# ALIGNING COGNITIVE AND NEURAL MECHANISMS IN TINY RECURRENT NEURAL NETWORKS

**Anonymous authors**
Paper under double-blind review

## ABSTRACT

A central goal of computational neuroscience is to develop models that both accurately predict behaviour and reveal the cognitive and neural mechanisms that generate it. Tiny recurrent neural networks (RNNs) with few hidden units have helped reconcile these demands, combining predictive flexibility with interpretable dynamics. However, although choice logits and their update dynamics are reproducible across fits, this does not imply reproducibility of the internal mechanisms that generate them. Here we show that tiny RNNs fit to subtly different training data (e.g. different cross-validation folds) can yield very different internal update mechanisms and hidden state trajectories, despite near-identical choice logits. This precludes using internal variables from tiny RNNs as candidate mechanisms underlying cognitive and neural computations. We characterise this problem across tiny RNNs trained on behavioural data from mice performing a probabilistic reversal learning task, then demonstrate how a series of biologically inspired constraints can greatly improve mechanistic reproducibility without compromising behavioural predictions. We were then able to interpret these mechanisms to explore possible computations in fronto-striatal circuits. Our results highlight mechanistic reproducibility as a complement to predictive accuracy and interpretability, and provide a framework for using RNN-derived variables to explore hypotheses about behavioural and neural dynamics during flexible decision-making.

## 1 INTRODUCTION

Understanding the mechanisms of human or animal behaviour requires models that are both predictive and interpretable (Eckstein et al., 2023; Summerfield & Miller, 2023). Fitting hand-crafted reinforcement learning (RL) (Miller et al., 2021; Niv, 2009) or Bayesian inference (Qü et al., 2025; Mishchanchuk et al., 2024; Blanco-Pozo et al., 2024) models to behavioural data has long offered quantitative and transparent insights into candidate cognitive mechanisms in the form of interpretable model parameters and decision variables, often supported by neural correlates in prefrontal-striatal circuits. However, such cognitive models can lead to incomplete or incorrect characterisation due to lack the expressivity (Eckstein et al., 2023), researcher bias and subjectivity (Eckstein et al., 2021; Schaaf et al., 2023).

A promising alternative approach to modelling adaptive behaviour is fitting recurrent neural networks (RNNs) directly to decision-making task data. As RNNs can approximate a diverse range of model structures in their weights, they afford the expressiveness to accurately predict behaviour. However, this increased flexibility has traditionally come at the cost of lacking interpretable internal decision variables and update mechanisms underpinning their choice behaviour. To combine the interpretability of cognitive modelling and flexibility of RNNs, several recent frameworks apply constraints to them: *hybrid neural-cognitive models* combine elements of both these approaches (Eckstein et al., 2026); *disentangled RNNs* enforce representational sparsity and information "bottlenecks" (Miller et al., 2023); *tiny RNNs* limit networks to a very small number of hidden units (Ji-An et al., 2025).

Whilst these frameworks make individual fitted networks more accessible to interpretation, they crucially do not consider whether these interpretations are *reproducible* across fits. Indeed, networks with similar training loss can learn different internal solutions, a phenomenon referred to



Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

as *solution degeneracy* (Huang et al., 2025; Maheswaranathan et al., 2019; Turner et al., 2021; Maheswaranathan et al., 2019). Solution degeneracy precludes using the internal computations of fitted models as candidate mechanisms for understanding choice behaviour or neural activity. Behavioural fit and interpretability provide necessary, but not sufficient, criteria for favouring one set of explanatory variables over another.

To address this, we make the following contributions:

- **Characterise mechanistic variability** across fits of *tiny RNNs* trained on behavioural data from mice performing a two-armed bandit probabilistic reversal learning task.
- **Improve mechanistic reproducibility** by applying biologically inspired constraints and architectural simplifications, without compromising predictive performance.
- **Interpret the reproducible mechanisms** of our two-dimensional "monoGRU", and generate trial-by-trial hidden unit and gate activations as candidate computations and predictors of dopamine dynamics recorded from the prefrontal cortex and striatum.

Taken together, we present a framework for enabling RNNs trained on behavioural data to be used as candidate mechanistic models of cognition, bridging computations and brain activity.

## 2 Reducing mechanistic variability with biologically constrained RNNs

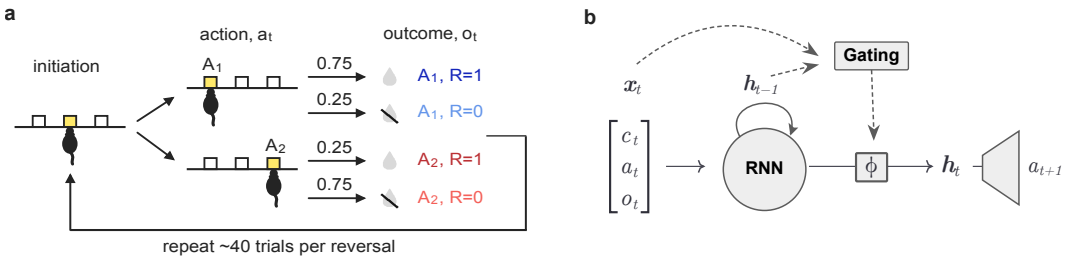



Figure 1: **Overview of probabilistic reversal learning and Tiny RNN framework. a)** Reversal learning task structure. Mouse chose between two options to obtain probabilistic rewards. Reward probabilities for the two options reversed at block transitions every ∼40 trials. **b)** RNN architecture overview. Inputs (cue $c_t$, action $a_t$, outcome $o_t$) are combined with past hidden states $\mathbf{h}_{t-1}$, including some gating mechanisms and a non-linear function $\phi$, to update the hidden state, $\mathbf{h}_t$, which is then decoded to predict the upcoming action $a_{t+1}$.

**Probabilistic reversal learning.** Probabilistic reversal learning tasks provide a well-characterised setting for studying adaptive reward-guided behaviour, supported by extensive literature using RL (Miller et al., 2021), Bayesian inference (Qü et al., 2025; Mishchanchuk et al., 2024) and data-driven cognitive model discovery (Kasenberg et al., 2026; Castro et al., 2025) methods. Here we trained mice (n=11) on a 2-armed bandit variant with forced choices. For each trial, $t$, the mice chose an action $a_t$, either left ($A_1$) or right ($A_2$), and observed an outcome, $o_t$, either reward ($R1$) or omission ($R0$). Reward probability was $75\%$ on the current 'good' side, and $25\%$ on the current bad side (Figure 2**a**). These probabilistic action-reward contingencies were reversed across blocks without an explicit cue. To encourage exploration across choices during learning, $25\%$ of trials were forced choices in which only one randomly selected option was available for selection.

**Predicting choices with tiny RNNs.** We followed the framework of Ji-An et al. (2025) to predict the mouse behaviour using tiny RNNs. To maximise interpretability, we trained recurrent neural networks with only 1 or 2 hidden units. For each trial, $t$, the RNN outputs a prediction of the upcoming action, $\hat{a}_{t+1}$, by decoding a hidden state, $\mathbf{h}_t = f(\mathbf{x}_t, \mathbf{h}_{t-1})$, which is recurrently updated using by combining inputs $\mathbf{x}_t$ and past hidden states, $\mathbf{h}_{t-1}$. Here the inputs are a unipolar encoding of the forced choice cue, the last action, and last outcome $\mathbf{x}_t = [c_t, a_t, o_t] \in \{0, 1\}^3$ (Figure 2**c**).

These RNNs predict upcoming actions by representing the action preference of the mouse in each trial, formally defined as the logit:

$$L_t = \log(P_t(A_1)/P_t(A_2)), \text{ where } [P_t(A_1), P_t(A_2)]^T = \text{Softmax}(W_{\text{decoder}}\mathbf{h}_t + b_{\text{decoder}}) = \hat{a}_{t+1}$$

and are trained with nested cross-validation to minimise the negative log-likelihood (NLL) against the upcoming action $a_{t+1}$. We also followed Ji-An et al. (2025) by using a Gated Recurrent Unit (Chung et al., 2014), which involves a per-unit update gate, $\mathbf{z}_t$, and reset gate, $\mathbf{r}_t$, which modulate how past hidden states and inputs combine to form the hidden state. We found that 2-unit RNNs significantly outperformed 1-unit models across subjects (subject-level paired Wilcoxon signed-rank test, $W = 1, n = 11, p < 0.01$; Figure B.1). All further analyses were therefore conducted on 2-unit RNNs.

**Mechanistic variability in tiny GRUs.** Independently trained RNNs, initialised with different parameters or trained on different subsets of the same dataset, often suffer from solution degeneracy: similar functional outputs with different internal representations and connectivity weights (Maheswaranathan et al., 2019; Huang et al., 2025). Here the internal representation is the hidden state, $\mathbf{h}_t$, which represents features of a dataset along the coordinate basis of each hidden unit (Godfrey et al., 2023). Before we can measure differences in internal representations and connectivity, it is crucial to account for symmetries which preserve functional equivalence. For example, we can always permute the identity of any two neurons without affecting the representation. Similarly for units with symmetric Tanh activations, flipping the sign of the neuron's input and output weights exactly preserves network function. Before comparing model $A$ to model $B$, we therefore align the model parameters up to functional equivalence, adapting recent methods (Ainsworth et al., 2023) to gated RNNs. Let $\mathcal{W}_A$ and $\mathcal{W}_B$ denote the set of weights for models $A$ and $B$ respectively. Let $T$ be a transformation that preserves functional equivalence of model $B$, then we align model $B$ to model $A$ by applying the transformation, $T^*$, which minimises the Frobenius norm between the weights (Figure 2**a**):

$$T^* = \arg\min_T \|\mathcal{W}_A - T(\mathcal{W}_B)\|_F^2 \tag{1}$$

For models with $\tanh$ activations, $T$ can be a combination of permutations or sign flips Chen et al. (1993). For models with ReLU activations we allow permutations and positive scaling Godfrey et al. (2023). By comparing aligned models we can measure differences in computational mechanisms.

We compared model fits within subject, where models were trained on slightly different subsets of data (different training folds), and across subjects. After aligning two models, we computed the cosine similarity between their flattened weight vectors. We then evaluated the two aligned models on the same data, resulting in trial-by-trial outputs (logits) and hidden state activations, which we then correlated to compute similarity (Pearson r). We also computed similarities for the trial-by-trial activations of update gates $\mathbf{z}_t$ and reset gates $\mathbf{r}_t$, as these are an important part of the computational mechanism which updates the internal representation $\mathbf{h}_t$.

Output logits were strikingly similar across fits, both within and between subjects. However, model weights and internal variables showed more variability, particularly between fits to different subjects (Table 1). This poses a challenge both for interpreting mechanisms, and characterising whether GRU solutions correspond to biological computations through comparing internal model variables to brain activity.

Table 1: Similarity (mean$\pm$ std) for GRU and monoGRU

| | **Within Subject** | | **Across Subjects** | |
|---|---|---|---|---|
| **Component** | **GRU** | **monoGRU** | **GRU** | **monoGRU** |
| Logit | $0.98 \pm 0.02$ | $0.99 \pm 0.01$ | $0.92 \pm 0.05$ | $\mathbf{0.94 \pm 0.03}^{**}$ |
| Weight | $0.77 \pm 0.26$ | $\mathbf{0.93 \pm 0.10}^{**}$ | $0.49 \pm 0.21$ | $\mathbf{0.84 \pm 0.12}^{***}$ |
| Hidden State | $0.87 \pm 0.17$ | $\mathbf{0.94 \pm 0.10}^{*}$ | $0.64 \pm 0.21$ | $\mathbf{0.87 \pm 0.13}^{***}$ |
| Update Gate | $0.79 \pm 0.27$ | $0.81 \pm 0.21$ | $0.55 \pm 0.30$ | $0.54 \pm 0.27$ |
| Reset Gate | $0.62 \pm 0.52$ | N/A | $0.21 \pm 0.49$ | N/A |

Bold values with asterisks denote a statistically significant difference from baseline GRU via subject-level paired $t$-test with FDR correction ($^{*}p_{\text{FDR}} < 0.05$, $^{**}p_{\text{FDR}} < 0.01$, $^{***}p_{\text{FDR}} < 0.001$).

**Minimising mechanistic variability with biological contraints and architectural simplification.** The variability we observed suggests that the GRU architecture is too expressive for the reversal learning task, so we added constraints and ablated components of the architecture (see Appendix B for full results). The resulting RNN architecture with highest similarity across models, both within and across subjects, involved 1) minimal biological constraints, 2) simplifying gating to only a single scalar update gate, and 3) removing the decoder bias term. We refer to this architecture as "monoGRU" (Table 2).

Recent work has shown that minimal biological constraints can recover features of single-cell tuning in biological networks in RNNs (Whittington et al., 2023; Whittington & Dorrell, 2026). We therefore trained models with non-negative firing rates, using a ReLU activation function instead of $\tanh$, and added energy minimisation and weight sparsity to the loss term (Appendix A.4). We expected this to reduce variability by breaking the degeneracy between otherwise equally good solutions.

In order to reduce variability in the update mechanisms, we simplified them by removing the reset gates, as in "Light GRU" and "Update-gate RNN" (Ravanelli et al., 2018; Collins et al., 2017). We then replaced unit-specific update gates with a single 1-dimensional update gate for all units, hence "monoGRU". This is related to leaky-integration units Jaeger et al. (2007) but with an adaptive gain, which was also motivated by principles of neuromodulatory gain modulation in the cortex (Ferguson & Cardin, 2020; Shine et al., 2021). Intuition for how the scalar update gate affects hidden state dynamics can be derived from considering the extreme cases of the gate being fully open, $z_t = 0$, or fully closed, $z_t = 1$, letting the hidden state either do a "recurrent update" or "persist":

$$\mathbf{h}_t = z_t \cdot \mathbf{h}_{t-1} + (1 - z_t)\mathbf{n}_t \quad \begin{cases} \xrightarrow{z_t = 0} & \mathbf{h}_t = \mathbf{n}_t \quad \textit{"recurrent update"} \\ \xrightarrow{z_t = 1} & \mathbf{h}_t = \mathbf{h}_{t-1} \quad \textit{"persist"} \end{cases}$$

Lastly, we found that removing the bias term in the decoder helps prevent representational collapse onto 1D in 2-unit networks. This collapse was partially encouraged by the energy minimisation and sparsity constraints introduced earlier. This increased the similarity of models even when compared across subjects. Update equations for the standard GRU and monoGRU are compared in Table 2.

Table 2: Comparison of standard GRU and monoGRU

| **Component** | **Standard GRU** | **monoGRU** |
|---|---|---|
| Hidden state | $\mathbf{h}_t = (1 - \mathbf{z}_t) \odot \mathbf{n}_t + \mathbf{z}_t \odot \mathbf{h}_{t-1}$ | $\mathbf{h}_t = (1 - z_t) \cdot \mathbf{n}_t + z_t \cdot \mathbf{h}_{t-1}$ |
| Update gate | $\mathbf{z}_t = \sigma(\mathbf{W}_{iz}\mathbf{x}_t + \mathbf{W}_{hz}\mathbf{h}_{t-1} + \mathbf{b}_z)$ | $z_t = \sigma(\mathbf{W}_{iz}\mathbf{x}_t + \mathbf{W}_{hz}\mathbf{h}_{t-1} + \mathbf{b}_z)$ |
| Candidate state | $\mathbf{n}_t = \phi(\mathbf{W}_{in}\mathbf{x}_t + \mathbf{W}_{hn}(\mathbf{r}_t \odot \mathbf{h}_{t-1}) + \mathbf{b}_n)$ | $\mathbf{n}_t = \phi(\mathbf{W}_{ih}\mathbf{x}_t + \mathbf{W}_{hh}\mathbf{h}_{t-1} + \mathbf{b}_h)$ |
| Reset gate | $\mathbf{r}_t = \sigma(\mathbf{W}_{ir}\mathbf{x}_t + \mathbf{W}_{hr}\mathbf{h}_{t-1} + \mathbf{b}_r)$ | None |
| Nonlinearity | $\phi(x) = \tanh(x) \in [-1, 1]$ | $\phi(x) = \text{ReLU}(x) \in \mathbb{R}^+$ |
| Decoder | $\hat{a}_{t+1} = \text{Softmax}(\mathbf{W}_{\text{decoder}}\mathbf{h}_t + \mathbf{b}_{\text{decoder}})$ | $\hat{a}_{t+1} = \text{Softmax}(\mathbf{W}_{\text{decoder}}\mathbf{h}_t)$ |

Our constraints and architectural simplification did not affect predictive performance, when comparing 2-unit GRU to 2-unit monoGRU with a subject-level paired Wilcoxon Signed rank test ($W = 26, n = 11, p = 0.58$; Figure B.1).

We measured mechanistic variability of the monoGRU by computing model similarity across fits as before, both within and between subject, and tested for significant differences compared to the standard GRU via subject-level paired t-tests (Table 1, Figure 2). Both model weights and the trial-by-trial hidden state dynamics were significantly more similar across fits for monoGRU compared to standard GRU, with the greatest improvement for between subject similarity where the standard GRU fits were most variable. While we found no significant difference in the similarity of update gate activations between GRU and monoGRU models, the reset gate - which showed the largest variability across fits in the standard GRU, is completely eliminated in the monoGRU. We note the high similarity in weights and hidden states across monoGRU fits, even for models trained on data from different subjects ($0.84 \pm 0.12$). This suggests that monoGRU models share a common

computational motif for the reversal learning task, which can be readily interpreted at the level of parameters and consistently compared to neural data.

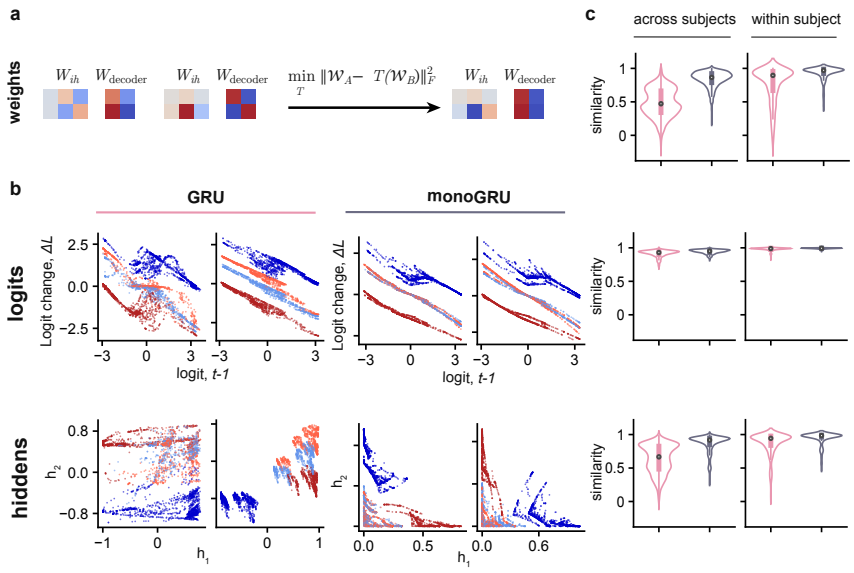



Figure 2: **Minimising mechanistic variability with monoGRU. a)** Example alignment of two models to maximise similarity of their weight matrices through transformations that exactly preserve the model dynamics. **b)** Example model fits for a single subject for two different cross validation folds. Standard GRU fits (left) have very similar logit dynamics (top) but different hidden state dynamics (bottom) across folds. monoGRU fits (right) have similar logit and hidden state dynamics across folds. Points show individual trials coloured by trial type. **c)** Similarity of weights (top) logit time series (middle) and hidden state time series (bottom) across fits to different subjects data (left) and to different cross-validation folds within subject (right).

## 3 Interpreting mechanisms of two-dimensional monoGRU

Having reduced mechanistic variability in the monoGRU's internal variables across fits, we examined them at two complementary levels of interpretation: how these variables represented action-reward histories, and how learned weights and state dynamics supported choice readout and state updating. Since there are only two hidden units, and a total of 22 parameters, we were able to examine these directly with no dimensionality reduction techniques. We highlight three key computational features.

**Hidden units learn an axis-aligned representation of action preference.** RNNs trained on next-choice prediction represent action preference in their hidden units, which can be done in a variety of ways. However, in monoGRU, each hidden unit consistently aligns with the mouse's preference for a specific action. When $h_1$ is high, the mouse will likely choose left, and when $h_2$ is high, the mouse will likely choose right. This is evident in the trial-by-trial activations of each unit (Figure 3**a**), and the representation geometry of the hidden units (Figure 3**a**). This axis-aligned representation geometry has the benefit of being easily decodable: simply take the difference of the two hidden units to read out the predicted action, $L_t \propto (h_{1,t} - h_{2,t})$. This is approximately what the decoder does (Figure 3**b, c**).

**Hidden units accumulate integrate reward history.** Mouse behaviour is characterised by seeking rewards, which is therefore a key feature that our models learn from predicting choices. Hidden unit activations across trials (Figure 3**a**) suggest that activity of hidden unit $h_1$ is increased by rewards on the left and decays on other trial types, while hidden unit $h_2$ shows the corresponding pattern for rewards on the right. This pattern is confirmed by an ordinary least squares regression predicting the activation of each hidden unit from the trial history (Figure 3**c**).

Taking a closer look at the weights from inputs to the candidate state, $\mathbf{n}_t$, (Figure 3**b, d**), there's a positive weight from reward outcome, $o_t \in 0, 1$, to both units, while there's an asymmetric contribution of action $a_t \in \{0, 1\}$. An analysis of the average parameters across subjects (Appendix D) suggests that the update of the each hidden unit approximates an action-specific exponential moving average (EMA) of rewards with passive forgetting. This relates monoGRU to an established Rescorla-Wagner model known as Q-learning with forgetting (Katahira, 2015), with a key difference of $z_t$ acting as an adaptive decay rate. Furthermore, the sum of the hidden units, $h_{1,t} + h_{2,t}$, approximates an EMA of total reward. We highlight that this is derived from strong modulation of reward on the inputs to the candidate state in combination with a global update gate.

**Attractor dynamics modulated by the global update gate.** Persistent activity in mPFC has been related to stable states in the dynamical system of neural activity, referred to as attractors Inagaki et al. (2019). Such dynamics are relevant to our models, since parameters are fixed after training and the hidden state is then be updated via recurrent dynamics. We therefore performed a linear stability analysis on the candidate state parameters under fixed inputs for each free choice trial type, resulting in stable dynamics in each case (Appendix A.6. Each trial type pushes the network towards a distinct fixed point aligned with the representation geometry (Figure 3**a**). Here the global adaptive update gate, $z_t$, has been formally related to changing the timescale of dynamics in RNNs (Tallec & Ollivier, 2018; Livi, 2025). In theory, when $z_t$ is low the dynamics are faster and when $z_t$ is high the dynamics are slower. This would allow the RNN to persist its hidden state even when the parameters of the candidate state would otherwise overwrite it. In practice, this is also case: around reversals where there is a transient increase in the update gate, which slows down the dynamics (Figure 3**b**). This is counter-intuitive, as one might expect the gate to encourage persistence during stable task states, where the mouse is consistently choosing the rewarded action. However, this persistence is supported by the attractor dynamics of the candidate state during rewarded trials. The update gate instead plays a role in slowing down forgetting during unrewarded trials. From gate activations across trials (Figure 3**b,c**) and the parameters (Figure 3**b**), we note that the update gate is strongly modulated by reward via $o_t$, as well as reward history via $(h_{1,t-1} + h_{2,t-1})$.

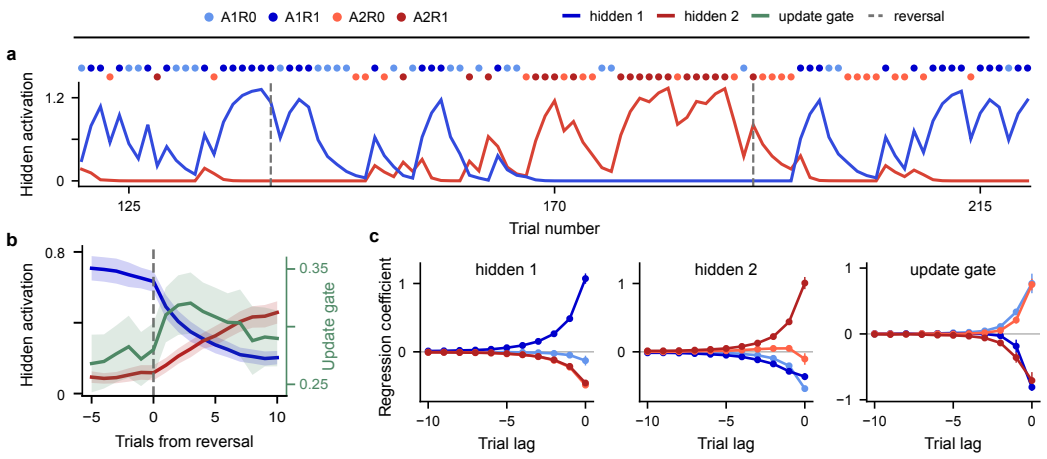



Figure 3: **Hidden state and gate dynamics of 2-unit monoGRU during reversal learning. a)** Trial-by-trial time-series of monoGRU hidden unit activations during an example session. Coloured dots indicate the four action-outcome trial-types; dashed vertical lines mark contingency reversals. **b)** Hidden unit activations (left axis) and update gate (right axis) around left to right reversals. For clarity, hidden units are aligned across subjects according to their decoder contributions favouring A1 or A2. **c)** Linear regressions predicting hidden unit and update gate activation from the trial event history of the previous 10 trials. Light dots represent individual subjects, and error bars indicate cross-subject mean ± s.e.m.

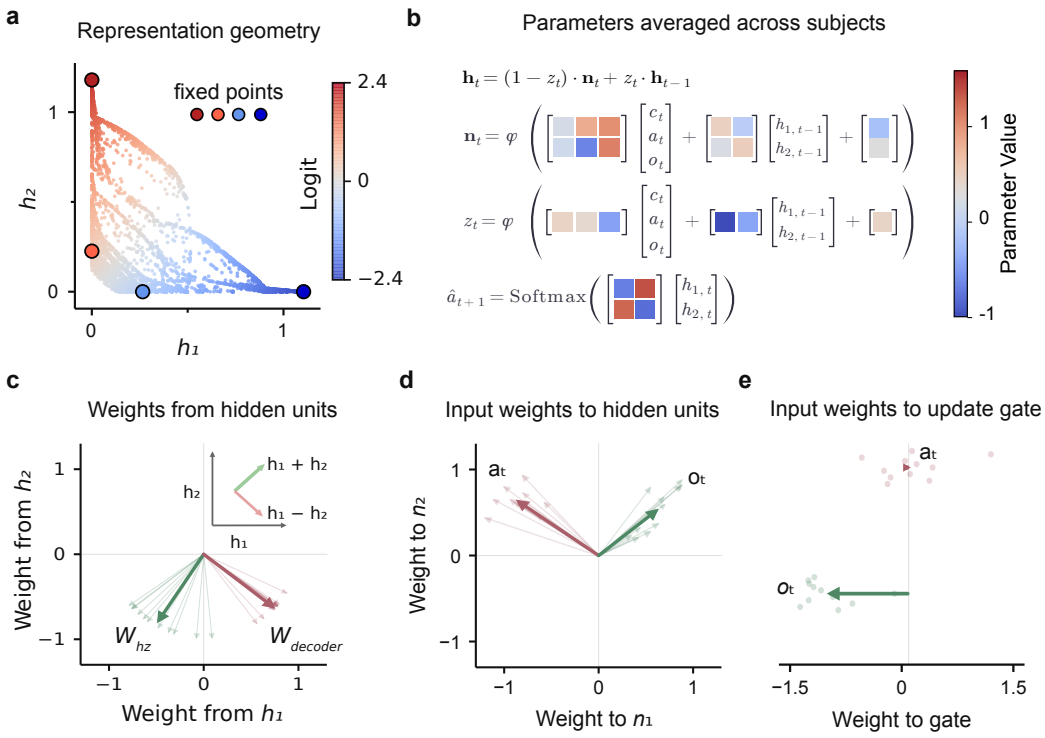



Figure 4: **Representation geometry and weights of 2-unit monoGRU. a)** Example hidden unit activations from a single subject, coloured by logit, $\log(P(A1)/P(A2))$, predicting the next-choice. Coloured circles show fixed points under a linear stability analysis of the candidate state parameters for each trial-type input. **b)** Weights of monoGRU, averaged across all subjects and folds. **c)** Readout weights of the hidden state from the decoder, $W_{decoder}$, and the update gate, $W_{hz}$, projected onto the hidden unit axes. In the inset, we plot the hidden state sum-axis and difference-axis. **d)** Input weights of action, $a_t$, and outcome, $o_t$, projected onto the hidden unit axes. **c,** Input weights of action, $a_t$, and outcome, $o_t$, projected along the update gate axis.

## 4 Aligning Behavioural and Neural Mechanisms with monoGRU

Having achieved reproducible trial-by-trial dynamics of monoGRU's internal mechanisms and characterised a computational motif linking hidden-state activity to choice and gating, we asked whether these variables could explain variation in behavioural and neural recordings *beyond* the task event data used to train the networks. Specifically, we examined behavioural response latencies, and dopamine dynamics recorded using fiber photometry from the medial frontal cortex (mFC) and nucleus accumbens (NAc) of the same mice performing the probabilistic reversal learning task (Fig. 6).

$h_1 + h_2$ **predicts initiation latency.** In natural foraging, mice must both adapt their action preferences and how quickly they take those actions. Classic theories of optimal foraging (Stephens et al., 2007) and average reward RL (Niv et al., 2007) connect this allocation of time to environmental reward rate. In reward-guided decision-making tasks, trial initiation latency provides a measure of motivation to work, or vigor. Given the sum of monoGRU's hidden units track recent reward rate, we tested whether $h_1 + h_2$ explained trial-by-trial fluctuations in how fast mice initiated trials. Indeed, even after accounting for the absolute difference, $|h_1 - h_2|$, regressions demonstrated that a higher sum of hidden units on the previous trial, $(h_{1,t-1} + h_{2,t-1})$, was associated with shorter latency to initiate (Fig. 6a).

$h_1 + h_2$ **and** $h_1 - h_2$ **explain prefrontal-striatal dopamine dynamics over different timescales.** To explore whether internal variables derived from monoGRU explain trial-by-trial variance in dopamine dynamics, we used linear regressions to model their influence across trials and time-

points (Appendix A). In running separate regression analyses for each timepoint, the time course of predictor loadings reflect the timing (e.g. phasic or tonic) and direction of variance in dopamine activity (Fig. 6b,c).

The phasic signalling of chosen action values has been observed in the ventral striatum as neuronal firing (Samejima et al., 2005) and dopamine release (Blanco-Pozo et al., 2024). We therefore examined whether monoGRU's chosen action preference - read out predominantly along $h_1 - h_2$ - explained event-related NAc dopamine fluctuations. The "chosen-logit" value had a significantly strong positive influence on NAc dopamine around the time of choice, followed by a slightly negative influence at outcome.

Given evidence PFC neurons maintain persistent representations of total expected value, $Q_l + Q_r$ (Bari et al., 2019), and dopamine concentrations in PFC covary with recent reward rate (Mohebi et al., 2019), we explored whether monoGRU's $h_1 + h_2$ - which tracks recent reward history - explains variance in PFC dopamine dynamics during behaviour. In contrast to the phasic dynamics observed in NAc, the loading for previous hidden-sum in mFC dopamine spanned the entire trial period - not only at decision or outcome events - reflecting slower, tonic dynamics.

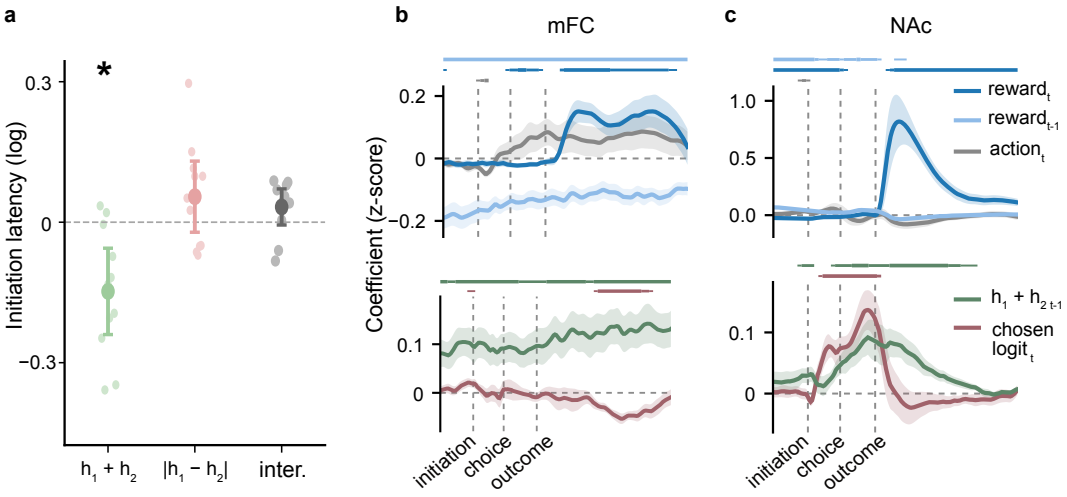



Figure 5: **monoGRU dynamics predict action latency and dopamine activity. a)** Regression predicting log latency to initiate trials as a function of monoGRU hidden unit activation sum, $h_1 + h_2$, absolute difference $|h_1 - h_2|$, and their interaction. Pale dots, individual mice (n=11); bold, across-subject mean and confidence intervals; $*$, two-sided t-test against zero, Benjamini–Hochberg-adjusted $P < 0.05$. **b, c)** Coefficients from the linear regression using monoGRU-derived trial-by-trial internal variables to predict variance in dopamine dynamics for each timepoint in the medial frontal cortex (mFC, n=9) and nucleus accumbens (NAc, n=10). Bold lines, across-subject means; shaded region, ±1 SEM across mice; dots above signals, coefficients different from zero at BH-adjusted $P < 0.05$, corrected across time separately for each predictor.

## 5 Discussion

In this work, we first characterised mechanistic variability in tiny RNNs for modelling choice behaviour, then introduced a framework for improving mechanistic reproducibility by applying biological constraints and simplifying the gating architecture to a single scalar update gate. This enabled us to identify a consistent and interpretable computational motif in our two-dimensional monoGRU, aligning cognitive and neural mechanisms: the hidden-state difference primarily represented action preference, and the hidden-state sum tracked recent reward history, explaining variance in behavioural response latencies and dopamine dynamics.

These results build on the predictive accuracy and interpretability of tiny RNNs for modelling decision-making data by dramatically improving the reproducibility of their internal mechanisms (Ji-An et al., 2025), thereby enabling examination. Our characterisation of mechanistic variability in

tiny RNNs relates to *solution degeneracy* in task-trained RNNs - where networks with similar training loss can have different internal representations and weights (Huang et al., 2025) - and *parameter ambiguity* in RL models fitted to behavioural data - where different combinations of parameters can yield similar predictive accuracy (Xiong et al., 2025). Our contribution bridges these literatures, characterising how and why mechanistic variability in hidden unit and gate activations become problematic in RNNs fitted to behavioural data. We then extended prior work by introducing constraints that limit degeneracy in hidden state representations and network-wide connectivity/weights.

The representational geometry of monoGRU, with behavioural features decodable via the sum and difference axis, may be relevant to understanding computations in prefrontal cortex (PFC) which support reversal learning behaviour (Cools et al., 2002; Clarke et al., 2004; Bari et al., 2019). Bari et al. (2019) show that instead of directly representing action values, $Q_l$ and $Q_r$, three populations of neurons PFC each align with relative and total value, $(Q_l - Q_r)$, $(Q_r - Q_l)$, and $(Q_l + Q_r)$. In their analysis they show that relative value was used to compute choice direction (Cai et al., 2011; Ito & Doya, 2009), and total value modulated choice vigor (Niv et al., 2007; Hamid et al., 2016). In monoGRU, we show that hidden units $h_1$ and $h_2$ which predict behaviour learn to approximate value signals, $Q_l$ and $Q_r$, using recurrent dynamics. This representation could be linearly decoded to recover representations similar to PFC. Our analyses of hidden states on behavioural and neural data complement the analysis of Bari et al. (2019): action-preference is read out along $h_1 - h_2$, whereas $h_1 + h_2$ tracks recent reward history, and is associated with subsequent trial-initiation latency, a behavioural measure of vigor. Future analyses could directly compare PFC electrophysiology data to monoGRU-derived features of relative and total value.

Inspired by principles of neuromodulatory gain modulation in the cortex (Ferguson & Cardin, 2020; Shine et al., 2021), the scalar update gate in monoGRU also offers a more biologically plausible implementation of gating in RNNs (Soo et al., 2023). Gating enables the *effective* weight of a connection to be modulated as a function of network state, flexibly reconfiguring circuit dynamics without changing synaptic weights (Costacurta et al., 2024; Foucault & Meyniel, 2021). In monoGRU, the previous-trial hidden sum can do this via its effect on the shared gate's activation, thereby modulating recurrent dynamics. In the brain, neuromodulators such as dopamine innervate the PFC and bind to receptors to change the gain of synapses' input-output transfer function, thereby shifting population activity dynamics (Ferguson & Cardin, 2020; Shine et al., 2021; Froudist-Walsh et al., 2021; Ott & Nieder, 2019). Taken together, this may support a computational mechanism whereby dopamine could support activity-based computations in recurrent PFC-based circuits via signalling of recent reward history (Wang et al., 2018).

By modelling choices and model discrete updates, $t$, at the level of each trial (as in Ji-An et al., 2025), our work is limited to computations that persist across trials, rather than computations within each trial. We could therefore extend our work to model discrete steps in the task, aligning the discrete updates $t$ with the order of experience: initiation cue $\rightarrow$ initiation action $\rightarrow$ free/forced choice cue $\rightarrow$ choice action. This could help us understand the computational mechanisms required to preserve reward histories across trials, while also switching between action contexts (initiation vs choice). Context-sensitive mechanisms would also be required to solve 2-step tasks, where the outcome of a second choice should inform the next first choice (Blanco-Pozo et al., 2024).

By ligning behavioural models of tasks with neural computations, under biological constraints,

## AI USE STATEMENT

In this work, we used generative AI tools to [implement methods, provide feedback on research methodology or experiments, review of writing, and LaTeX formatting,] We have not used generative AI tools to [help develop theoretical models or conceptual frameworks, formulate mathematical claims, propose or refine hypotheses, interpret results, clean and reformat dataset], and [assist with translation, generate synthetic data sets, assist in the writing of proofs, provide critical ingredients for proving mathematical claims, support qualitative and thematic data analysis], are not applicable to this work.

Additionally, we used generative AI tools for [create or modify scientific figures or images, create or edit software code, summarize or analyse existing literature, sourcing/searching for infor- mation, edit a research paper to improve readability, identify relevant literature]. We have reviewed all AI-assisted work. All generated code was reviewed and tested for correctness by at least one Author. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

## ETHICS STATEMENT

All experimental procedures were performed in line with the UK Animal (Scientific Procedure) Act 1986 and in accordance with the University of Oxford animal use guidelines. They were approved by the local ethical review panel at the Department of Experimental Psychology, University of Oxford, and performed under UK Home Office Project Licence PP6834709.

## REPRODUCIBILITY STATEMENT

Further methodological details experiments and analyses have been included in Appendix A. All code and behavioural data required to reproduce RNN results will be made available upon publication. Neural data of dopamine will not be made public.

## AUTHOR CONTRIBUTIONS

[Anonymised for review]

## ACKNOWLEDGMENTS

[Anonymised for review]

## REFERENCES

Samuel K. Ainsworth, Jonathan Hayase, and Siddhartha Srinivasa. Git Re-Basin: Merging Models modulo Permutation Symmetries, March 2023.

Bilal A. Bari, Cooper D. Grossman, Emily E. Lubin, Adithya E. Rajagopalan, Jianna I. Cressy, and Jeremiah Y. Cohen. Stable Representations of Decision Variables for Flexible Behavior. *Neuron*, 103(5):922–933.e7, September 2019. ISSN 0896-6273. doi: 10.1016/j.neuron.2019.06.001.

Marta Blanco-Pozo, Thomas Akam, and Mark E. Walton. Dopamine-independent effect of rewards on choices through hidden-state inference. *Nature Neuroscience*, January 2024. ISSN 1097-6256, 1546-1726. doi: 10.1038/s41593-023-01542-x.

Xinying Cai, Soyoun Kim, and Daeyeol Lee. Heterogeneous Coding of Temporally Discounted Values in the Dorsal and Ventral Striatum during Intertemporal Choice. *Neuron*, 69(1):170–182, January 2011. ISSN 0896-6273. doi: 10.1016/j.neuron.2010.11.041.

Pablo Samuel Castro, Nenad Tomasev, Ankit Anand, Navodita Sharma, Rishika Mohanta, Aparna Dev, Kuba Perlin, Siddhant Jain, Kyle Levin, Noémi Éltető, Will Dabney, Alexander Novikov, Glenn C Turner, Maria K Eckstein, Nathaniel D Daw, Kevin J Miller, and Kimberly L Stachenfeld. Discovering Symbolic Cognitive Models from Human and Animal Behavior, February 2025.

An Mei Chen, Haw-minn Lu, and Robert Hecht-Nielsen. On the Geometry of Feedforward Neural Network Error Surfaces. *Neural Computation*, 5(6):910–927, November 1993. ISSN 0899-7667, 1530-888X. doi: 10.1162/neco.1993.5.6.910.

Junyoung Chung, Caglar Gulcehre, KyungHyun Cho, and Yoshua Bengio. Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling, December 2014.

H. F. Clarke, J. W. Dalley, H. S. Crofts, T. W. Robbins, and A. C. Roberts. Cognitive inflexibility after prefrontal serotonin depletion. *Science (New York, N.Y.)*, 304(5672):878–880, May 2004. ISSN 1095-9203. doi: 10.1126/science.1094987.

Jasmine Collins, Jascha Sohl-Dickstein, and David Sussillo. Capacity and Trainability in Recurrent Neural Networks, March 2017.

Roshan Cools, Luke Clark, Adrian M. Owen, and Trevor W. Robbins. Defining the neural mechanisms of probabilistic reversal learning using event-related functional magnetic resonance imaging. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 22(11): 4563–4567, June 2002. ISSN 1529-2401. doi: 10.1523/JNEUROSCI.22-11-04563.2002.

Julia C. Costacurta, Shaunak Bhandarkar, David M. Zoltowski, and Scott W. Linderman. Structured flexibility in recurrent neural networks via neuromodulation, July 2024.

Maria K Eckstein, Linda Wilbrecht, and Anne Ge Collins. What do reinforcement learning models measure? Interpreting model parameters in cognition and neuroscience. *Current Opinion in Behavioral Sciences*, 41:128–137, October 2021. ISSN 23521546. doi: 10.1016/j.cobeha.2021.06.004.

Maria K. Eckstein, Christopher Summerfield, Nathaniel D. Daw, and Kevin J. Miller. Predictive and Interpretable: Combining Artificial Neural Networks and Classic Cognitive Models to Understand Human Learning and Decision Making, May 2023.

Maria K. Eckstein, Christopher Summerfield, Nathaniel D. Daw, and Kevin J. Miller. Hybrid neural–cognitive models reveal how memory shapes human reward learning. *Nature Human Behaviour*, pp. 1–16, February 2026. ISSN 2397-3374. doi: 10.1038/s41562-025-02324-0.

Katie A. Ferguson and Jessica A. Cardin. Mechanisms underlying gain modulation in the cortex. *Nature Reviews Neuroscience*, 21(2):80–92, February 2020. ISSN 1471-003X, 1471-0048. doi: 10.1038/s41583-019-0253-y.

Cédric Foucault and Florent Meyniel. Gated recurrence enables simple and accurate sequence prediction in stochastic, changing, and structured environments. *eLife*, 10:e71801, December 2021. ISSN 2050-084X. doi: 10.7554/eLife.71801.

Sean Froudist-Walsh, Daniel P. Bliss, Xingyu Ding, Lucija Rapan, Meiqi Niu, Kenneth Knoblauch, Karl Zilles, Henry Kennedy, Nicola Palomero-Gallagher, and Xiao-Jing Wang. A dopamine gradient controls access to distributed working memory in the large-scale monkey cortex. *Neuron*, 109(21):3500–3520.e13, November 2021. ISSN 08966273. doi: 10.1016/j.neuron.2021.08.024.

Charles Godfrey, Davis Brown, Tegan Emerson, and Henry Kvinge. On the Symmetries of Deep Learning Models and their Internal Representations, March 2023.

Arif A Hamid, Jeffrey R Pettibone, Omar S Mabrouk, Vaughn L Hetrick, Robert Schmidt, Caitlin M Vander Weele, Robert T Kennedy, Brandon J Aragona, and Joshua D Berke. Mesolimbic dopamine signals the value of work. *Nature Neuroscience*, 19(1):117–126, January 2016. ISSN 1097-6256, 1546-1726. doi: 10.1038/nn.4173.

Ann Huang, Satpreet H. Singh, Flavio Martinelli, and Kanaka Rajan. Measuring and Controlling Solution Degeneracy across Task-Trained Recurrent Neural Networks, November 2025.

Hidehiko K. Inagaki, Lorenzo Fontolan, Sandro Romani, and Karel Svoboda. Discrete attractor dynamics underlies persistent activity in the frontal cortex. *Nature*, 566(7743):212–217, February 2019. ISSN 0028-0836, 1476-4687. doi: 10.1038/s41586-019-0919-7.

Makoto Ito and Kenji Doya. Validation of Decision-Making Models and Analysis of Decision Variables in the Rat Basal Ganglia. *Journal of Neuroscience*, 29(31):9861–9874, August 2009. ISSN 0270-6474, 1529-2401. doi: 10.1523/JNEUROSCI.6157-08.2009.

Eugene M Izhikevich. Solving the distal reward problem through linkage of STDP and dopamine signaling. *BMC Neuroscience*, 8(S2):S15, July 2007. ISSN 1471-2202. doi: 10.1186/1471-2202-8-S2-S15.

Herbert Jaeger, Mantas Lukoševičius, Dan Popovici, and Udo Siewert. Optimization and applications of echo state networks with leaky- integrator neurons. *Neural Networks*, 20(3):335–352, April 2007. ISSN 08936080. doi: 10.1016/j.neunet.2007.04.016.

Li Ji-An, Marcus K. Benna, and Marcelo G. Mattar. Discovering Cognitive Strategies with Tiny Recurrent Neural Networks, July 2025.

Daniel Kasenberg, Pablo Samuel Castro, Maria K Eckstein, Noemi Elteto, Will Dabney, Caroline Wang, Martin Engelcke, Rishika Mohanta, Aparna Dev, Matthew M Botvinick, Nenad Tomasev, Glenn C Turner, Vincent Costa, D Daw, Kimberly L Stachenfeld, and Kevin J Miller. AI-Discovered Cognitive Models Reveal Novel Insights into Human and Animal Learning. May 2026.

Kentaro Katahira. The relation between reinforcement learning parameters and the influence of reinforcement history on choice behavior. *Journal of Mathematical Psychology*, 66:59–69, June 2015. ISSN 00222496. doi: 10.1016/j.jmp.2015.03.006.

Lorenzo Livi. Time-Scale Coupling Between States and Parameters in Recurrent Neural Networks, August 2025.

Ilya Loshchilov and Frank Hutter. Decoupled Weight Decay Regularization, January 2019.

Niru Maheswaranathan, Alex H. Williams, Matthew D. Golub, Surya Ganguli, and David Sussillo. Universality and individuality in neural dynamics across large populations of recurrent networks, December 2019.

Kevin J. Miller, Matthew M. Botvinick, and Carlos D. Brody. From predictive models to cognitive models: Separable behavioral processes underlying reward learning in the rat, February 2021.

Kevin J. Miller, Maria Eckstein, Matthew M. Botvinick, and Zeb Kurth-Nelson. Cognitive Model Discovery via Disentangled RNNs, June 2023.

Karyna Mishchanchuk, Gabrielle Gregoriou, Albert Qü, Alizée Kastler, Quentin J. M. Huys, Linda Wilbrecht, and Andrew F. MacAskill. Hidden state inference requires abstract contextual representations in the ventral hippocampus. *Science*, 386(6724):926–932, November 2024. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.adq5874.

Ali Mohebi, Jeffrey R. Pettibone, Arif A. Hamid, Jenny-Marie T. Wong, Leah T. Vinson, Tommaso Patriarchi, Lin Tian, Robert T. Kennedy, and Joshua D. Berke. Dissociable dopamine dynamics for learning and motivation. *Nature*, 570(7759):65–70, June 2019. ISSN 0028-0836, 1476-4687. doi: 10.1038/s41586-019-1235-y.

Yael Niv. Reinforcement learning in the brain. *Journal of Mathematical Psychology*, 53(3):139–154, June 2009. ISSN 00222496. doi: 10.1016/j.jmp.2008.12.005.

Yael Niv, Nathaniel D. Daw, Daphna Joel, and Peter Dayan. Tonic dopamine: Opportunity costs and the control of response vigor. *Psychopharmacology*, 191(3):507–520, April 2007. ISSN 0033-3158. doi: 10.1007/s00213-006-0502-4.

Torben Ott and Andreas Nieder. Dopamine and Cognitive Control in Prefrontal Cortex. *Trends in Cognitive Sciences*, 23(3):213–234, March 2019. ISSN 1364-6613. doi: 10.1016/j.tics.2018.12.006.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Köpf, Edward Yang, Zach DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. PyTorch: An Imperative Style, High-Performance Deep Learning Library, December 2019.

Albert J. Qü, Lung-Hao Tai, Christopher D. Hall, Emilie M. Tu, Maria K. Eckstein, Karyna Mishchanchuk, Wan Chen Lin, Juliana B. Chase, Andrew F. MacAskill, Anne G. E. Collins, Samuel J. Gershman, and Linda Wilbrecht. Nucleus accumbens dopamine release reflects Bayesian inference during instrumental learning. *PLOS Computational Biology*, 21(7):e1013226, July 2025. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1013226.

Mirco Ravanelli, Philemon Brakel, Maurizio Omologo, and Yoshua Bengio. Light Gated Recurrent Units for Speech Recognition. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 2(2):92–102, April 2018. ISSN 2471-285X. doi: 10.1109/TETCI.2017.2762739.

Kazuyuki Samejima, Yasumasa Ueda, Kenji Doya, and Minoru Kimura. Representation of Action-Specific Reward Values in the Striatum. *Science*, 310(5752):1337–1340, November 2005. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1115270.

Jessica V. Schaaf, Laura Weidinger, Lucas Molleman, and Wouter Van Den Bos. Test–retest reliability of reinforcement learning parameters. *Behavior Research Methods*, 56(5):4582–4599, September 2023. ISSN 1554-3528. doi: 10.3758/s13428-023-02203-4.

James M. Shine, Eli J. Müller, Brandon Munn, Joana Cabral, Rosalyn J. Moran, and Michael Breakspear. Computational models link cellular mechanisms of neuromodulation to large-scale neural dynamics. *Nature Neuroscience*, 24(6):765–776, June 2021. ISSN 1546-1726. doi: 10.1038/s41593-021-00824-6.

Wayne W.M. Soo, Vishwa Goudar, and Xiao-Jing Wang. Training biologically plausible recurrent neural networks on cognitive tasks with long-term dependencies, October 2023.

David W. Stephens, Joel S. Brown, and Ronald C. Ydenberg (eds.). *Foraging: Behavior and Ecology*. University of Chicago Press, Chicago, IL, September 2007. ISBN 978-0-226-77264-6.

Christopher Summerfield and Kevin Miller. Computational and systems neuroscience: The next 20 years. *PLOS Biology*, 21(9):e3002306, September 2023. ISSN 1545-7885. doi: 10.1371/journal.pbio.3002306.

Corentin Tallec and Yann Ollivier. Can recurrent neural networks warp time?, March 2018.

Elia Turner, Kabir Dabholkar, and Omri Barak. Charting and navigating the space of solutions for recurrent neural networks, November 2021.

Jane X. Wang, Zeb Kurth-Nelson, Dharshan Kumaran, Dhruva Tirumala, Hubert Soyer, Joel Z. Leibo, Demis Hassabis, and Matthew Botvinick. Prefrontal cortex as a meta-reinforcement learning system. *Nature Neuroscience*, 21(6):860–868, June 2018. ISSN 1097-6256, 1546-1726. doi: 10.1038/s41593-018-0147-8.

James C. R. Whittington and William Dorrell. How much neuroscience does a neuroscientist need to know?, January 2026.

James C. R. Whittington, Will Dorrell, Surya Ganguli, and Timothy E. J. Behrens. Disentanglement with Biological Constraints: A Theory of Functional Cell Types, March 2023.

Hua-Dong Xiong, Li Ji-An, Marcelo G. Mattar, and Robert C. Wilson. Deep Learning Improves Parameter Estimation in Reinforcement Learning Models, March 2025.