Striatal reinforcement learning modeling predicts perseveration in a two-arm bandit task

Vijeeth Guggilla[1,3], Eric Yttri[2,3] - [1]Grinnell College, [2]Carnegie Mellon University, [3]Center for the Neural Basis of Cognition

BACKGROUND

Reversal learning is a crucial method of adapting to one's environment and is often studied by using two-arm bandit tasks. These tasks can model dynamic environments by shifting reward probabilities. Performance in these tasks relies on striatal direct (D) and indirect (I) pathway activity, which contributes to action selection and reinforcement.

QUESTION

Dual opponent actor reinforcement learning (OpAL) can model such pathways in a novel manner to further explore reversal learning. However, over long blocks in dynamic environments, the model generates an inexplicable performance decay. **How can this model be modified to ameliorate such artifacts, and what insight do these modifications provide? What do the changes imply when extrapolated to a biological context?**

SOLUTION

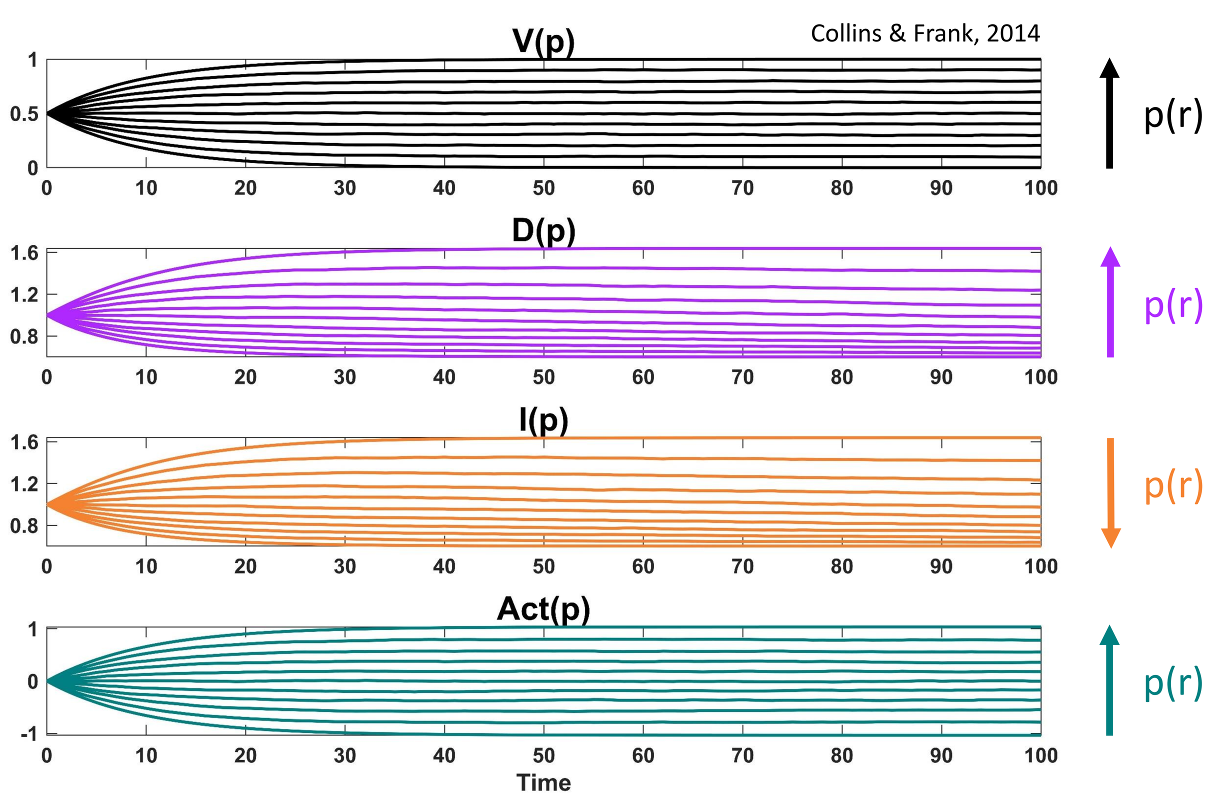
Discounting the learning rate for direct pathway in trials with negative reward predictions erases long-term task independent performance decay. Accordingly, implementing this change reveals perseveration in reversal learning tasks with long blocks. The perseveration observed in such contexts is relevant to biological systems in explaining fixation behavior in terms of giving less weight to learning from negative reward prediction.

APPLICATION

Exploring reversal learning task performance in animals is hypothesized to reveal similar perseverative behavior and elucidate the striatal dopamine dependency of such actions. A Y-maze task with mice was used to generate preliminary data in a reversal learning task with long blocks. Dopamine dependency of this task was studied via direct and indirect pathway inhibitors.

1

Opponent Process Actor Learning (OpAL) is a reinforcement learning model based on striatal direct (D) and indirect (I) activity. It separates actor and critic learning to simultaneously account for incentive and learning effects of dopamine.



Critic Learning

$$V(t+1) = V(t) + \alpha_C \times \delta(t)$$

$$\delta(t) = r(t) - V(t)$$

V = critic weight
α_C = critic learning rate
δ = prediction error
r = reinforcement received

Actor Learning

$$D_a(t+1) = D_a(t) + [\alpha_D D_a(t)] \times \delta(t)$$

$$I_a(t+1) = I_a(t) + [\alpha_I I_a(t)] \times [-\delta(t)]$$

D = direct pathway weight
I = indirect pathway weight
α_D = D learning rate
α_I = I learning rate

Policy

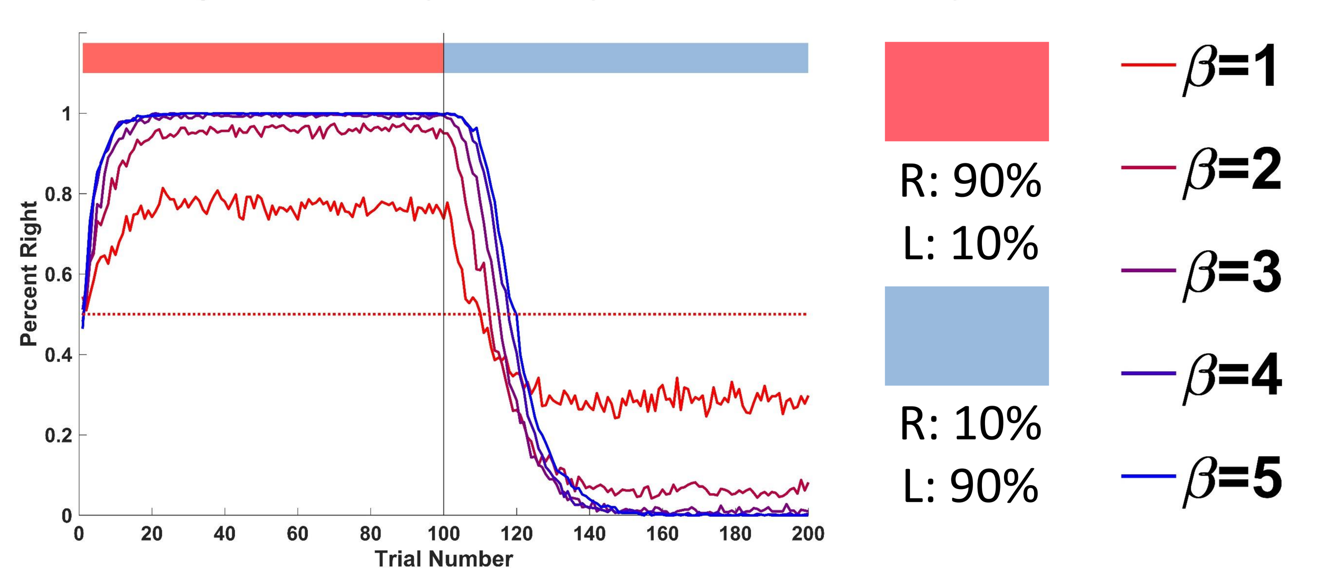
$$Act_a(t) = \beta_D D_a(t) - \beta_I I_a(t)$$

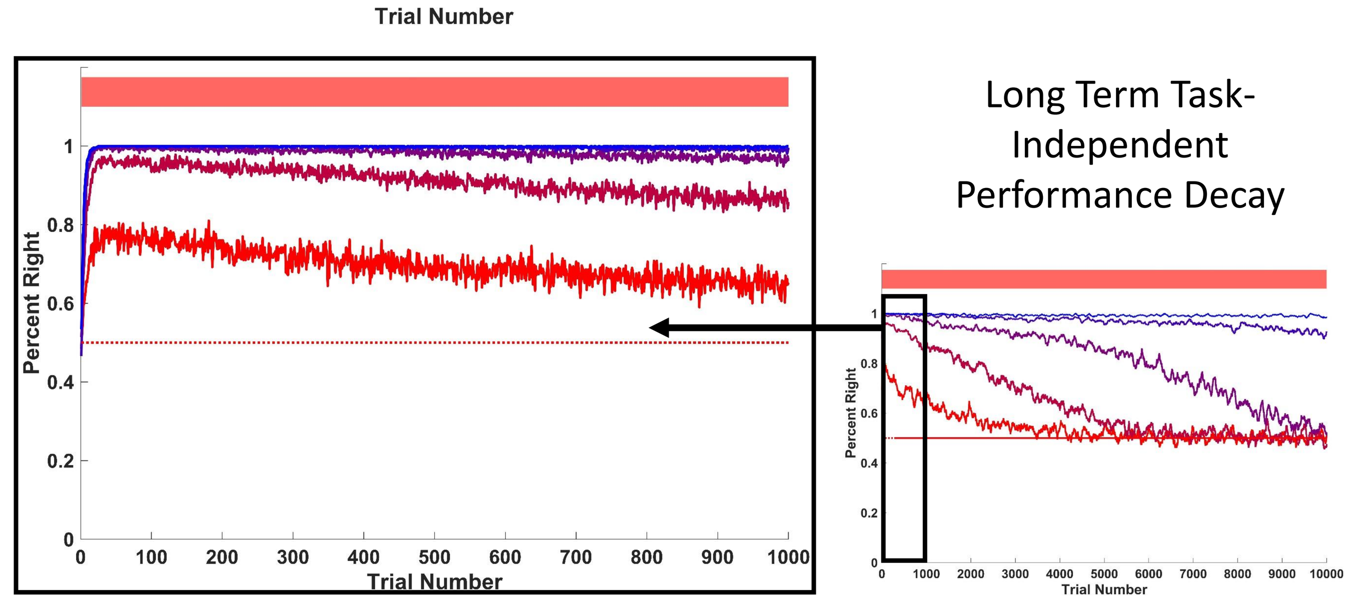
$$p(a) = \frac{e^{Act_a(t)}}{\sum_i e^{Act_{a_i}(t)}}$$

Act_a = combined actor weight
p = probability of choosing action a
β_D = D dopamine, β_I = I dopamine

2

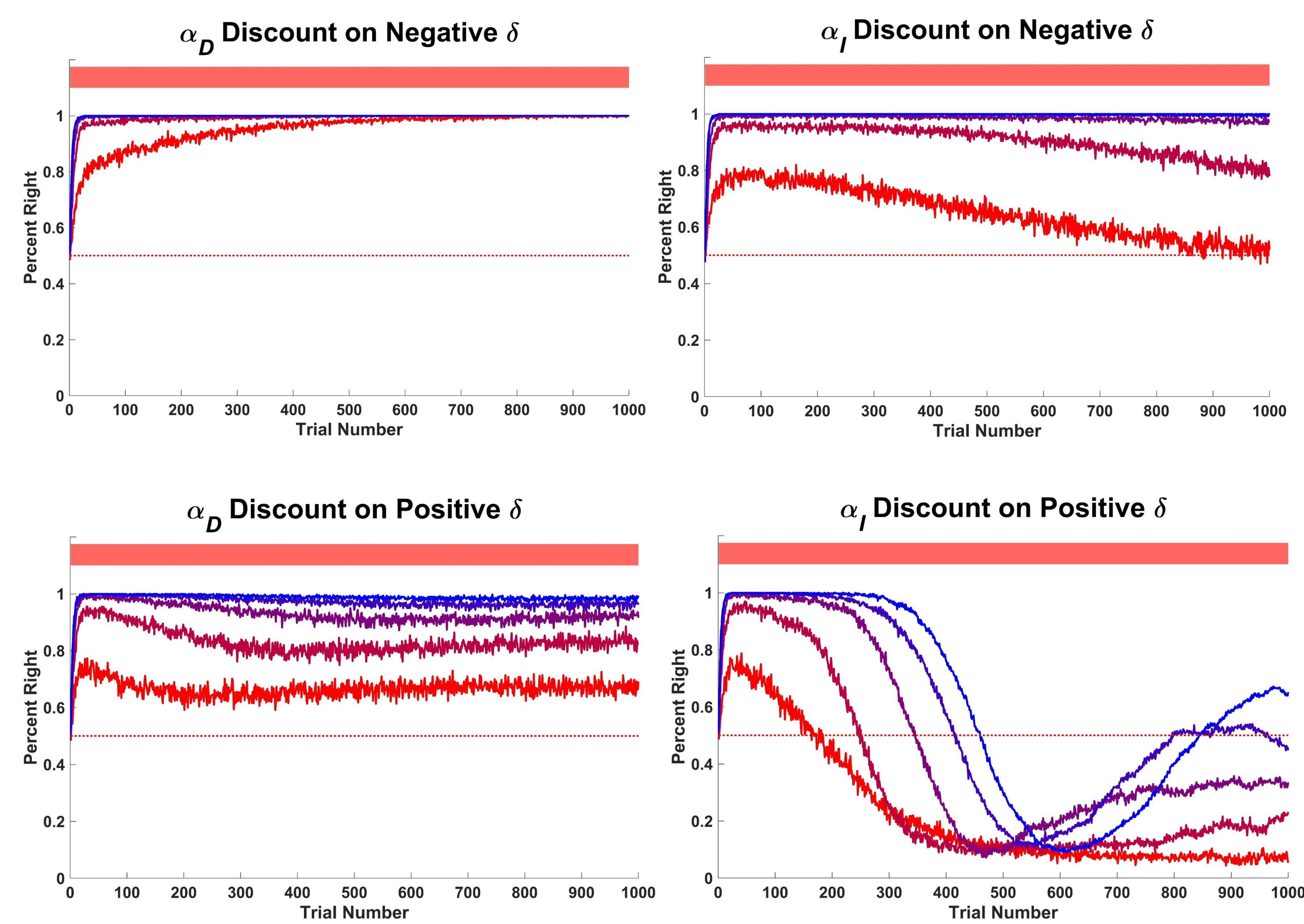
Reversal learning tasks appear to be effectively modeled by OpAL upon first glance, showing the expected relationship between learning and dopamine levels. However, upon extending the task block length, an inexplicable performance decay is observed.





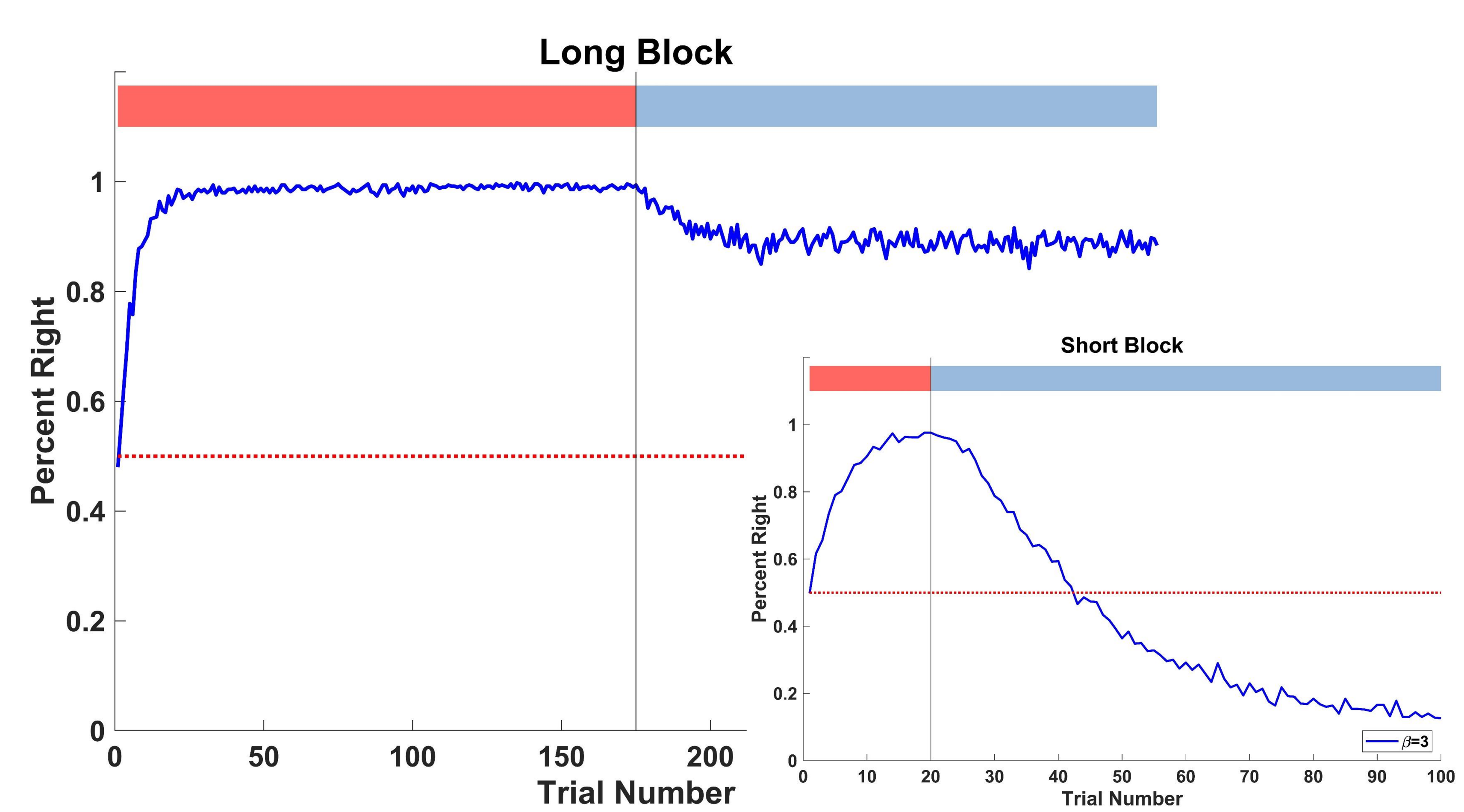
3

Discounting the D learning rate for all trials with a negative prediction error ameliorates the performance decay. Changing other learning rates under different conditions does not solve the problem, making the solution specific to placing less emphasis on direct learning from negative reward predictions.



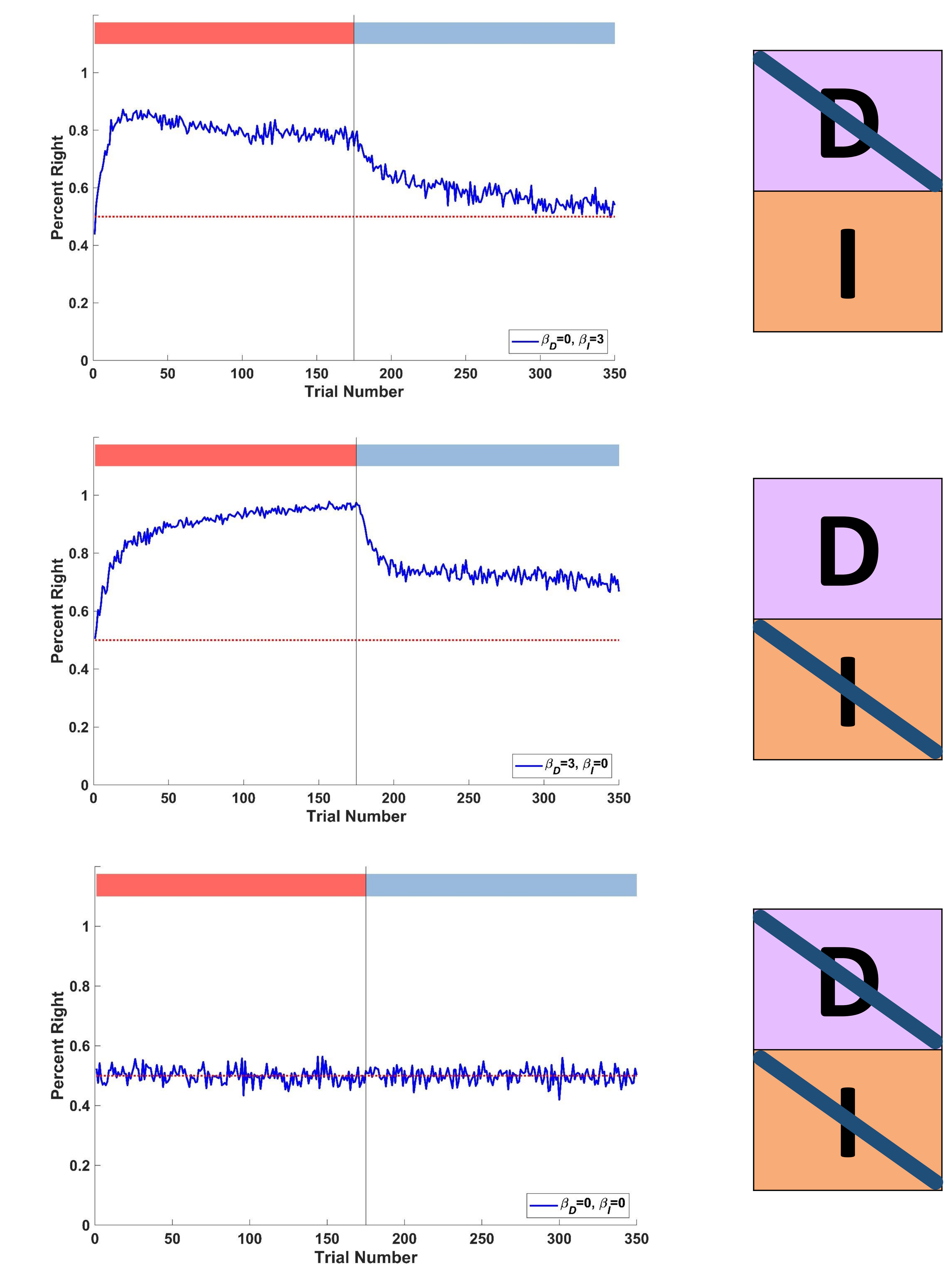
4

Lower integration of negative prediction errors in the D weight caused the model to predict perseverative behaviors for long blocks. This correlates with the biological context in explaining how fixative behavior can occur due to an impaired ability to directly process the lack of gain from an action.



5

Asymmetrically modifying β_D and β_I to reflect the influence of direct and indirect pathway inhibition on action selection affects the observed perseverative behaviors.



6

Mice performed a Y-maze task to demonstrate reversal learning for comparisons with the model. D (SCH-23390) and I (raclopride) inhibitors were used to study the effects of direct and indirect pathway inhibition.