



Feasibility of dopamine as a vector-valued feedback signal in the basal ganglia

Emil Wärnberg^{a,b} and Arvind Kumar^{b,1}

Edited by Terrence Sejnowski, Salk Institute for Biological Studies, San Diego, CA; received December 29, 2022; accepted June 8, 2023

It is well established that midbrain dopaminergic neurons support reinforcement learning (RL) in the basal ganglia by transmitting a reward prediction error (RPE) to the striatum. In particular, different computational models and experiments have shown that a striatum-wide RPE signal can support RL over a small discrete set of actions (e.g., no/no-go, choose left/right). However, there is accumulating evidence that the basal ganglia functions not as a selector between predefined actions but rather as a dynamical system with graded, continuous outputs. To reconcile this view with RL, there is a need to explain how dopamine could support learning of continuous outputs, rather than discrete action values. Inspired by the recent observations that besides RPE, the firing rates of midbrain dopaminergic neurons correlate with motor and cognitive variables, we propose a model in which dopamine signal in the striatum carries a vector-valued error feedback signal (a loss gradient) instead of a homogeneous scalar error (a loss). We implement a local, “three-factor” corticostriatal plasticity rule involving the presynaptic firing rate, a postsynaptic factor, and the unique dopamine concentration perceived by each striatal neuron. With this learning rule, we show that such a vector-valued feedback signal results in an increased capacity to learn a multidimensional series of real-valued outputs. Crucially, we demonstrate that this plasticity rule does not require precise nigrostriatal synapses but remains compatible with experimental observations of random placement of varicosities and diffuse volume transmission of dopamine.

basal ganglia | feedback alignment | learning | dopamine

The basal ganglia are thought to be the main locus of reinforcement learning (RL) in the brain (1). In particular, dopamine-modulated long-term plasticity in the corticostriatal synapses is crucial for learning and fine-tuning skilled movements based on environmental feedback (2). Combined with the striking observation that midbrain dopaminergic cells transmit a reward prediction error (RPE) to the striatum (3, 4), this has inspired a plethora of computational models of the basal ganglia implementing various forms of RL. Notably, however, virtually all these models assume the set of actions that can be selected—the action space in RL terminology—is small and discrete (see e.g., refs. 5–14). Practically, this means that each action can be exclusively represented by a disjoint group of striatal neurons, sometimes called action channels (15). At their core, in each of these models, there is some sort of competition between the action channels so that the selected action (or likely selected in probabilistic models) corresponds to the channel with the highest activity. This is consistent with a global RPE transmitted by dopamine that reinforces or depresses the corticostriatal synapses of the active channel.

However, there is now accumulating evidence that the action space of the basal ganglia is not small and discrete, but rather multidimensional and continuous (16–20). For a multidimensional output, a global RPE is not as effective at driving learning as it is for discrete action channels. For example, a mouse learning to reach for a food pellet may need to learn to control its paw in the x , y , and z directions. Intuitively, a single error, perhaps proportional to the final distance to the target, would be less efficient than having a three-dimensional error signal representing the error in the three directions.

More formally, producing a continuous and multidimensional output requires the basal ganglia to learn a function approximation rather than tabular values (21). Although simple function approximators (e.g., single-layer networks) can be successfully trained with a scalar global error, that strategy rapidly becomes untenable with increasing network depth and complexity (22). In practice, contemporary RL algorithms for continuous action spaces, e.g., A3C (23), DDPG (24) and PPO (25) rely on artificial neural networks trained with backpropagation to approximate the continuous policy.

Therefore, we asked whether it would be possible for dopamine to support function approximation learning in the basal ganglia by carrying a vector-valued feedback signal

Significance

The striatum is thought to learn to select actions based on environmental feedback and rewards using a dopamine feedback signal. When the action is continuous and multidimensional, e.g., a reaching movement, it is beneficial that the dopamine feedback has the same dimensionality as the task. However, although the dopaminergic cell bodies exhibit multidimensional responses, it has been unclear whether they can be effectively used by the striatum given their broad and unspecific axonal arbors. We present a simplified simulation model where multidimensional dopamine activity improves learning compared to a single signal, despite the nonspecificity of the projections. Thus, we demonstrate the feasibility of multidimensional feedback using dopamine in the basal ganglia and make testable predictions.

Author affiliations: ^aDepartment of Neuroscience, Karolinska Institutet, 171 77 Stockholm, Sweden; and ^bDivision of Computational Science and Technology, School of Electrical Engineering and Computer Science, KTH Royal Institute of Technology, 114 28 Stockholm, Sweden

Author contributions: E.W. designed research; E.W. performed research; E.W. contributed new reagents/analytic tools; E.W. analyzed data; A.K. supervised the research; and E.W. and A.K. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

Copyright © 2023 the Author(s). Published by PNAS. This open access article is distributed under Creative Commons Attribution License 4.0 (CC BY).

¹To whom correspondence may be addressed. Email: arvkumar@kth.se.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2221994120/-DCSupplemental>.

Published August 1, 2023.

from the midbrain back to the striatum. Such a feedback signal would manifest in the VTA and SNC in terms of cell tunings to various task-related variables, consistent with recent observations (26–28). However, one apparent problem with dopamine transmitting a vector-valued error is that dopaminergic axons do not precisely target specific neurons in the striatum, but instead release dopamine from a large number of varicosities that can then diffuse over a short distance through extracellular space (29), thereby mixing any individual error components. In principle, one could imagine this problem being solved by representing each action dimension in a spatially compact region that receives its private dopamine channel (see refs. 30 and 31 for similar ideas). However, experimental evidence suggests that individual dopaminergic neurons (32) have mixed tuning rather than responding to a single task variable. Moreover, although there is a coarse-grain somatotopic organization of the striatum (33) as well as the substantia nigra (34), the axonal arborizations of individual SNC neurons are huge and cover large portions of the striatum (up to 5%; 35). Therefore, a large set of isolated parallel channels without cross-talk appears unlikely.

In this work, rather than separating the entire basal ganglia into fine-grained parallel channels, we propose that the mixing of multiple error components can be undone downstream from the striatum. In particular, we propose that if the striatofugal projections were to be subjected to similar systematic long-term plasticity as the corticostriatal projections, then we can make use of feedback alignment (36) to have striatum learn continuous outputs efficiently. We show that by using a stylized model of diffuse mixing of dopaminergic feedback as the random feedback,

the Random Feedback Local Online (RFLO) learning rule (37) can be employed in a recurrent neural network model of the basal ganglia. This results in significantly improved learning compared to a model with homogeneous/scalar dopamine in the entire striatum. Thus, we connect two seemingly unrelated observations: heterogeneous dopamine response and the involvement of the striatum in learning complex and multidimensional continuous actions.

Network Model

As a model of the basal ganglia learning a skilled movement (such as an animal reaching for a food pellet or pressing a lever), we constructed a task wherein a recurrent neural network must learn to repeatedly output a given trajectory in d -dimensional space. The output trajectory is defined as the activity of d readout neurons. In our idealized model of a small piece of basal ganglia, we take the readout population to be either the internal globus pallidus (GPi) or the substantia nigra pars reticulata (SNr) (see ref. 17, for experimental support).

This striatum projects to the readout population (GPi/SNr) and receives excitatory inputs from two input populations: cortex and thalamus (Fig. 1A). The task of the network is to adjust the input and recurrent synaptic weights in the striatum so that the readout matches the desired d -dimensional target $T(t)$ as closely as possible (Fig. 1I).

In the cortex and striatum, we model the subthreshold membrane potential $V(t)$ of each neuron (or small group of neurons) as

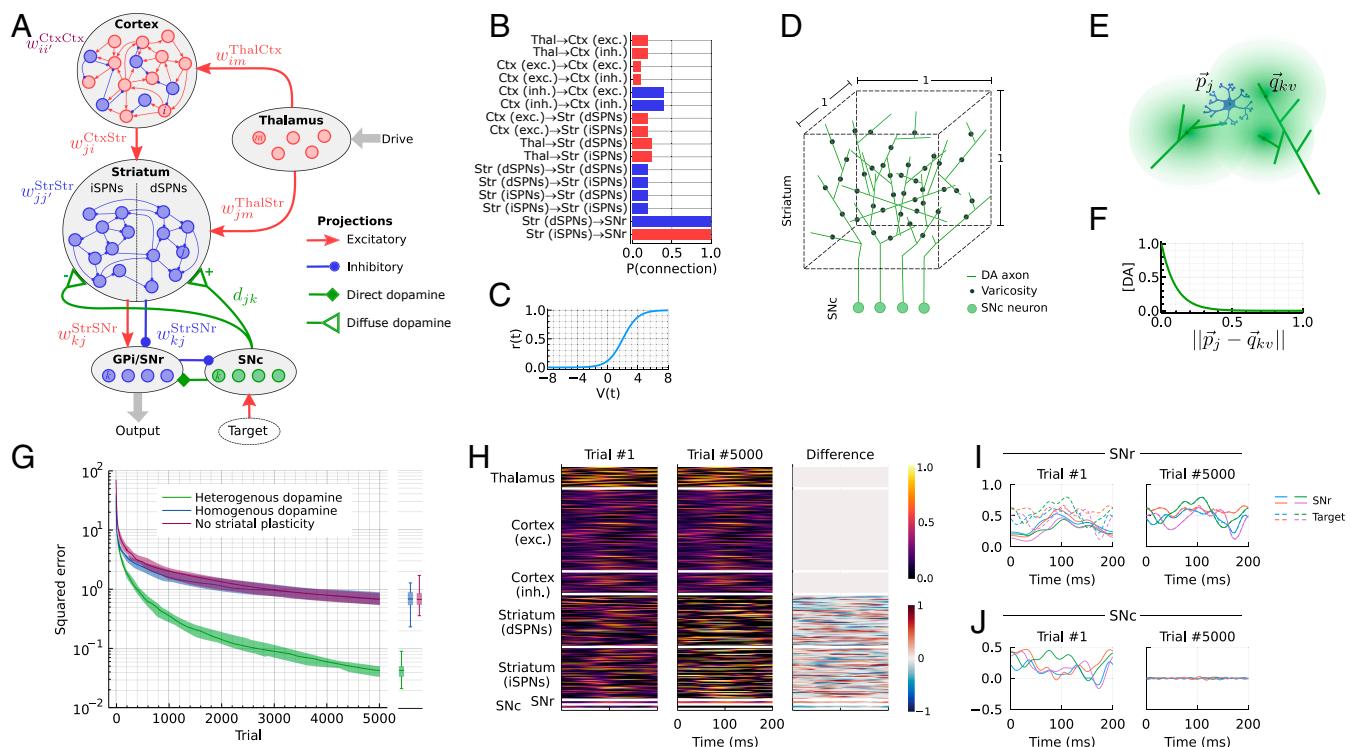


Fig. 1. Heterogeneous dopamine improves learning in a RNN. (A) Network architecture. (B) Connection probability of a synapse for any pair of neurons in the network model. (C) Transfer function of each unit. (D) Illustration of placement of varicosities in a unit cube. (E) Illustration of striatal projection neurons (SPN) (located at $\vec{p}_j = [p_x^j, p_y^j, p_z^j]$) receiving a mix of dopamine from three close-by varicosities. (F) Relative dopamine concentration as a function of distance from the varicosity. (G) Convergence of the loss (L2) over trials for the model (heterogeneous dopamine) as well as two control models (homogeneous dopamine and no striatal plasticity). Each trial is one presentation of the pattern. Solid lines indicate the median of 25 runs; shaded areas show the first and third quartiles. Boxplots show median, quartiles, and min/max of the same 25 runs at the final (5,000th) training trial. (H) Example activity in the network in the first and last trials of training. (I) Comparison of the 4-dimensional readout (i.e., SNr/GPi activity) with the target in the first and last trial of training. (J) The difference between the readout and the target (i.e., SNC activity) in the first and last trial of training.

$$\frac{dV(t)}{dt} = -\frac{1}{\tau_m} V(t) + \sum_{\text{Pre}} \sum_{\text{pre}} w_{\text{post},\text{pre}}^{\text{PrePost}} r_{\text{pre}}^{\text{Pre}}(t), \quad [1]$$

where $\tau_m = 10\text{ms}$ is the membrane time constant. We write the synaptic weights as $w_{\text{post},\text{pre}}^{\text{PrePost}}$ where $\text{Pre} \in \{\text{Thal, Ctx, Str}\}$ is the presynaptic population and $\text{pre} \in \{m, i, j\}$ is the index of the neuron in that population*. The firing rate $r(t)$ of each neuron is calculated as:

$$r(t) = \phi(V(t)) = \frac{1}{1 + e^{-V(t)+b}}, \quad [2]$$

where the term $b = 2$ shifts the sigmoid to the right so that the firing rates are sparser when the inputs are balanced and the membrane potentials fluctuate around 0 (Fig. 1C). In the readout population (GPi/SNr), we index the neurons by k and let their firing rates be given by

$$r_k^{\text{SNr}}(t) = \phi \left(\sum_{j=1}^{N^{\text{Str}}} w_{kj}^{\text{StrSNr}} r_j^{\text{Str}}(t) \right). \quad [3]$$

The purpose of our model is to demonstrate that mixing of dopamine is not detrimental to vector-valued feedback and not to capture every detail of the basal ganglia. Nevertheless, to make sure the learning setup is fair, we added a number of biological constraints to the model. First, all connections except the readout are sparse, i.e., only a fraction of pairs of neurons are allowed to connect (Fig. 1B). Second, we required the signs of the weights to match the sign of the projection (excitatory or inhibitory) throughout learning (Fig. 1A and B). For simplicity, we omitted the external globus pallidus (GPe) and subthalamic nucleus (STN) and modeled the indirect pathway as a direct excitatory projection from the striatum to GPi/SNr. The sign of dopamine-driven plasticity was reversed for the striatal projection neurons in the indirect pathway (iSPNs; Fig. 1A). Because of our focus on dopamine-dependent learning in the striatum, we only include the cortex as a reservoir of rich but task-aligned dynamics and do not consider any learning that might take place in the cortex itself. However, we include dopamine-dependent plasticity in all basal ganglia synapses: corticostriatal, thalamostriatal, striatostriatal, and striatofugal.

Derivation of the Synaptic Plasticity Rule. To construct normatively appropriate learning rules for the plastic synapses, we note the task is to minimize the loss

$$\ell(t) = \frac{1}{2} \sum_{k=1}^d \left(r_k^{\text{SNr}}(t) - T_k(t) \right)^2. \quad [4]$$

For the error to decrease over time, we would like the plasticity rule to change the weight of each corticostriatal synapse (w_{ji}^{CtxStr}) such that

$$\frac{dw_{ji}^{\text{CtxStr}}(t)}{dt} \propto -\frac{\partial \ell(t)}{\partial w_{ji}^{\text{CtxStr}}(t)}. \quad [5]$$

In *SI Appendix, Text S1*, we show that by expanding this partial derivative with two simplifications, considering ℓ only at t (i.e., not backward or forward in time) and treating the firing rates of

other striatal projection neurons (SPNs) as fixed as in ref. 37, we arrive at the following plasticity rule:

$$\frac{dw_{ji}^{\text{CtxStr}}(t)}{dt} = -\alpha \gamma_j(t) p_{ji}(t), \quad [6]$$

$$\tau^{\text{Str}} \frac{dp_{ji}}{dt} = -p_{ji}(t) + r_j^{\text{Str}}(t) \left(1 - r_j^{\text{Str}}(t) \right) r_i^{\text{Ctx}}(t), \quad [7]$$

$$\gamma_j(t) = \sum_{k=1}^d \epsilon_k(t) w_{kj}^{\text{StrSNr}}(t), \quad [8]$$

$$\epsilon_k(t) = \left(r_k^{\text{SNr}}(t) - T_k(t) \right) r_k^{\text{SNr}}(t) \left(1 - r_k^{\text{SNr}}(t) \right). \quad [9]$$

The plasticity rules for the thalamostriatal and striatostriatal synaptic weights are fully analogous. We interpret this plasticity rule in biological terms as follows. From Eq. 6, we see that the weight update depends on a neuron-specific factor γ_j and a synapse-specific factor p_{ji} . The latter is a low-pass filtered trace of a Hebbian-like product between pre- and postsynaptic firing. This could be identified as an eligibility trace (9, 38), and we note it could be represented, for example, by the local concentration of calcium in the spine.

The eligibility trace p_{ji} is multiplied by a “third factor” γ_j . Experimental results suggest plasticity in corticostriatal synapses depends on three factors: presynaptic activity, postsynaptic activity, and dopamine (39, 40). Given that the two former are captured by p_{ji} , we would like to associate γ_j in Eq. 6 with dopamine. If we assume the number of dopaminergic neurons to be the same as the number of read-out neurons,

$$N^{\text{SNC}} = N^{\text{SNr}} = d, \quad [10]$$

we can assign ϵ_k to the firing rate r_k^{SNC} :

$$r_k^{\text{SNC}}(t) = \epsilon_k(t). \quad [11]$$

That is, we assume that the SNC has access to the vector-valued error (*Discussion*). Note that we can handle negative values of ϵ_k by loosely interpreting r_k^{SNC} as the deviation from some baseline firing rate. This leaves just one problem: The coefficients used to sum the contribution of the dopaminergic cells in Eq. 8 should be the downstream striatofugal weights w_{kj}^{StrSNr} , which are not available to the corticostriatal synapses. However, following a similar derivation as above, Murray (37) showed that, if the readout weights w_{kj}^{StrSNr} themselves are plastic, we can replace w_{kj}^{StrSNr} in Eq. 8 with a random value at only a minor cost to the convergence of the loss, thanks to a phenomenon called feedback alignment (36). Therefore, we next asked whether the dopaminergic nigrostriatal projection could form such a random feedback matrix.

Dopamine Diffusion. To investigate whether dopaminergic feedback could be used to communicate the third factor needed for the corticostriatal plasticity, we set up a stylized model of dopamine diffusion. We assumed each striatal neuron had a position $\vec{p}_j = [p_x^j, p_y^j, p_z^j]$ where $p_x^j, p_y^j, p_z^j \in [0, 1]$. Second, we assumed that each SNC neuron sent axonal projections that covered the entire cube and that axonal arbor of each SNC neuron has $N^{\text{var}} = 10$ varicosities randomly placed in the unit cube (Fig. 1D). Third, we assumed the dopamine released from each varicosity is proportional to the firing rate at the soma in the SNC

*We use the order post, pre for the subscript to follow the matrix-friendly convention where the first letter indexes the row and the second the column.

and that the dopamine concentration decreases exponentially with distance from the varicosity. This gives the dopamine concentration $C_j(t)$ at striatal neuron j as

$$C_j(t) = \sum_{k=1}^{N^{\text{SNC}}} \sum_{v=1}^{N^{\text{Var}}} r_k^{\text{SNC}}(t) \frac{1}{\lambda} e^{-\frac{\|\vec{p}_j - \vec{q}_{kv}\|}{\lambda}}, \quad [12]$$

where $\vec{q}_{kv} = [q_x^{kv}, q_y^{kv}, q_z^{kv}]$ is the position of the v th varicosity of SNC neuron k . λ controls the rate of decay with distance and was set to 0.1 (so that dopamine concentration decreases to $1/e \approx 37\%$ after diffusing a distance equivalent to 10% of the side of the cube). This model defines an effective nigrostriatal weight

$$d_{jk} = \sum_{v=1}^{N^{\text{Var}}} \frac{1}{\lambda} e^{-\frac{\|\vec{p}_j - \vec{q}_{kv}\|}{\lambda}}, \quad [13]$$

that does not vary with time, so we can write

$$C_j(t) = \sum_{k=1}^{N^{\text{SNC}}} r_k^{\text{SNC}}(t) d_{jk}. \quad [14]$$

Remember we assumed the vector-valued error $\epsilon_k(t)$ is present in the SNC (Eq. 11), and note the similarity between Eqs. 8 and 14. If we introduce plasticity in the striatofugal projection, feedback alignment will cause $w_{kj}^{\text{StrSNr}}(t) \rightarrow d_{jk}$. Therefore, we again start with

$$\frac{dw_{kj}^{\text{StrSNr}}(t)}{dt} \propto -\frac{\partial \ell(t)}{\partial w_{kj}^{\text{StrSNr}}(t)}, \quad [15]$$

and arrive at

$$\frac{dw_{kj}^{\text{StrSNr}}(t)}{dt} = -\beta \epsilon_k(t) r_j^{\text{Str}}(t), \quad [16]$$

with $\epsilon_k(t) = r_k^{\text{SNC}}(t)$ as before and $\beta = 10^{-3}$.

In summary, we set the corticostriatal plasticity update to

$$\frac{dw_{ji}^{\text{CtxStr}}(t)}{dt} = -\alpha p_{ji}(t) \sum_{k=1}^{N^{\text{SNC}}} r_k^{\text{SNC}}(t) d_{jk}, \quad [17]$$

where d_{jk} is given by Eq. 13 and $p_{ji}(t)$ is given by Eq. 7. We set $\alpha = -2.5 \cdot 10^{-2}$ for direct pathway neurons and $\alpha = 2.5 \cdot 10^{-2}$ for indirect pathway neurons to capture the different effects of D1 and D2 receptors. The minus sign helps $w_{kj}^{\text{StrSNr}}(t) \rightarrow d_{jk}$ for direct pathway striatal neurons, because for these neurons $w_{kj}^{\text{StrSNr}}(t) < 0$. We used analogous plasticity rules for thalamostriatal and striatostriatal connections.

Learning with Vector-Valued Dopamine Feedback

Having set up the model, we simulated the dynamics for 5,000 presentations of the target output (Fig. 1H). Each target was a 4-dimensional 200-ms time series drawn from a Gaussian process (Fig. 1I; see *Methods*). In the first trial, the output does not match the target (Fig. 1I, Left), but after 5,000 trials, the plasticity rules

have driven the network to produce GPi/SNr output that closely matches with the targets (Fig. 1I, Right). This is achieved both by plasticity in the readout population (GPi/SNr; Eq. 17) and plasticity in the striatum that adapts the SPN firing rates (Eq. 17; Fig. 1H).

We next asked how this learning depends on the nature of the dopamine feedback. It is well known that for recurrent networks with rich dynamics, plasticity in the readout is sufficient to learn complex patterns (41, 42). Therefore, we first compared our model to a reduced model that only had plasticity in the readout (striatofugal) projection according to Eq. 16, i.e., no plasticity in the striatum. We found that learning with dopamine feedback was faster (Fig. 1G). Next, we compared our model to a model in which each SPN at every time point received the same dopamine feedback, i.e., the striatum receives a homogeneous, scalar dopamine signal. Strikingly, this model performed no better than the reduced model that only has learning in the readout layer (Fig. 1G). The increase in learning performance persisted with different sizes of the striatal (*SI Appendix*, Fig. S1A) and the readout (*SI Appendix*, Fig. S1B) populations, as well as for faster and slower timescales of the target (τ_{task} in Eq. 22; *SI Appendix*, Fig. S1C). These observations show that vector-valued dopamine feedback is crucial for the improvement in learning.

The Improvement in Learning Is because of Feedback Alignment. Next, we further explored the conditions during which vector-valued feedback improved learning of the targets. First, we considered two additional alternative models: i) the feedback is random, i.e., we still have vector-valued feedback, but shuffle the coefficients d_{jk} of Eq. 14 and ii) we use the “ideal” feedback $d_{jk} = w_{kj}^{\text{StrSNr}}(t)$. Note that locking the feedback weights to the feed-forward in the second model means the feedback matrix is time-dependent. Performance of both of these models was similar to our dopamine model (Fig. 2A). This suggests that the main criterion for the feedback to be effective is that feedback matrix $D = [d_{jk}]$ is nondegenerate.

With this hypothesis in mind, we tested varying the spatial scale λ of dopamine diffusion (Eq. 13). Note that this spatial scale was measured as a fraction of the side of the cube (Fig. 1D). When $\lambda \ll 1$, almost no dopamine reaches any SPN from the varicosities, and the network reverts to the No feedback control model (Fig. 2B). On the other hand, when $\lambda \gg 1$, the dopamine from each varicosity covers the entire cube so that all SPNs effectively receive the sum of the dopamine released anywhere. This causes the network to revert to the Homogeneous dopamine control model. In between these two extremes, where the dopamine scale is intermediate, there is a sweet spot where each SPN receives dopamine corresponding to a unique random linear projection of the 4-dimensional error (Fig. 2C). In this regime, the benefit of the feedback is the largest (Fig. 2B). Finally, we also varied the number of varicosities N^{Var} in Eq. 13 and found that with a larger number N^{Var} , a smaller spatial scale λ becomes viable (Fig. 2D). This is also consistent with the creation of a nondegenerate D .

To illustrate the importance of striatofugal plasticity for learning, we simulated the network model with “lesioned” basal ganglia projections by either removing the plasticity (static) or clamping them to 0 (absent). As expected, when the plasticity of the striatofugal projection was turned off, no feedback alignment could take place and the striatal plasticity could not contribute to learning the targets (Fig. 2E, Str→SNr). Turning off plasticity of the striatal projections (corticostriatal, thalamostriatal, and striatostriatal) on the other hand has a more moderate impact.

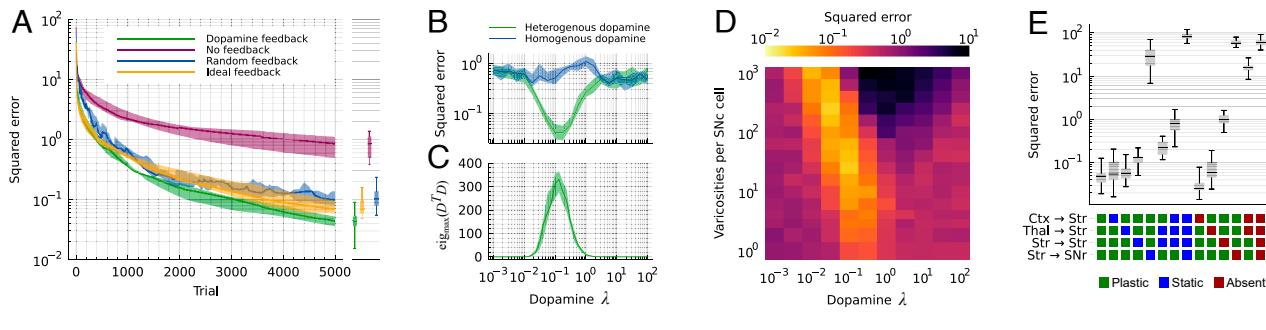


Fig. 2. Learning improvement depends on feedback alignment. (A) Convergence of the loss (L2) for different types of feedback: dopamine feedback as described by our model, no feedback at all (effectively making all synapses except the striatofugal static), random feedback where the elements of the feedback matrix D are shuffled randomly and ideal feedback where $D = W_{Str, SNr}^T$. (B) Squared error after 5,000 trials as a function of the spatial constant of dopamine diffusion (λ). λ was measured relative to the side of the cube in Fig. 1D. (C) Linear independence of the rows of D as a function of the dopamine spatial constant λ . (D) Squared loss after 5,000 trials as a joint function of the dopamine spatial constant λ and the number of varicosities per SNC cell. (E) Squared error after 5,000 trials when some of the four plastic projections are impaired: either set to static, i.e., no plasticity, or removed from the model altogether (absent). See SI Appendix, Fig. S1D for the complete set of manipulations. (All panels) errors are reported as the median over 25 trials. Shaded areas indicate the first and third quartiles of the 25 trials. Boxplots show median, first, and third quartiles; whiskers indicate min and max.

This is because even with all of them fixed, we can still have echo-state-like learning in the striatofugal weights (see the No feedback null model in Figs. 1G and 2A). Similarly as when fixing the weights, removing either the cortical or the thalamic projection does not change the eventual error much, as both projections play similar and mostly interchangeable roles in our model, whereas removing both silences the striatum completely and hence gives a very large error (Fig. 2E). See *SI Appendix*, Fig. S1D for a systematic investigation of how removing or blocking plasticity on different projections affects learning.

Fast Synaptic Dynamics Can Compensate for Slow Dopamine. So far, we have assumed that dopamine is diffuse in space but delivered instantly to the receiving SPNs. While this allows the synaptic weights to be updated correctly on every time step, it neglects the temporal dynamics of dopamine release, diffusion, reuptake, etc. A faithful quantitative model of these processes is beyond the scope of our abstract rate network, but it is nevertheless important to determine how dependent our dopamine-based learning rule is on the assumption of instantaneous dopamine release. We simplified all temporal dynamics of dopamine into a simple exponential low-pass filter with time constant τ_{DA} . That is, we changed the equation for dopamine concentration at striatal neuron j (Eq. 12) to

$$C'_j(t) = \int_0^t e^{\frac{t'-t}{\tau_{DA}}} C_j(t') dt'. \quad [18]$$

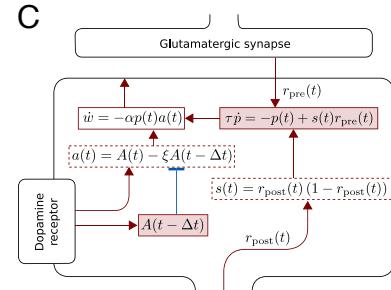
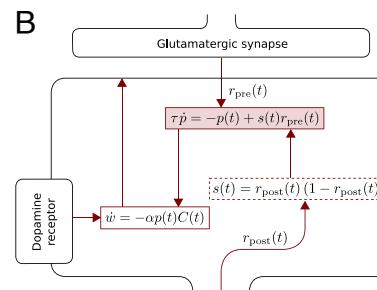
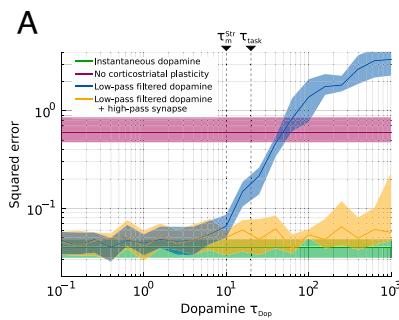


Fig. 3. Slow dopamine dynamics. (A) When the dopamine signal is low-pass filtered with a time constant slower than the time constant of the target output, the error (blue) increases. For comparison, the errors of previous models with instantaneous (green) and absent (purple) dopamine are also shown (these do not depend on the τ_{Dop}). Finally, the orange line shows the error when using the extended synaptic dynamics shown in panel C. All errors are reported as the median over 25 trials. Shaded areas indicate the first and third quartiles of the 25 trials. (B) The synaptic dynamics of the standard RFLO learning rule. (C) Adding an extra forward-inhibition motif to the synaptic dynamics to compensate for temporal smoothing of dopamine.

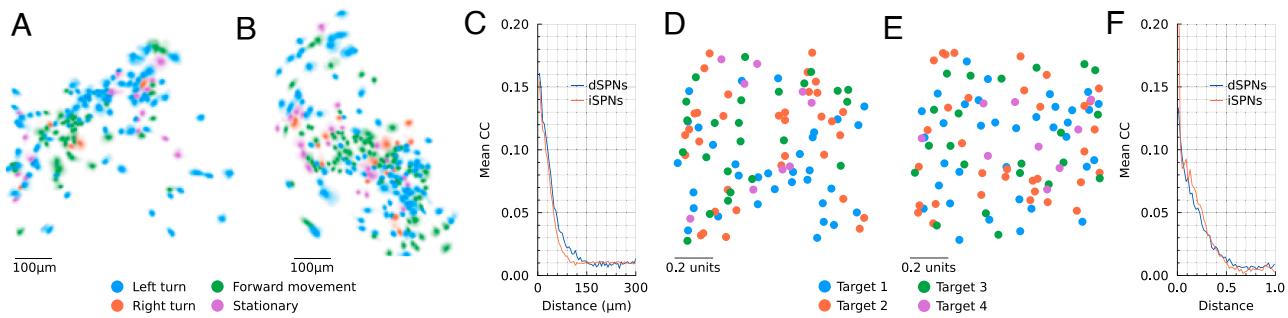


Fig. 4. Spatial correlations. (A) Calcium-imaging (1-photon) footprints of dSPNs in the *Right* dorsomedial striatum of a mouse exploring an open field. The footprints of the neurons are colored according to which basic movement they respond most strongly to. The most common movement is turns contralateral (*Left*) to the recorded hemisphere (*Right*), but all four movements are represented by some neurons. (B) Same as A, but iSPNs from a different mouse. (C) Mean Pearson's correlation coefficient between the deconvolved calcium traces of pairs of SPNs as a function of the distance between the center of their footprints in the field-of-view. Means were calculated by pooling data from 7 (dSPNs) and 8 (iSPNs) recording sessions. Data for A–C was adapted from ref. 47. (D) The x and y coordinate of simulated dSPNs, with colors indicating which target dimension it was most correlated with after 5,000 trials of training. The z coordinate was omitted, as if the cells are orthographically seen from above. (E) The iSPNs from the same simulation as in (D). (F) Mean Pearson's correlation coefficient between the firing rates of pairs of SPNs in the simulation as a function of the distance between them. The means are calculated across all pairs of SPNs pooled from 1,000 simulations. Distance is measured relative to the side of the cube.

error was consistently similar to instantaneous even for large τ_{DA} (Fig. 3A, yellow line).

Spatial Dopamine Induces Spatial Structure of SPN Responses.

A ubiquitous observation in 1-photon calcium imaging in rodents is that SPNs that are active during some action tend to be spatially close to other SPNs responding to the same action (43–47, Fig. 4 A and B). At the same time, the actions are not represented by entirely isolated clusters because examples of SPNs preferring any given action can always be found in all parts of the miniscopes' fields-of-view (44–47, Fig. 4 A and B). Mechanistically, this arrangement could be simply explained by spatially neighboring SPNs sometimes receiving inputs from shared cortical or thalamic axons, as well as from other SPNs or interneurons close-by. However, the algorithmic and computational significance of this spatial arrangement has been controversial, with some authors interpreting it in favor of spatially compact action channels (43, 46) and some authors in favor of a distributed representation of actions (44, 47).

Therefore, we asked whether the spatial structure of dopamine diffusion in our model was sufficient to induce the weak but significant distance-dependent correlation between pairs of SPNs together with a spatially distributed representation of the action. We found both phenomena (Fig. 4 D–F). In our model, target dimension are equivalent to the "actions". Consistent with the experimental data, we found examples of both dSPNs and iSPNs preferring all four target dimensions in all parts of a constructed field-of-view (Fig. 4 D–E). In addition, correlation between SPN firing rates ($r_j^{\text{Str}}(t)$) was around 0.15 for very close SPNs and decayed to zero for SPNs further apart (Fig. 4F), which closely resembles the shape of distance-dependent correlation between deconvolved calcium fluorescence traces (Fig. 4C). Note that we only used the spatial position of the SPNs (\vec{y}_j) to construct the matrix D (Eq. 13) and not to initialize the synaptic weights. Therefore, Fig. 4F shows exclusively the spatial structure induced by dopamine-dependent plasticity and we note that this is sufficient to reach experimental-level strength of the correlations. This is not the case for a much smaller or larger dopamine spatial constant λ (SI Appendix, Fig. S2). Nevertheless, we note that space-dependent connectivity probabilities and shared inputs will also likely contribute to the spatial dependence of the correlations in real experiments. In conclusion, we have demonstrated that the experimentally observed level of spatial correlations is compatible

also with a distributed representation of actions in the striatum and could be a consequence of dopamine being used as a vector-valued error.

Discussion

The broad, unspecific dopaminergic axonal projections have been argued to only allow for the transmission of a scalar homogeneous feedback signal (48). Here, we provide a tenable counterexample of this view, even if speculative and highly idealized. We demonstrate that in a reduced model of a piece of basal ganglia that a heterogeneous, vector-valued feedback signal could in fact be transmitted by dopamine, even if the dopaminergic projections in the striatum are random. We have identified four key requirements for effective use of vector-valued dopamine, which also serve as predictions that can be verified experimentally:

1. At least one projection downstream of the striatum must be plastic.
2. The vector-valued error must be available to both the dopaminergic (here, SNc) and the readout (here, SNr) populations.
3. Striatal dopamine dynamics must be at least as fast as the targeted movements, or, alternatively, be high-pass filtered by feed-forward inhibition in the synaptic biochemical pathways.
4. The dopamine received by each SPN must be sufficiently independent, or, stated formally, the effective connectivity matrix arising from summing contribution of individual varicosities (Eq. 13) must not be degenerate.

Direct experimental evidence for any downstream plasticity in the basal ganglia (Requirement 1) is scarce, but a recent study by González-Rodríguez et al. (49) showed that dopamine depletion in the SNr plays a larger role than striatal dopamine in producing motor deficits in Parkinson's disease. This is qualitatively consistent with the relative importance of striatonigral over corticostriatal plasticity in our model (Fig. 2F). However, we note that although we placed this plasticity in the striatofugal projection(s), it could in principle also be met by plasticity in the nigrothalamic or nigrocollicular projections.

A vector-valued error would likely appear as tunings to various motor and task variables in experimental animals (Requirement 2), especially in the phase before the animals are so overtrained their error is zero. Indeed, cells in the SNr (17, 26, 50)

as well as the SNC (27, 32, 51, 52) respond to a plethora of behavioral and task variables. We deliberately excluded the details of how this error may be computed in the brain, but we speculate at least three possible algorithmic ways in which it could appear:

1. Brain regions such as the motor cortex or cerebellum could have a forward model of the world as well as the target and thus can directly compute the error and send it to the midbrain.
2. The brain could be wired as a set of hierarchical control loops, in which each loop provides the target for the level below (as proposed by ref. 16). Each such loop could stretch throughout the cortex and basal ganglia.
3. If the executed action has more variability than the command read out by the SNr, the policy gradient theorem (21) states that the gradient for the update should be

$$r^{\text{SNC}}(t) \propto \delta(t) \nabla \ln p\left(a(t) \mid r^{\text{SNr}}(t)\right), \quad [20]$$

where a is the vector-valued action taken, δ is the temporal difference (TD) error as predicted by a critic, and p is the probability density function of a . Note that this suggests that SNC cells should fire proportionally to both the TD error and to (the gradient of) some behavioral variables, which could explain why many SNC cells appear tuned to both (53, 54). Lindsey and Litwin-Kumar (55) have proposed that the dorsal striatum could make use of such a policy gradient but nonetheless argue that dopamine itself is a scalar proportional to the squared norm of the policy gradient.

Importantly, none of these ways requires the vector-valued error to be provided directly from a supervisor external to the brain. Including one or more brain regions that use some of these principles to translate external rewards and internal goals into a vector-valued error or policy gradient will be a critical future extension of our model.

There are several ways the brain could implement filters that allow extraction of faster fluctuations in dopamine concentration (Requirement 3). Our example in Fig. 3 is highly idealized and makes arguably unfair use of our idealized perfect exponential decay of the dopamine. In reality, the journey of dopamine molecule from a varicosity to a dopamine receptor depends on the local geometry and dopamine reuptake so that the dependence on both time and distance is most likely complicated and nonlinear (although these effects might be less pronounced at very short distances; 56, 57). Nevertheless, evolution has had a good opportunity to tweak the biochemical pathways to compensate for these effects as far as permitted by the signal-to-noise ratio. Whether this is tenable in a realistic model of dopamine diffusion and biochemical cascades remains an open question, but we predict that there is at least one node in the biochemical cascade of dopamine-induced synaptic plasticity that is sensitive to fast fluctuations in local dopamine concentration.

The spatial frequency of the dopamine landscape in the striatum must be high enough so that even neighboring SPNs do not sense the exact same dopamine concentration (Requirement 4). This can be achieved by having a short spatial constant of dopamine diffusion, and possibly compensating with a larger number of varicosities (Fig. 2D). Consistent with our model, Cragg and Rice (56) estimated the diffusion distance of dopamine following release to a few microns.

The main goal of our work was to demonstrate that the broad and unspecific nigrostriatal dopaminergic projection can in principle transfer a usable vector-valued error to the striatum; our ambition was not to provide a complete biological account

of the process. For this reason, there are many likely very important features of basal ganglia anatomy and physiology we did not include, for example, dorsolateral/dorsomedial functional division in the striatum (58), the different roles of the matrix and the striosome (59), axonally initiated dopaminergic release by cholinergic interneurons (60, 61), saturating dopamine receptors (57), etc., Similarly, our primary goal was not to introduce a new algorithm for training recurrent neural networks; the network setup and plasticity rule is an application of the RFLO rule (37). Nevertheless, we show that the RFLO rule is applicable in a basal ganglia-like network with multiple inhibitory synapses and with our reduced model of dopamine feedback and propose vector-valued error feedback as a candidate functional role of dopamine.

Whether the striatum actually takes advantage of the vector-valued dopamine in the manner we proposed here could in principle be tested by building on existing experiments. For example, Bova et al. (62) have shown that optogenetically stimulating SNC of rats during a reaching task results in impairment also in subsequent, nonstimulated trials. This demonstrates the key role of midbrain dopamine in motor learning. A similar, but technically more challenging, experiment could be used to test the importance of the vector-valued nature of the dopamine signals in the striatum. If two opsins with opposing effects are simultaneously expressed in the dopaminergic axons in the striatum, they could be carefully stimulated using two-colored patterned illumination so that some terminals are excited and some are inhibited, while the net dopamine release in the area remains the same. Our prediction is that in this scenario, motor learning is impaired because even though the precise temporal nature of the net (that is, scalar) dopamine signal remains intact, its vector-valued nature is scrambled.

Previous proposals for use of heterogeneous dopamine (30, 31) assume that the heterogeneous responses of dopaminergic cells are transmitted to the striatum through private parallel channels without any cross-talk. However, this is not easily reconciled with functional and anatomical findings (*Introduction*). Similarly, Gardner et al. (63) have proposed that dopamine conveys a vector-valued sensory prediction error (SPE) that can be used to update a parameterized successor representation. Although such an SPE could explain the heterogeneous firing pattern of midbrain dopaminergic neurons, it would require the vector-valued signal to be transmitted intact to the receiving area (striatum). Another proposed use of heterogeneous firing in the midbrain dopaminergic neurons is to support a distributional coding of value (64). However, a distributional value code only explains different gains in the coding of the reward prediction error, not why the neurons respond to unrewarded task variables. Nevertheless, it is entirely possible that the brain simultaneously employs a distributional value code (perhaps most strongly in the VTA) for a “critic” subregion and a vector-valued error code (perhaps most strongly in the SNC) for an “actor” subregion.

In conclusion, we propose that the experimental observation of heterogeneous responses of dopamine cells (26–28) can represent a vector-valued error. By providing this type of error, the SNC supports the basal ganglia learning to select actions from a continuous action space in continuous time, thereby providing the animal with vital behavioral flexibility, control, and adaptability.

Materials and Methods

The dynamics of neurons, network structure, and learning rule is already described in the Results section. Here, we describe only the technical details needed to run the simulations.

Network Simulations. The network was simulated in a custom simulator written in Julia (65). The dynamics were simulated with forward-Euler with $\text{dt} = 1 \text{ ms}$. The number of units in each population is shown in *SI Appendix, Table S1*.

Simulations consisted of multiple trials concatenated after each other without any reset of the network in between. The current time in the current trial was signaled to the network by setting the thalamic firing rates to

$$r_m^{\text{Thal}}(t) = \phi \left(A_m \cos \frac{2\pi t}{T} + B_m \sin \frac{2\pi t}{T} \right), \quad [21]$$

where $T = 200 \text{ ms}$ is the duration of a single trial, and A_m and B_m are constants drawn randomly from a circle with radius 4 (i.e., $A_m^2 + B_m^2 = 4^2$ for all m). ϕ is the logistic transfer function (Eq. 2).

Initializing the Weights. For each pair of cells in each projection, there was a fixed probability (Fig. 1B) of a synapse being inserted. If a synapse was inserted, its weight was drawn from a uniform distribution $[0, w_{\max}/\sqrt{N_{\text{post}}}]$ (*SI Appendix, Table S2*) and then multiplied by -1 for the inhibitory projections. The weights in *SI Appendix, Table S2* were chosen for the network to have close to chaotic trajectories before training.

After all synapses were created, the sum of the weight of all incoming synapses was calculated for each neuron. If this sum was greater than 0, all the inhibitory synapses were slightly increased so that the new sum was exactly 0. Conversely, if the sum was less than 0, all the excitatory weights were slightly increased to reach sum 0. This ensured that each neuron had roughly balanced excitation and inhibition, which in turn created rich dynamics from the start.

1. Y. Niv, A. Langdon, Reinforcement learning with Marr. *Curr. Opin. Behav. Sci.* **11**, 67–73 (2016).
2. E. Perrin, L. Venance, Bridging the gap between striatal plasticity and learning. *Curr. Opin. Neurobiol.* **54**, 104–112 (2019).
3. W. Schultz, P. Dayan, R. M. Montague, A neural substrate of prediction and reward. *Science* **275**, 1593 (1997).
4. H. G. R. Kim *et al.*, A unified framework for dopamine signals across timescales. *Cell* **183**, 1600–1616.e25 (2020).
5. M. D. Humphries, R. D. Stewart, K. N. Gurney, A physiologically plausible model of action selection and oscillatory activity in the basal ganglia. *J. Neurosci.* **26**, 12921–12942 (2006).
6. T. C. Stewart, T. Bekolay, C. Eliasmith, Learning to select actions with spiking neurons in the basal ganglia. *Front. Neurosci.* **6**, 1–14 (2012).
7. P. Berthet, J. Hellgren-Kotaleski, A. Lansner, Action selection performance of a reconfigurable basal ganglia inspired model with Hebbian-Bayesian Go-NoGo connectivity. *Front. Behav. Neurosci.* **6** (2012).
8. R. Bolado-Gomez, K. Gurney, A biologically plausible embodied model of action discovery. *Front. Neuror.* **7**, 1–24 (2013).
9. K. N. Gurney, M. D. Humphries, P. Redgrave, A new framework for cortico-striatal plasticity: Behavioural theory meets in vitro data at the reinforcement-action interface. *PLoS Biol.* **13**, e1002034 (2015).
10. C. Baston, M. Ursino, A biologically inspired computational model of basal ganglia in action selection. *Comput. Intell. Neurosci.*, 1–24 (2015).
11. R. Bogacz, E. Martin Moraud, A. Abdi, P. J. Magill, J. Baufreton, Properties of neurons in external globus pallidus can support optimal action selection. *PLoS Comput. Biol.* **12**, 1–28 (2016).
12. P. Berthet, M. Lindahl, P. J. Tully, J. Hellgren-Kotaleski, A. Lansner, Functional Relevance of different basal ganglia pathways investigated in a spiking model with reward dependent plasticity. *Front. Neural Circ.* **10** (2016).
13. K. Dunovan, C. Vich, M. Clapp, T. Verstynen, J. Rubin, Reward-driven changes in striatal pathway competition shape evidence evaluation in decision-making. *PLOS Comput. Biol.* **15**, e1006998 (2019).
14. J. Bahuguna, P. Weidel, A. Morrison, Exploring the role of striatal D1 and D2 medium spiny neurons in action selection using a virtual robotic framework. *Euro. J. Neurosci.* **49**, 737–753 (2019).
15. P. Redgrave, T. Prescott, K. Gurney, The basal ganglia: A vertebrate solution to the selection problem? *Neuroscience* **89**, 1009–1023 (1999).
16. H. H. Yin, Action, time and the basal ganglia. *Philos. Trans. R. Soc. B: Biol. Sci.* **369** (2014).
17. J. W. Barter *et al.*, Basal ganglia outputs map instantaneous position coordinates during behavior. *J. Neurosci.* **35**, 2703–2716 (2015).
18. P. E. Rueda-Orozco, D. Robbe, The striatum multiplexes contextual and kinematic information to constrain motor habits execution. *Nat. Neurosci.* **18**, 453 (2015).
19. J. Park, L. T. Coddington, J. T. Dudman, Basal ganglia circuits for action specification. *Ann. Rev. Neuro.* **43**, 485 (2020).
20. A. K. Dhawale, S. B. E. Wolff, R. Ko, B. P. Ölvéczky, The basal ganglia control the detailed kinematics of learned motor skills. *Nat. Neuro.* **24**, 1256 (2021).
21. R. S. Sutton, D. McAllester, S. Singh, Y. Mansour, “Policy gradient methods for reinforcement learning with function approximation” in *Advances in Neural Information Processing Systems*, S. Solla, T. Leen, K. Müller, Eds. (MIT Press, 1999), vol. 12, pp. 1057–1063.
22. B. A. Richards *et al.*, A deep learning framework for neuroscience. *Nat. Neurosci.* **22**, 1761–1770 (2019).
23. V. Mnih *et al.*, “Asynchronous methods for deep reinforcement learning” in *Proceedings of The 33rd International Conference on Machine Learning, Proceedings of Machine Learning Research*, M. F. Balcan, K. Q. Weinberger, Eds. (PMLR, New York, New York, USA, 2016), vol. **48**, pp. 1928–1937.
24. T. P. Lillicrap *et al.*, “Continuous control with deep reinforcement learning” in *4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico, May 2–4, 2016, Conference Track Proceedings*, Y. Bengio, Y. LeCun, Eds. (2016), pp. 2–4.
25. J. Schulman, F. Wolski, P. Dhariwal, A. Radford, O. Klimov, Proximal policy optimization algorithms. *CoRR* abs/1707.06347 (2017).
26. D. Fan, M. A. Rossi, H. H. Yin, Mechanisms of action selection and timing in substantia nigra neurons. *J. Neurosci.* **32**, 5534–5548 (2012).
27. M. W. Howe, D. A. Dombek, Rapid signalling in distinct dopaminergic axons during locomotion and reward. *Nature* **535**, 505 (2016).
28. B. Engelhard *et al.*, Specialized coding of sensory, motor and cognitive variables in VTA dopamine neurons. *Nature* **570**, 509 (2019).
29. G. W. Arbutnott, J. Wickens, Space, time and dopamine. *Trends Neurosci.* **30**, 62–69 (2007).
30. A. A. Hamid, M. J. Frank, C. I. Moore, Wave-like dopamine dynamics as a mechanism for spatiotemporal credit assignment. *Cell* **184**, 2733–2749.e16 (2021).
31. R. S. Lee, B. Engelhard, I. B. Witten, N. D. Daw, A vector reward prediction error model explains dopaminergic heterogeneity. *bioRxiv* p. 2022.02.28.482379 (2022).
32. R. Avisati *et al.*, Distributional coding of associative learning within projection-defined populations of midbrain dopamine neurons. *bioRxiv* (2022).
33. B. J. Hunnicutt *et al.*, A comprehensive excitatory input map of the striatum reveals novel functional organization. *eLife* **5**, e19103 (2016).
34. N. N. Foster *et al.*, The mouse cortico “basal ganglia” thalamic network. *Nature* **598**, 188–194 (2021).
35. W. Matsuda *et al.*, Single nigrostriatal dopaminergic neurons form widely spread and highly dense axonal arborizations in the neostriatum. *J. Neurosci.* **29**, 444–453 (2009).
36. T. P. Lillicrap, D. Cownden, D. B. Tweed, C. J. Akerman, Random synaptic feedback weights support error backpropagation for deep learning. *Nat. Comm.* **7** (2016).
37. J. M. Murray, Local online learning in recurrent networks with random feedback. *eLife* **8**, e43299 (2019).
38. E. M. Izhikevich, Solving the distal reward problem through linkage of STDP and dopamine signaling. *Cerebral Cortex* **17**, 2443–2452 (2007).
39. S. D. Fisher *et al.*, Reinforcement determines the timing dependence of corticostriatal synaptic plasticity in vivo. *Nat. Commun.* **8** (2017).
40. W. Gerstner, M. Lehmann, V. Liakoni, D. Corneil, J. Brea, Eligibility traces and plasticity on behavioral time scales: Experimental support of neohebbian three-factor learning rules. *Front. Neural Circ.* **12** (2018).
41. H. Jaeger, H. Haas, Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication. *Science* **304**, 78–80 (2004).
42. W. Maass, H. Markram, On the computational power of circuits of spiking neurons. *J. Comput. Syst. Sci.* **69**, 593–616 (2004).
43. G. Barbera *et al.*, Spatially compact neural clusters in the dorsal striatum encode locomotion relevant information. *Neuron* **92**, 202–213 (2016).
44. A. Klaus *et al.*, The spatiotemporal organization of the striatum encodes action space. *Neuron* **95**, 1171–1180.e7 (2017).
45. J. G. Parker *et al.*, Diametric neural ensemble dynamics in parkinsonian and dyskinetic states. *Nature* **557**, 177–182 (2018).
46. J. H. Shin, M. Song, S. B. Paik, M. W. Jung, Spatial organization of functional clusters representing reward and movement information in the striatal direct and indirect pathways. *Proc. Natl. Acad. Sci. U.S.A.* **117** (2020).
47. M. Weglage *et al.*, Complete representation of action space and value in all dorsal striatal pathways. *Cell Rep.* **36**, 109437 (2021).

Target Signals. The targets were drawn from a Gaussian process with mean 0.5 and variance given by

$$\sigma^2(t_1, t_2) = 0.15^2 \exp \left(-\frac{\delta(t_1, t_2)^2}{\tau_{\text{task}}^2} \right), \quad [22]$$

where $\delta(t_1, t_2)$ is the smallest difference between t_1 and t_2 when including wrap-around, i.e.

$$\delta(t_1, t_2) = \min(|t_1 - t_2|, |t_1 - t_2 + T|, |t_1 - t_2 - T|), \quad [23]$$

where $T = 200 \text{ ms}$ is the duration of a single trial. The periodic kernel is to avoid discontinuities when running consecutive trials without resetting the network. For all experiments in the main figures, $\tau_{\text{task}} = 20 \text{ ms}$.

Data, Materials, and Software Availability. Code for reproducing the simulations can be found at GitHub at <https://github.com/emiwar/diffuse-dopamine-rnn> (66).

ACKNOWLEDGMENTS. We thank Dr. Moritz Weglage and Prof. Erik Fransén for helpful discussions and Lihao Guo and Hauke Wernecke for proofreading the manuscript. Partial funding from StratNeuro (to A.K.), Swedish Research Council (to A.K.), Digital Futures (to A.K.), and Karolinska Institutet (KID doctoral funding to E.W.) is gratefully acknowledged.

48. R. Chen, J. H. Goldberg, Actor-critic reinforcement learning in the songbird. *Curr. Opin. Neurobiol.* **65**, 1–9 (2020).
49. P. González-Rodríguez *et al.*, Disruption of mitochondrial complex i induces progressive parkinsonism. *Nature*, 1476 (2021).
50. Y. Tang *et al.*, Opposing regulation of short-term memory by basal ganglia direct and indirect pathways that are coactive during behavior. *bioRxiv* (2021).
51. P. D. Dodson *et al.*, Representation of spontaneous movement by dopaminergic neurons is cell-type selective and disrupted in parkinsonism. *Proc. Natl. Acad. Sci. U.S.A.* **113**, E2180–E2188 (2016).
52. L. T. Coddington, J. T. Dudman, The timing of action determines reward prediction signals in identified midbrain dopamine neurons. *Nat. Neurosci.* **21**, 1563–1573 (2018).
53. N. F. Parker *et al.*, Reward and choice encoding in terminals of midbrain dopamine neurons depends on striatal target. *Nat. Neurosci.* **19**, 845–854 (2016).
54. R. S. Lee, M. G. Mattar, N. F. Parker, I. B. Witten, N. D. Daw, Reward prediction error does not explain movement selectivity in DMS-projecting dopamine neurons. *eLife* 1–16 (2019).
55. J. Lindsey, A. Litwin-Kumar, Action-modulated midbrain dopamine activity arises from distributed control policies. *arXiv* pp. 1–17 (2022).
56. S. J. Cragg, M. E. Rice, DAncing past the DAT at a DA synapse. *Trends Neurosci.* **27**, 270–277 (2004).
57. C. Liu, P. Goel, P. S. Kaeser, Spatial and temporal scales of dopamine transmission. *Nat. Rev. Neurosci.* **22**, 345–358 (2021).
58. B. W. Balleine, J. P. O'Doherty, Human and rodent homologies in action control: Corticostriatal determinants of goal-directed and habitual action. *Neuropsychopharmacology* **35**, 48–69 (2010).
59. B. Bloem, R. Huda, M. Sur, A. M. Graybiel, Two-photon imaging in mice shows striosomes and matrix have overlapping but differential reinforcement-related responses. *eLife* **6** (2017).
60. S. Threlfell *et al.*, Striatal dopamine release is triggered by synchronized activity in cholinergic interneurons. *Neuron* **75**, 58–64 (2012).
61. C. Liu *et al.*, An action potential initiation mechanism in distal axons for the control of dopamine release. *Science* **375**, 1387–1385 (2022).
62. A. Bova *et al.*, Precisely timed dopamine signals establish distinct kinematic representations of skilled movements. *eLife* **9**, e61591 (2020).
63. M. P. Gardner, G. Schoenbaum, S. J. Gershman, Rethinking dopamine as generalized prediction error. *Proc. R. Soc. B: Biol. Sci.* **285** (2018).
64. W. Dabney *et al.*, A distributional code for value in dopamine-based reinforcement learning. *Nature* **577**, 671 (2020).
65. J. Bezanson, A. Edelman, S. Karpinski, V. B. Shah, Julia: A fresh approach to numerical computing. *SIAM Rev.* **59**, 65–98 (2017).
66. E. Wärnberg, Code for "Feasibility of dopamine as a vector-valued feedback signal in the basal ganglia." GitHub. <https://github.com/emiwar/diffuse-dopamine-rnn>. Deposited 24 April 2023.