{t \rightarrow \infty} \|s_t - s_t^*\| \leq \frac{\gamma \tau}{\mu}.$$

Furthermore, this bound is tight, that is we can find a f for which the upper bound is reached.

Proof. Throughout the proof, we drop the θ subscript in f_θ for simplicity. We consider the function $g : \alpha \mapsto -s_t^* - \alpha(s_t - s_t^*) + f(s_t^* + \alpha(s_t - s_t^*), t)$ and integrate it between 0 and 1 to obtain the following relationship:

$$-s_t + f(s_t, t) = g(1) \quad (40)$$

$$= g(0) + \int_0^1 \frac{dg}{d\alpha}(\alpha) d\alpha \quad (41)$$

$$= 0 + \left[\int_0^1 \left(-\text{Id} + \frac{\partial f}{\partial s}(s_t^* + \alpha(s_t - s_t^*), t) \right) d\alpha \right] (s_t - s_t^*). \quad (42)$$

In the last line, we used that s_t^* is an equilibrium satisfying $s_t^* = f(s_t^*, t)$ so that $g(0) = 0$. It follows that

$$(-s_t + f(s_t, t))^\top (s_t - s_t^*) \leq -\mu \|s_t - s_t^*\|^2. \quad (43)$$

To prove this, we first define J_t as the integral from the equation above, and, multiplying the previous equation by $(s_t - s_t^*)$, we have

$$(s_t - s_t^*)^\top (-s_t + f(s_t, t)) = (s_t - s_t^*)^\top J_t (s_t - s_t^*) \quad (44)$$

$$= (s_t - s_t^*)^\top \frac{J_t + J_t^\top}{2} (s_t - s_t^*) \quad (45)$$

$$\leq -\mu \|s_t - s_t^*\|^2 \quad (46)$$

The second equality holds as a quadratic form with a non-symmetric matrix takes the same values as the quadratic form with its symmetric part. The last inequality comes from the assumption on the eigenvalues of the symmetric part of $\partial_s f(s, t)$ being smaller than $1 - \mu$.

Let us now look at the Lyapunov function $V(t) := \frac{1}{2} \|s_t - s_t^*\|^2$ that measures how far the current estimate s_t is to the equilibrium s_t^* . Its temporal derivative is equal to

$$\dot{V}(t) = (s_t - s_t^*)^\top \dot{s}_t - (s_t - s_t^*)^\top \dot{s}_t^* \quad (47)$$

$$= \tau^{-1} (s_t - s_t^*)^\top (-s_t + f(s_t, t)) - (s_t - s_t^*)^\top \dot{s}_t^* \quad (48)$$

$$\leq -\tau^{-1} \mu \|s_t - s_t^*\|^2 + \|s_t - s_t^*\| \|\dot{s}_t^*\| \quad (49)$$

$$= -2\tau^{-1} \mu V(t) + \sqrt{2V(t)} \gamma. \quad (50)$$

The inequality is justified by the inequality we proved above and the Cauchy-Schwartz inequality. We can now study the sign of the upper bound on $\dot{V}(t)$ and observe that $\dot{V}(t) < 0$ whenever $V(t) > \frac{\tau^2 \gamma^2}{2\mu^2}$. It then implies that $\limsup_{t \rightarrow \infty} V(t) = \frac{\tau^2 \gamma^2}{2\mu^2}$, which is the desired result.

We now show that the bound is tight by exhibiting an example for which the inequality in the Theorem statement is an equality. For that, we take $t = t$ and follow the dynamics $\dot{s}_t = -s_t/2 + t$ ($f(s, t) = s/2 + t$, $\tau = 1, \mu = 1/2$), we obtain $s_t^* = 2t$, so that $\frac{\gamma}{\mu} = 4$. Alternatively, solving the differential equation gives $s_t = (s_0 + 4) \exp(-t/2) + 2t - 4$ so $\|s_t^* - s_t\| \rightarrow 4$. \square

Remarks. Let us remark on several things:

- The result and the proof are inspired by a result from [Polyak, 1987, Chapter 6] for time-varying optimization, which assumes $f(s, t) = \nabla_s E(s, t)$. We extend this result to general f , which requires changing the μ strong convexity assumption needed in the time-varying optimization setting to the assumption we have on the eigenvalues of the Jacobian in Theorem 1. A discrete-time version of this result can be obtained by adapting the proof of Popkov [2005].
- In the main text, we claimed that \dot{s}_t^* both increases with how fast external input varies and depends on the geometry of $\partial_s f_\theta$. This is because the implicit function

theorem applied to $s_t^* - f_\theta(s_t^*, t) = 0$ gives

$$\dot{s}_t^* = - \left(\text{Id} - \frac{\partial f_\theta}{\partial s}(s_t^*, t) \right)^{-1} \frac{\partial f_\theta}{\partial t}(s_t^*, t). \quad (51)$$

The inverse matrix reflects the geometry of the network and $\partial_t f_\theta$ how fast the inputs to the neurons change.

A.3 Proof of Theorem 2

Theorem 2. *Let s follow the prospective dynamics (5). Then, assuming that $\partial_s f_\theta(s_t, t)$ is always invertible and f_θ is Lipschitz continuous in t on that trajectory, we have*

$$\|s_t - f_\theta(s_t, t)\| = \|s_0 - f_\theta(s_0, 0)\| \exp\left(-\frac{t}{\tau}\right)$$

and

$$\limsup_{t \rightarrow \infty} \|s_t - s_t^*\| = 0,$$

i.e., after an initial exponential convergence phase, s_t and s_t^* coincide.

Proof. The first result follows from rewriting Equation 5 as $\tau d_t [s_t - f_\theta(s_t, t)] = -[s_t - f_\theta(s_t, t)]$. It follows that $s_t - f_\theta(s_t, t) = c_0 e^{-\frac{t-t_0}{\tau}}$, $\lim_{t \rightarrow \infty} \|s_t - f_\theta(s_t, t)\| = 0$ and $\limsup_{t \rightarrow \infty} \|s_t - s_t^*\| = 0$. The assumptions on f ensure that s_t^* is well defined and evolves continuously. \square

B Physical implementation of prospective neurons

B.1 Adaptive neurons can be prospective

In the main text, we have claimed that for u and a following the adaptive neuron dynamics

$$\begin{aligned} \tau \dot{u}_t &= -u_t + \left(1 + \frac{\tau}{\tau_a}\right) f_\theta(u_t, t) - \frac{\tau}{\tau_a} a_t \\ \tau_a \dot{a}_t &= -a_t + f_\theta(u_t, t), \end{aligned} \quad (52)$$

we have

$$\tau \dot{u}_t = -u_t + f_\theta(u_t, t) + \tau d_t f_\theta(u_t, t) + \tau \tau_a d_t^2 f_\theta(u_t, t) + O(\tau \tau_a^2). \quad (53)$$

We prove this statement. First, remark that a_t is a low-pass filter version of $f_\theta(u_t, t)$, with a time-constant τ_a so that

$$a_t = \frac{1}{\tau_a} \int_0^t \exp\left(-\frac{t-t'}{\tau_a}\right) f_\theta(u_{t'}, t') dt'. \quad (54)$$

Integration by parts (integrating the exponential, differentiating f_θ) then gives

$$a_t = \left[\exp\left(-\frac{t-t'}{\tau_a}\right) f_\theta(u_{t'}, t') \right]_0^t - \int_0^t \exp\left(-\frac{t-t'}{\tau_a}\right) \frac{d}{dt} [f_\theta(u_{t'}, t')] dt' \quad (55)$$

$$= f_\theta(u_t, t) - \exp\left(-\frac{t}{\tau_a}\right) f_\theta(u_0, 0) - \int_0^t \exp\left(-\frac{t-t'}{\tau_a}\right) \frac{d}{dt} [f_\theta(u_{t'}, t')] dt'. \quad (56)$$

Therefore, assuming that t is sufficiently far from the boundary condition $t = 0$,

$$\frac{f_\theta(u_t, t) - a_t}{\tau_a} = \frac{1}{\tau_a} \int_0^t \exp\left(-\frac{t-t'}{\tau_a}\right) \frac{d}{dt} [f_\theta(u_{t'}, t')] dt', \quad (57)$$

holds. We can now use this result to derive Eq. 53. Let us first remark that $\tau_a^{-1} (f_\theta(u_t, t) - a_t)$ converges to $d_t f_\theta(u_t, t)$ when τ_a goes to 0. We are interested in the first-order error in τ_a between these two quantities:

$$\frac{f_\theta(u_t, t) - a_t}{\tau_a} - \frac{d}{dt} [f_\theta(u_t, t)] \quad (58)$$

$$= \frac{1}{\tau_a} \int_0^t \exp\left(-\frac{t-t'}{\tau_a}\right) \left(\frac{d}{dt} [f_\theta(u_{t'}, t')] - \frac{d}{dt} [f_\theta(u_t, t)] \right) dt' \quad (59)$$

$$= \frac{1}{\tau_a} \int_0^t \exp\left(-\frac{t-t'}{\tau_a}\right) \left((t' - t) \frac{d^2}{dt^2} [f_\theta(u_t, t)] + O((t' - t)^2) \right) dt' \quad (60)$$

$$= \tau_a \frac{d^2}{dt^2} [f_\theta(u_t, t)] + O(\tau_a^2) \quad (61)$$

Note that in all these calculations, we have assumed $t \gg 0$. The first equation uses the fact that the integral of $\tau_a^{-1} \exp(-\tau_a^{-1}(t-t'))$ between 0 and ∞ is 1. The second one leverages the Taylor expansion of $d_t f_\theta(u_t, t)$. For it to be mathematically rigorous, one must assume that the third-order time derivative of $f_\theta(u_t, t)$ is uniformly bounded and make use of the Lagrange formulation of Taylor's remainder. Finally, the last one comes from the standard integral values

$$\frac{1}{\tau_a} \int_0^\infty (t' - t) \exp\left(-\frac{t-t'}{\tau_a}\right) dt' = \tau_a \quad (62)$$

and

$$\frac{1}{\tau_a} \int_0^\infty (t' - t)^2 \exp\left(-\frac{t-t'}{\tau_a}\right) dt' = 2\tau_a^2. \quad (63)$$

It follows that the adaptive neuron dynamics become

$$\tau \dot{u}_t = -u_t + f_\theta(u_t, t) + \tau d_t f_\theta(u_t, t) + \tau \tau_a d_t^2 f_\theta(u_t, t) + O(\tau \tau_a^2). \quad (64)$$

B.2 Time constant mismatch

Here we derive the first order deviation in $(\tau' - \tau)$ of the trajectory u_t when there is a mismatch between the prospective input and the neuron time constants:

$$\tau \dot{u}_t = -u_t + f_\theta(u_t, t) + \tau' \frac{d}{dt} [f_\theta(u_t, t)]. \quad (65)$$

We start by making the difference $(\tau' - \tau)$ appear in Eq. 65:

$$\tau \dot{u}_t = -u_t + f_\theta(u, t) + \tau \frac{d}{dt} [f_\theta(u_t, t)] + (\tau' - \tau) \frac{d}{dt} [f_\theta(u_t, t)]. \quad (66)$$

We see that the equation without mismatch appears, and we know by Theorem 2 that u_t^* is the solution of this equation after a transitory period. Therefore, we assume that $u_0 = u_0^*$ at initialization, which allows us to expand the mismatched trajectory as:

$$u_t = u_t^* + (\tau' - \tau) d_\tau u_t + O((\tau - \tau')^2). \quad (67)$$

We can perform the same expansion for the other terms appearing in Eq. 65:

$$f_\theta(u_t, t) = f_\theta(u_t^*, t) + (\tau' - \tau) \frac{\partial f_\theta}{\partial u}(u_t^*, t) \frac{du_t}{d\tau} + O((\tau' - \tau)^2), \quad (68)$$

and

$$\frac{d}{dt} [f_\theta(u_t, t)] = \frac{d}{dt} [f_\theta(u_t^*, t)] + (\tau' - \tau) \frac{\partial}{\partial u} \left[\frac{d}{dt} [f_\theta(u_t^*, t)] \right] \frac{du_t}{d\tau} + O((\tau' - \tau)^2). \quad (69)$$

Then, by plugging Eq. 67, 68, and 69 into Eq. 65, and keeping only the terms in $(\tau' - \tau)$ we obtain the following differential equation on $d_\tau u_t$:

$$\tau \frac{d \dot{u}_t}{d\tau} = \left(-\text{Id} + \frac{\partial}{\partial u} \left[f_\theta(u_t^*, t) + \tau \frac{d}{dt} [f_\theta(u_t^*, t)] \right] \right) \frac{du_t}{d\tau} + \frac{d}{dt} [f_\theta(u_t^*, t)]. \quad (70)$$

We now proceed to compute the terms appearing in (70) in order to solve it. First, by definition of equilibrium we have

$$\frac{d}{dt} [f_\theta(u_t^*, t)] = \dot{u}_t^*. \quad (71)$$

However, replacing $f_\theta(u_t^*, t)$ by u_t^* cannot be used when we compute partial derivatives with respect to u . This is because both functions are equal uniformly across time, while having different variations with respect to u . To keep our analysis simple, we now turn to the case $f_\theta(u_t, t) = W\rho(u_t) + W_{\text{in}}x_t$. We have that the Jacobian at the equilibrium trajectory is equal to

$$\frac{\partial f_\theta}{\partial u}(u_t^*, t) = W \odot \text{diag}(\rho'(u_t^*)) := W'(t). \quad (72)$$

Finally, the Jacobian of the time derivative of $f_\theta(u_t, t)$ is

$$\tau \frac{\partial}{\partial u} \left[\frac{d}{dt} [f_\theta(u_t^*, t)] \right] \quad (73)$$

$$= \tau \frac{\partial}{\partial u} [(W_{\text{in}} \dot{x}_t + W \rho'(u_t^*) \odot \dot{u}_t^*)] \quad (74)$$

$$= \tau W \frac{\partial}{\partial u} [\rho'(u_t^*) \odot \dot{u}_t^*] \quad (75)$$

$$= \tau W \text{diag}(\rho''(u_t^*) \odot \dot{u}_t^*) + \tau W'(t) \frac{\partial \dot{u}_t^*}{\partial u} \quad (76)$$

$$= W \text{diag} \left(\rho''(u_t^*) \odot \left(-u_t^* + W_{\text{in}} x_t + W \rho(u_t^*) + \tau \frac{d}{dt} [f_\theta(u_t^*, t)] \right) \right) \quad (77)$$

$$+ W'(t) \frac{\partial}{\partial u} \left[-u_t^* + W_{\text{in}} x_t + W \rho(u_t^*) + \tau \frac{d}{dt} [f_\theta(u_t^*, t)] \right]. \quad (78)$$

Here we used the following matrix calculus identity

$$\frac{\partial}{\partial x} [f(x) \odot g(x)] = \text{diag}(g(x)) \cdot \frac{\partial f}{\partial x} + \text{diag}(f(x)) \cdot \frac{\partial g}{\partial x}. \quad (79)$$

We call $W''(t)$ the term of (77). Then, we continue with

$$\tau \frac{\partial}{\partial u} \left[\frac{d}{dt} [f_\theta(u_t^*, t)] \right] = W''(t) \quad (80)$$

$$+ W'(t) \frac{\partial}{\partial u} \left[-u_t^* + W_{\text{in}} x_t + W \rho(u_t^*) + \tau \frac{d}{dt} [f_\theta(u_t^*, t)] \right] \quad (81)$$

$$= W''(t) - W'(t) + W'(t)^2 + \tau W'(t) \frac{\partial}{\partial u} \left[\frac{d}{dt} [f_\theta(u_t^*, t)] \right] \quad (82)$$

so that

$$\tau (\text{Id} - W'(t)) \frac{\partial}{\partial u} \left[\frac{d}{dt} [f_\theta(u_t^*, t)] \right] = W''(t) - (\text{Id} - W'(t)) W'(t) \quad (83)$$

and

$$\tau \frac{\partial}{\partial u} \left[\frac{d}{dt} [f_\theta(u_t^*, t)] \right] = (\text{Id} - W'(t))^{-1} W''(t) - W'(t). \quad (84)$$

Here, we assume that the activation function ρ has a negligible second order derivative $\rho'' \approx 0$. This approximation is actually an equality in cases where the activation function is piece-wise linear, such as the hard sigmoid or the ReLU. This assumption makes $W''(t) = 0$ for all t and we simply have:

$$\tau \frac{\partial}{\partial u} \left[\frac{d}{dt} [f_\theta(u_t^*, t)] \right] = -W'(t). \quad (85)$$

We are now in a position to replace all the terms in (70):

$$\tau \frac{d\dot{u}_t}{d\tau} = (-\text{Id} + W'(t) - W'(t)) \frac{du_t}{d\tau} + \dot{u}_t^* \quad (86)$$

$$\tau \frac{d\dot{u}_t}{d\tau} = -\frac{du_t}{d\tau} + \dot{u}_t^*. \quad (87)$$

Solving this differential equation, and after a transitory period ($t \gg 0$), the mismatch is given by:

$$\frac{du_t}{d\tau} = \frac{1}{\tau} \int_0^t \exp\left(-\frac{t-t'}{\tau}\right) \dot{u}_t^* dt'. \quad (88)$$

Assuming $\gamma := \max_t \|\dot{u}_t^*\|$, we thus have

$$\left\| \frac{du_t}{d\tau} \right\| \leq \gamma, \quad (89)$$

that is, using Equation 67,

$$\|u_t - u_t^*\| \leq |\tau' - \tau| \gamma + O((\tau - \tau')^2). \quad (90)$$

We emphasize that this is a local result around $\tau' = \tau$ and that the picture may be qualitatively different further away from τ . Let us consider the following example:

$$\tau \dot{u} = -u + wu + w\tau' \dot{u}, \quad (91)$$

which is obtained by setting $f(u) = wu$, so that

$$(\tau - w\tau') \dot{u} = -(1 - w)u \quad (92)$$

For $\tau' = \tau$, u converges to 0 as long as $w < 1$. However, whenever $\tau' > \tau/w$, the dynamics becomes unstable and u exponentially diverges. This highlights that the bound derived above no longer holds in this region. Finally, we note that the case $\tau' = \tau/w$ is degenerate, as only a constant state equal to 0 satisfies the differential equation (assuming $w \neq 1$). Our analysis does not consider this case.

C Additional details on holomorphic equilibrium propagation

Building on holomorphic equilibrium propagation [Laborieux and Zenke, 2022, 2024], we demonstrate how prospective dynamics enable neurons to track oscillating feedback signals. In this framework, error terms δ are computed by introducing finite-size oscillating feedback β_t in the complex plane. The key challenge is that oscillations must be slow enough for neurons to equilibrate, introducing a timescale constraint. We now

show that prospective dynamics circumvent this limitation, allowing effective tracking regardless of the neurons’ intrinsic timescales or the oscillation frequency.

We consider the following neural dynamics:

$$\tau_u \dot{u}_t = -u_t + W\rho(u_t) + \beta_t(u_t^{\text{out}} - y_t), \quad \text{with } u_t^{\text{in}} = x_t. \quad (93)$$

Here, u and β are complex-valued vectors that evolve over time. Compared to the one we used as an example in Section 1, the dynamics contains two additional elements: First, the weight matrix includes both feedforward connections that are responsible for processing the input and backward connections that send teaching signals from the output to hidden neurons. Second, the dynamics includes a nudging force $\beta_t(u_t^{\text{out}} - y_t)$ that pushes output neurons towards the target y_t . The error gradient is multiplied by the oscillating nudging strength $\beta_t = |\beta| \exp(\frac{2i\pi t}{\tau_\beta})$. Provided that u_t tracks the equilibrium points u_t^* imposed by the feedback β_t , and that β -oscillations are much faster than x and y , the error is encoded in the first mode of the oscillations:

$$\delta_t = \frac{1}{\tau_\beta |\beta|} \int_{t-\tau_\beta}^t u_{t'} e^{-2i\pi t'/\tau_\beta} dt'. \quad (94)$$

The weight update is the same as Eq. 13, where the pre-synaptic term $\rho(u)_i$ is measured as the average of the oscillations induced by β_t . [Laborieux and Zenke, 2024] show that this update closely approximates the gradient. Whether such complex neurons can be implemented in the brain remains an open question. Alternatively, Anisetti et al. [2024] have shown that oscillations can exactly compute the gradient in the weak nudging limit. Finally, we argue that the separation of timescales needed here is less restrictive than the one assuming infinitely fast neurons; here, only the teaching signal, a small part of the network, has to be fast.

D Additional simulation details for the inverted pendulum

We use `gymnax` [Lange, 2022] to simulate the inverted pendulum system through Euler integration ($\Delta t = 1\text{ms}$). At each timestep, the agent receives the state $s_t \in \mathbb{R}^4$ consists of position, velocity, angle, angular velocity ($s_t = [x_t, \dot{x}_t, \theta_t, \dot{\theta}_t]$). An episode terminates when the state vector is outside of the range $-2.4 \leq x \leq 2.4, -12^\circ \leq \theta \leq 12^\circ$. Typically, for the task, the agent receives a reward of 1 at every time step within the admissible region. Due to our online near-continuous learning regime, we reformulate this as a reward rate

$$r(s(t)) = \begin{cases} 1, & \text{if } -2 \leq x \leq 2, -8.5^\circ \leq \theta \leq 8.5^\circ \\ 0, & \text{otherwise} \end{cases} \quad (95)$$

Note that prospective neurons take time to react to discontinuity. As a result, it is impossible to learn the right policy solely due to the lack of reward at the last timestep before the end of the episode. Therefore, we use a reward rate that restricts the agent to a smaller rewarded region than the environment’s termination condition, allowing for more time steps spent with a reward of 0.

To train the agent, we adopt the continuous RL formalization, where the value function at time t is defined by

$$V(s(t)) = \int_t^\infty \exp\left(-\frac{t' - t}{\tau_v}\right) r(s(t')) dt' \quad (96)$$

τ_v controls the discounting of future rewards.

$$\dot{V}(t) = \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \left[\int_{t+\Delta t}^\infty \exp\left(-\frac{t' - t - \Delta t}{\tau_v}\right) r(t') dt' \right. \quad (97)$$

$$\left. - \int_t^\infty \exp\left(-\frac{t' - t}{\tau_v}\right) r(t') dt' \right] \quad (98)$$

$$= \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \left[- \int_t^{t+\Delta t} \exp\left(-\frac{t' - t}{\tau_v}\right) r(t') dt' \right. \quad (99)$$

$$\left. + \exp\left(\frac{\Delta t}{\tau_v}\right) \int_{t+\Delta t}^\infty \exp\left(-\frac{t' - t}{\tau_v}\right) r(t') dt' \right] \quad (100)$$

$$= \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \left(- \int_t^{t+\Delta t} \exp\left(-\frac{t' - t}{\tau_v}\right) r(t') dt' + \exp\left(\frac{\Delta t}{\tau_v}\right) V(t) \right) \quad (101)$$

$$= -r(t) + \frac{1}{\tau_v} V(t) \quad (102)$$

Therefore, the continuous TD error is

$$\delta(t) = r(t) - \frac{1}{\tau_v} V(t) + \dot{V}(t) \quad (103)$$

To train the critic, we want to minimize

$$E(t) = \frac{1}{2} |\delta(t)|^2$$

Discretizing Eq 103, we estimate \dot{V} using finite differences

$$\delta(t) = r(t) - \frac{1}{\tau} V(t) + \frac{V(t) - V(t - \Delta t)}{\Delta t} \quad (104)$$

$$= r(t) + \frac{1}{\Delta t} \left(\left(1 - \frac{\Delta t}{\tau_v}\right) V(t) - V(t - \Delta t) \right) \quad (105)$$

$$\hat{\delta}_t = \delta(t)\Delta t = r(t)\Delta t + \left(1 - \frac{\Delta t}{\tau_v}\right) V(t) - V(t - \Delta t) \quad (106)$$

We thus obtain the appropriate scaling for the reward per timestep $\hat{r}_t = r(t)\Delta t$ and the discount rate $\gamma = 1 - \frac{\Delta t}{\tau_v}$. We train the agent with the prospective version of the online A2C as detailed in Algorithm 1. Conceptually, we denote the activations (including the network outputs) of the policy and value networks as column vectors $u_{p,t} \in \mathbb{R}^{66}$ (64 hidden neurons, 2 action neurons), $u_{v,t} \in \mathbb{R}^{65}$ (64 hidden neurons, 1 value neuron) respectively. The feedforward inputs at each step are then given by

$$f_{p/v}(s_t, u_{p/v,t}) = W_{p/v} \left(\text{ReLU}^{s_t}(u_{p/v,t}) \right) + b_{p/v} \quad (107)$$

where W are lower triangular block matrices parameterized appropriately to carry out feedforward computation.

The loss functions are

$$\hat{\delta}_t = r(t)\Delta t + \left(1 - \frac{\Delta t}{\tau_v}\right) u_{v,t}^{\text{out}} - u_{v,t-1}^{\text{out}} \quad (108)$$

$$\mathcal{L}_{\text{actor}}(u_{p,t}, a) = -\hat{\delta}_t \log \left([\text{Softmax}(u_{p,t}^{\text{out}})]_a \right) \quad (109)$$

$$\mathcal{L}_{\text{critic}}(u_{v,t}, r) = \hat{\delta}_t^2 \quad (110)$$

The optimizer we use is SGD with early stopping (`critic_lr`= $1e^{-6}$, `actor_lr`= $1e^{-4}$, $\tau_v = 1$, $\tau_u = \tau_\delta = 1$) with a batch size of 1. We train the agent for 10000 episodes, and each episode has a maximum time of $T = 4\text{s}$ (4000 steps).

Algorithm 1 Online advantage actor critic (A2C) with prospective neurons

```

1: Initialize policy and value networks  $f_{p,v}$  with parameters  $\theta_0, \phi_0$ 
2: Initialize environment with appropriate  $\Delta t$ 
3: for each training episode do
4:   Initialize state variables  $u_{p/v,0}$ 
5:   Initialize gradient variables  $\delta_{p/v,0}$ 
6:   Reset env and observe the initial state  $s_0$ 
7:   while not done and  $t < \text{MAX\_LENGTH}$  do
8:      $a_t \sim u_{p,t}^{\text{out}}$   $\triangleright$  Sample action
9:      $r_t, s_{t+1} \leftarrow \text{env}(s_t, a_t)$   $\triangleright$  Step environment

10:    Estimate TD error with prospective dynamics of  $u_v$ :
11:    
$$u_{v,t+1} \leftarrow u_{v,t} + \frac{\Delta t}{\tau} (-u_{v,t} + f_v(s_{t+1}, u_{v,t}))$$

 $+ (f_v(s_{t+1}, u_{v,t}) - f_v(s_t, u_{v,t-1}))$ 
12:    
$$A_t \leftarrow r_t + \left(1 - \frac{\Delta t}{\tau_v}\right) u_{v,t+1}^{\text{out}} - u_{v,t}^{\text{out}}$$


13:    Estimate prospective dynamics for  $u_p$ :
14:    
$$u_{p,t+1} \leftarrow u_{p,t} + \frac{\Delta t}{\tau} (-u_{p,t} + f_p(s_{t+1}, u_{p,t}))$$

 $+ (f_p(s_{t+1}, u_{p,t}) - f_p(s_t, u_{p,t-1}))$ 

15:    Backpropagated error dynamics:
16:    
$$\delta_{p,t+1} \leftarrow \delta_{p,t} + \frac{\Delta t}{\tau} (-\delta_{p,t} + \nabla_u \mathcal{L}_{\text{actor}}(u_{p,t}, a_t, A_t))$$

 $+ (\nabla_u \mathcal{L}_{\text{actor}}(u_{p,t}, a_t, A_t) - \nabla_u \mathcal{L}_{\text{actor}}(u_{p,t-1}, a_{t-1}, A_{t-1}))$ 
17:    
$$\delta_{v,t+1} \leftarrow \delta_{v,t} + \frac{\Delta t}{\tau} (-\delta_{v,t} + \nabla_u \mathcal{L}_{\text{critic}}(u_{v,t}, r_t))$$

 $+ (\nabla_u \mathcal{L}_{\text{critic}}(u_{v,t}, r_t) - \nabla_u \mathcal{L}_{\text{critic}}(u_{v,t-1}, r_{t-1}))$ 

18:    Update parameters:
19:    
$$\theta_{t+1} \leftarrow \theta_t - \frac{\Delta t}{\tau_\theta} (\partial_\theta f_p^\top \delta_{p,t+1})$$

20:    
$$\phi_{t+1} \leftarrow \phi_t - \frac{\Delta t}{\tau_\phi} (\partial_\phi f_v^\top \delta_{v,t+1})$$

21:  end while
22: end for

```
