Simultaneous modeling of behavior and dopamine with disentangled RNNs

Siddhant Jain (siddhantjain@deepmind.com)
Google DeepMind
Mountain View, CA, USA

Nathaniel D Daw (ndaw@deepmind.com)
Google DeepMind and Princeton University
Princeton, NJ, USA

Kimberly L Stachenfeld (stachenfeld@deepmind.com)
Google DeepMind and Columbia University
New York, NY, USA

Kevin J Miller (kevinjmiller@deepmind.com)
Google DeepMind and University College London
London, United Kingdom

Abstract

Understanding how neural activity relates to cognitive processes during learning requires methods that can jointly model brain signals and behavior. Here, we extend Disentangled RNNs (DisRNN), an approach for using constrained recurrent neural networks to discover cognitive models, to the case of jointly modeling behavioral data and measurements of neural activity. We augment a DisRNN trained on choice prediction with a separate subnetwork to predict neural activity. We apply this approach to datasets from a simple reward-learning task, consisting of choices, rewards, and a scalar measure of dopamine responses to reward. First, using synthetic data from a Q-learning agent, we demonstrate the approach is able to capture both choices and reward prediction errors with a single set of internal variables, consistent with the groundtruth. Next, we apply this approach to laboratory data from mice performing a similar task, successfully modeling both choice behavior and nucleus accumbens dopamine responses. Analysis of the fit DisRNNs confirms that the same interpretable latent variables are utilized for both choice prediction and dopamine signal prediction, demonstrating the model's potential to uncover cognitive models that bridge behavior and neural data through shared representations.

Keywords: reinforcement learning; data-driven modeling;

Introduction

Understanding the neural mechanisms of reward-guided learning behavior is a long-standing goal of cognitive neuroscience. A very common workflow in computational cognitive neuroscience is to fit computational models to behavior, then to treat these models as hypotheses about the neural mechanisms that underlie that behavior. These hypotheses can then be tested by comparing them to measurements of neural activity (O'Doherty, Hampton, & Kim, 2007; Daw et al., 2011). Another possibility is to fit models jointly to behavior and neural activity, allowing each modality to constrain the interpretation of the other (Dezfouli, Morris, Ramos, Dayan, & Balleine, 2018; Dommanget-Kott et al., 2024). This approach is much less widely-used, perhaps because fitting neural data typically requires highly flexible models, which are often difficult to interpret as cognitive hypotheses.

Disentangled RNNs (DisRNNs) are a recently-developed method for automatically learning interpretable cognitive models from data (Miller, Eckstein, Botvinick, & Kurth-Nelson, 2023). While standard RNNs often learn complex, high-dimensional internal representations that are difficult to map onto specific cognitive processes, DisRNNs employ architectural constraints as well as information bottlenecks which impose a cost on information flow in order to encourage sparse latent representations where key variables capture distinct elements of the underlying cognitive process.

In this work, we extend DisRNNs to simultaneously fit both choice behavior and measurements of neural activity. We apply this approach to synthetic and laboratory datasets of both choices and dopamine responses in a classic reinforcement learning task. We first validate our approach using synthetic data generated by a known cognitive model (Q-Learning), demonstrating that our method successfully recovers the key structure and variables of the ground-truth model within interpretable latent variables (Q-values and reward prediction error simulated as dopamine). Next, we apply the framework to laboratory data from mice performing a two-armed bandit task with concurrent dopamine recordings (Parker et al., 2016). We find that the model can simultaneously fit behavioral choices and neural activity, using the same underlying latent representations. This indicates that DisRNN, fit jointly in this way, is able to discover from data computational models that bridge behavior and neural activity via shared, interpretable computations.

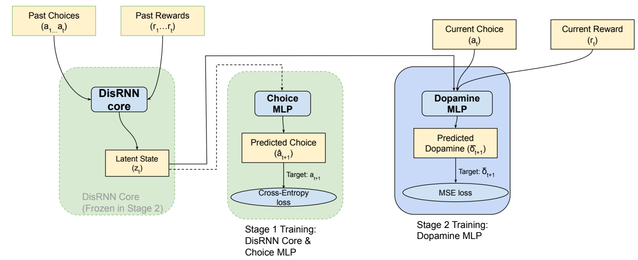


Figure 1: The DisRNN-Dopamine model architecture. A core DisRNN predicts choice (trained in Stage 1), while a separate neural network ("Dopamine MLP") predicts dopamine responses using the DisRNN latent state and information about the current trial (trained in Stage 2).

Methods

Our network design (illustrated in Figure 1) builds upon the Disentangled RNN (DisRNN) proposed by (Miller et al., 2023). The core DisRNN is trained to predict choices. We augment it with an additional neural network ("Dopamine MLP") head that takes as input the DisRNN's latent state along with the current choice and reward, and is trained to predict a measure of dopamine response.

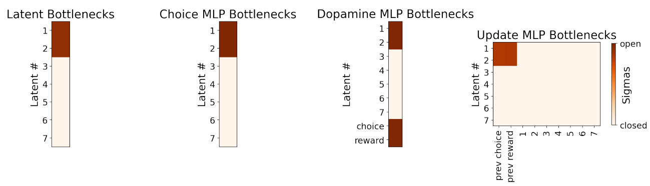
We employ a two-stage training procedure. Initially, we train only the parameters of the core DisRNN component and choice MLP to predict the agent's choices (a_{t+1}) based on past choices and rewards. This stage uses a standard cross-entropy loss function. After the first stage, we freeze the parameters of the core DisRNN and train *only* the parameters of the newly added dopamine MLP readout. This stage uses the dopamine values as the target, minimizing a Mean Squared Error (MSE) loss between these and the MLP's output. For training the dopamine MLP readout, we open up all the latent bottlenecks which might have been closed in core DisRNN training.

This sequential training approach first allows the DisRNN to learn latent dynamics relevant to behavior, and then trains a separate readout to map these learned dynamics onto the associated dopamine signal.

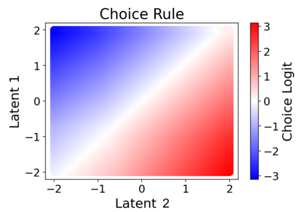
Results

Synthetic Data

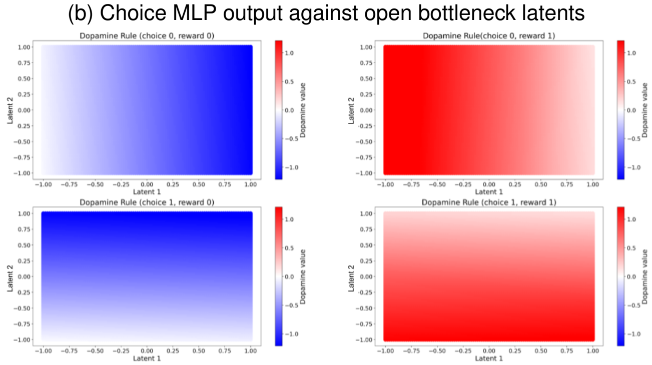
We simulate behavioral and neural data using a Q-Learning agent performing a two-armed bandit task with drifting reward probabilities similar to (Miller et al., 2023). As a proxy for dopamine signal, we compute the reward prediction error (RPE) on each trial t as the difference between the received reward r_t and the value of the chosen action $Q_t(a_t)$. The objective



(a) Sigma parameters visualized for the bottlenecks. Darker colors indicate open bottlenecks



(b) Choice MLP output against open bottleneck latents



(c) Dopamine MLP output against bottleneck latents for different combination of choice and reward.

Figure 2: DisRNN-Dopamine recovers latent dynamics of Q-Learning.

for our network is to jointly predict the agent's trial-by-trial choices *and* the corresponding RPE_t. We analyzed the internal representations learned by the disRNN trained on the synthetic Q-learning dataset. After training, only two latents exhibited open bottlenecks, indicating their active use by the network (Figure 2a). These latents show a pattern consistent with representation of the Q-values associated with the two actions.

For the separately trained dopamine MLP we observe that the *same* two latent variables have open bottlenecks (Figure 2a, Dopamine MLP Bottlenecks), indicating that the dopamine prediction mechanism leverages the learned Q-value representations. The predicted dopamine response is contingent on both the reward and the choice made (Figure 2c) and consistent with a representation of reward prediction error signal.

Laboratory Dataset

We next evaluated our networks on a dataset collected from mice performing a similar two-armed bandit task while measurements of dopamine neuron activity in the nucleus acucmbens were recorded using fiber photometry (Parker et al., 2016).

We trained the DisRNN model using the two-stage procedure described in Figure 1 and compare the performance

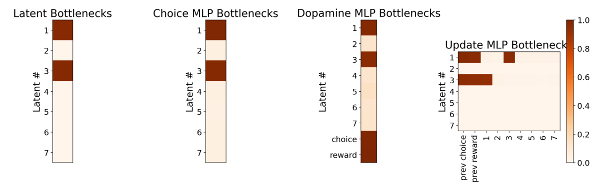


Figure 3: DisRNN bottleneck analysis on experimental data from Parker et al. (2016). Similar to synthetic data, latents used for choice are also utilized by the dopamine prediction head.

against a Differential Forgetting Q Learning (DFQ) model (Ito & Doya, 2009) along with a naive baseline with no memory of trial history in Table 1.

Model	Choice NLL ↑	Dopamine MSE ↓
DFQ (Choice Only)	**0.6168**	2.9351
DFQ (DA only)	0.5	1.6721
Naive / Chance	0.5	2.2181
DisRNN (Ours)	0.6023	**1.6553**

Table 1: Results from fitting our model on (Parker et al., 2016) dataset. DFQ (choice only) is trained with choice and hence doesn't perform so well on dopamine. DFQ (DA only) is trained only with dopamine. The naive baseline for dopamine is simply the average rewarded and unrewarded dopamine recordings from the train set. DisRNN provides a unified model for predicting both choice and dopamine simultaneously. It has competitive results on the normalized log likelihood (NLL) for choice and outperforms the baselines for the mean squared error (MSE) for dopamine.

Crucially, analysis of the internal representations learned by the DisRNN from the experimental data mirrors the findings from our synthetic experiments. Figure 3 visualizes the bottleneck activations for the core DisRNN latents, the Choice MLP readout, and the Dopamine MLP readout. Consistent with the synthetic results (Figure 2), we observe sparse activation, with specific latents being utilized by the choice prediction mechanism (Figure 3, Choice MLP Bottlenecks). Importantly, the *same* subset of latent variables with open bottlenecks for choice prediction also exhibit open bottlenecks for the Dopamine MLP (Figure 3, Dopamine MLP Bottlenecks). This suggests that the DisRNN framework learns behaviorally relevant representations that are subsequently read out to predict neural activity, supporting the model's interpretability and its potential for uncovering shared neural and behavioral computations.

References

Daw, N. D., et al. (2011). Trial-by-trial data analysis using computational models. *Decision making, affect, and learning: Attention and performance XXIII*, *23*(1), 3–38.

Dezfouli, A., Morris, R., Ramos, F. T., Dayan, P., & Balleine, B. (2018). Integrated accounts of behavioral and neuroimaging data using flexible recurrent neural network models. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, & R. Garnett (Eds.), *Advances in neural information processing systems* (Vol. 31). Curran Associates, Inc.

Dommanget-Kott, M., Fernandez-de Cossio-Diaz, J., Coraggioso, M., Bormuth, V., Monasson, R., Debrégeas, G., & Cocco, S. (2024). Linking brain and behavior states in zebrafish larvae locomotion using hidden markov models. *bioRxiv*.

Ito, M., & Doya, K. (2009). Validation of decision-making models and analysis of decision variables in the rat basal ganglia. *Journal of Neuroscience*, *29*(31), 9861–9874.

Miller, K. J., Eckstein, M., Botvinick, M., & Kurth-Nelson, Z. (2023). Cognitive model discovery via disentangled rnns. In *Advances in neural information processing systems.*

O'Doherty, J. P., Hampton, A., & Kim, H. (2007). Model-based fMRI and its application to reward learning and decision making. *Annals of the New York Academy of Sciences*, *1104*, 35–53.

Parker, N. F., Cameron, C. M., Taliaferro, J. P., Lee, J., Choi, J. Y., Davidson, T. J., ... Witten, I. B. (2016, June). Reward and choice encoding in terminals of midbrain dopamine neurons depends on striatal target. *Nature Neuroscience*, *19*(6), 845–854. doi: 10.1038/nn.4287