A biologically and behaviorally inspired RL model: Mean Shift with Homeostasis (MeSH)

Claudia Li [1,4], **Syan Timothy Lopez [2,4]**, Eric Yttri[3,4] – [1]Purdue University, [2]University of California, Berkeley, [3]Carnegie Mellon University, [4]Center for the Neural Basis of Cognition

BACKGROUND

Reinforcement learning is one method that organisms use to alter their performance by changing their behavior in response to reward. For instance, they can learn to modify movement amplitude in order to achieve success in a variable amplitude operant (VAO) task. The basal ganglia have been shown to facilitate reinforcement learning through two opponent pathways, direct and indirect.

QUESTION

We initially generated a model based upon the results of a VAO mouse reaching experiment (Yttri and Dudman 2016) and sought to answer the following questions:
How well can the model perform, how can we improve its performance, and can we break up the learning component to more accurately represent the direct and indirect basal ganglia pathways?

SOLUTION

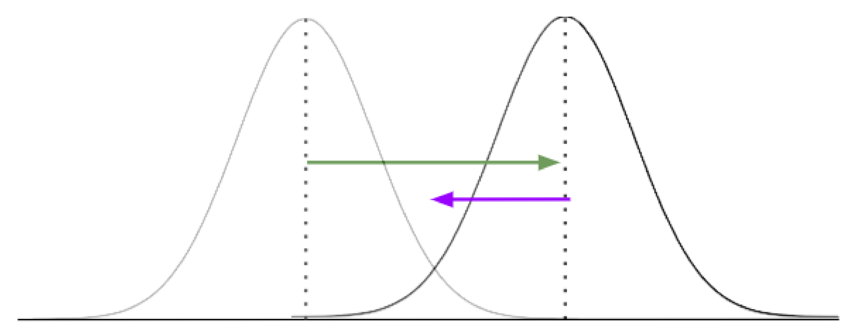
We explored changing the model parameters and computations. We found that a graded reward boundary paradigm was more efficient than the original model's success threshold paradigm and that an implementation of the indirect pathway increased model performance.

FUTURE DIRECTIONS

For future directions, the accuracy of the model's performance under different reward paradigms and with the indirect pathway implementation can be tested against empirical psychophysical data.

1

Mean Shift with Homeostasis (MeSH) is a biologically and behaviorally based reinforcement learning rule in which a distribution of action values is updated by reward reinforcement and a counteractive homeostatic drive.



$$M_{i+1} = M_i + \boxed{\omega_r\,(m_i - \bar{M}_i)} - \boxed{\omega_p\,(m_i - P)\,/\,|m_i - P|}$$

Distribution of possible reach amplitudes — Reinforcement from reward — Homeostatic drive

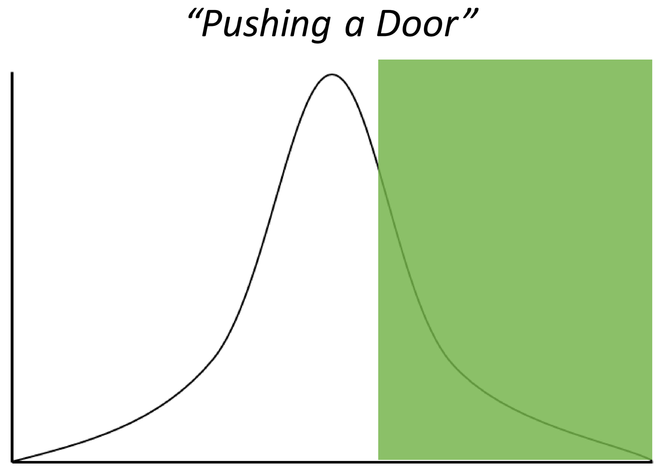
Yttri and Dudman 2018

M = Gaussian distribution of movement parameter values
m = Current movement parameter value
ω_r = Weight reward
ω_p = Weight set
P = Setpoint (starting mean of M)

2

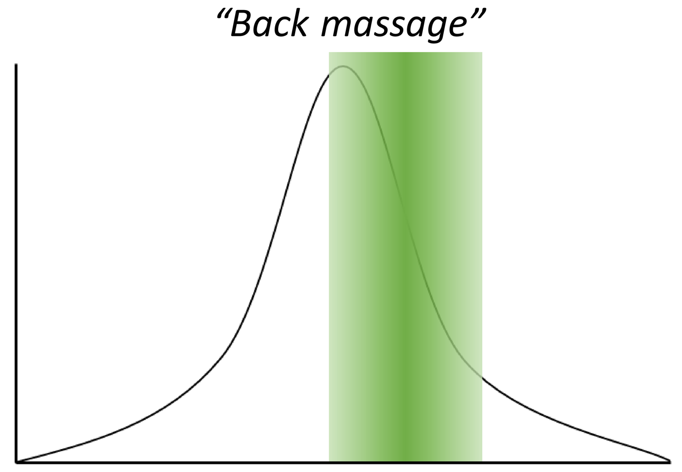
Since the requirement for task success will be task-specific, we tested the MeSH rule with different reward paradigms. Each has a different requirement for success. A reach value is randomly chosen from the distribution. If the reach is within the green area, the reach succeeds, and the model is rewarded. Three example paradigms are given below.

"Pushing a Door"



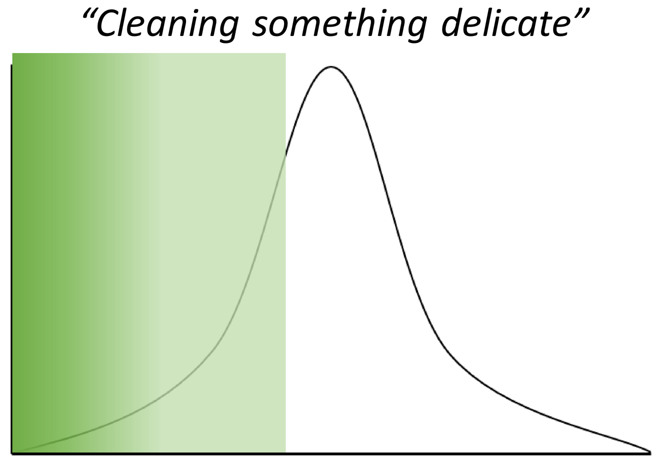
Success Reward Threshold

"Back massage"



Graded Reward Boundary

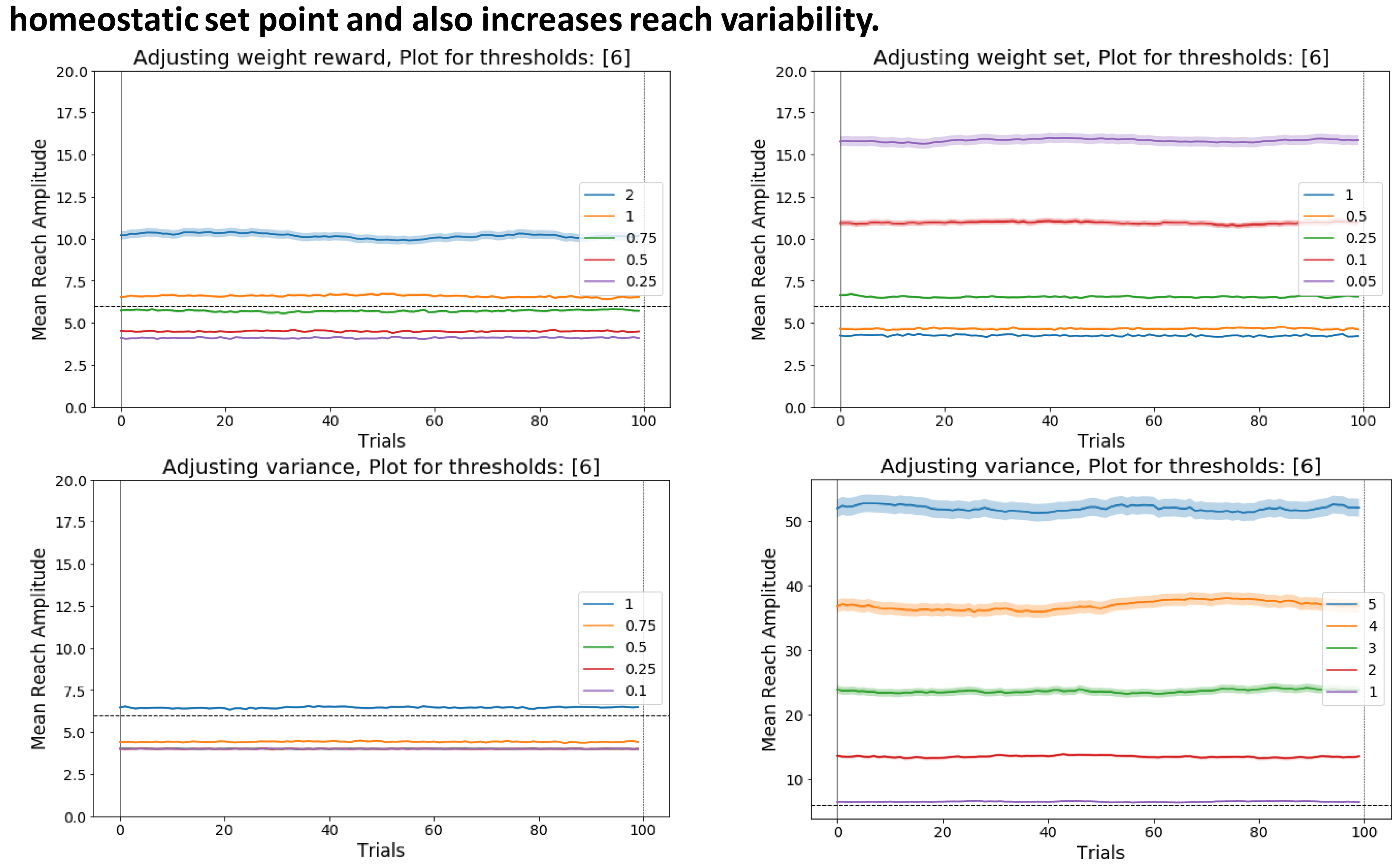
"Cleaning something delicate"



Graded Reward Threshold

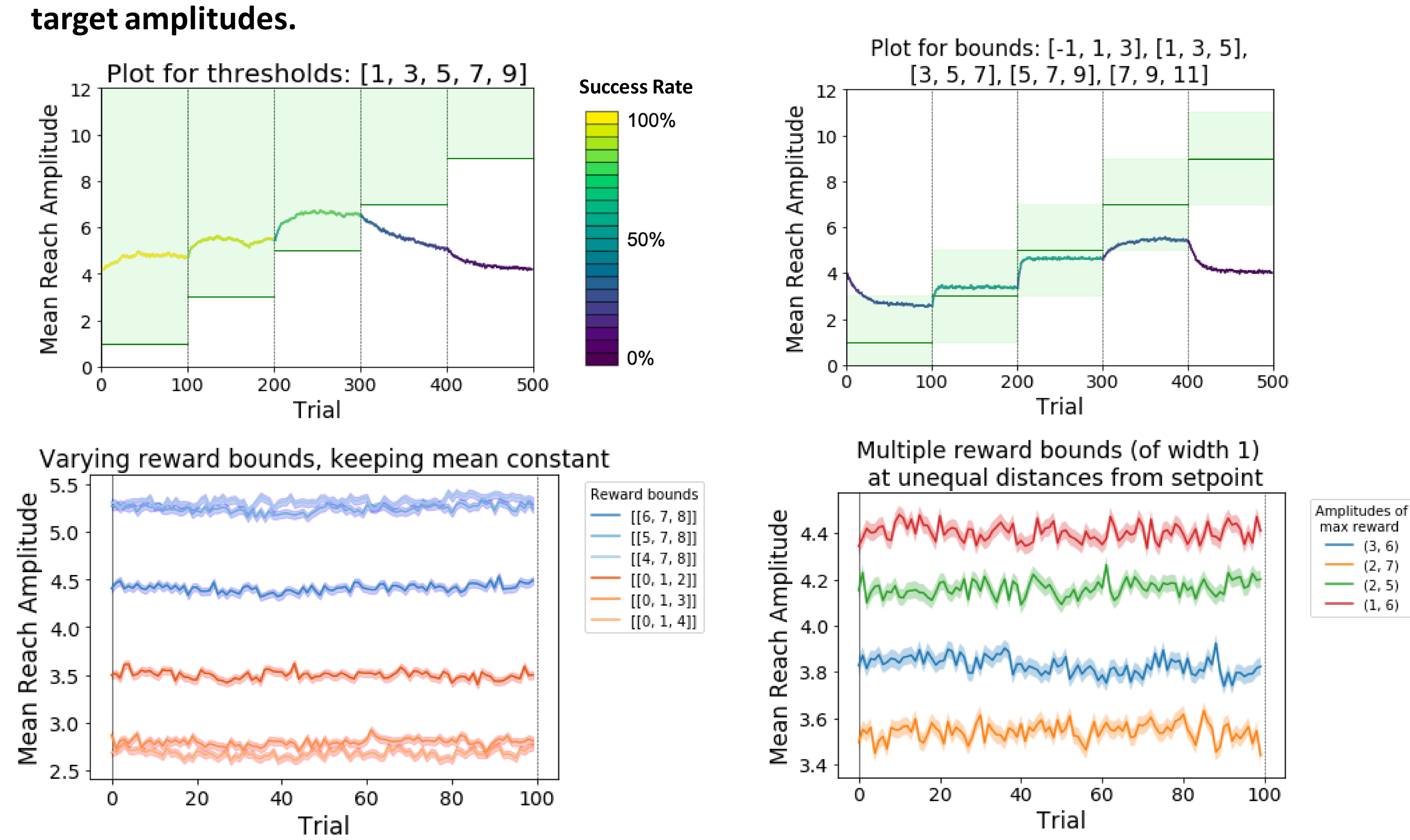
3

We systematically evaluated the contribution and optimal value for each parameter of the MeSH model. ω_r determines the extent to which the reaches move away from the homeostatic set point, while ω_p contributes to the tendency of reaches to return to the homeostatic set point. Like ω_r, variance contributes to the tendency of reaches to move away from the homeostatic set point and also increases reach variability.



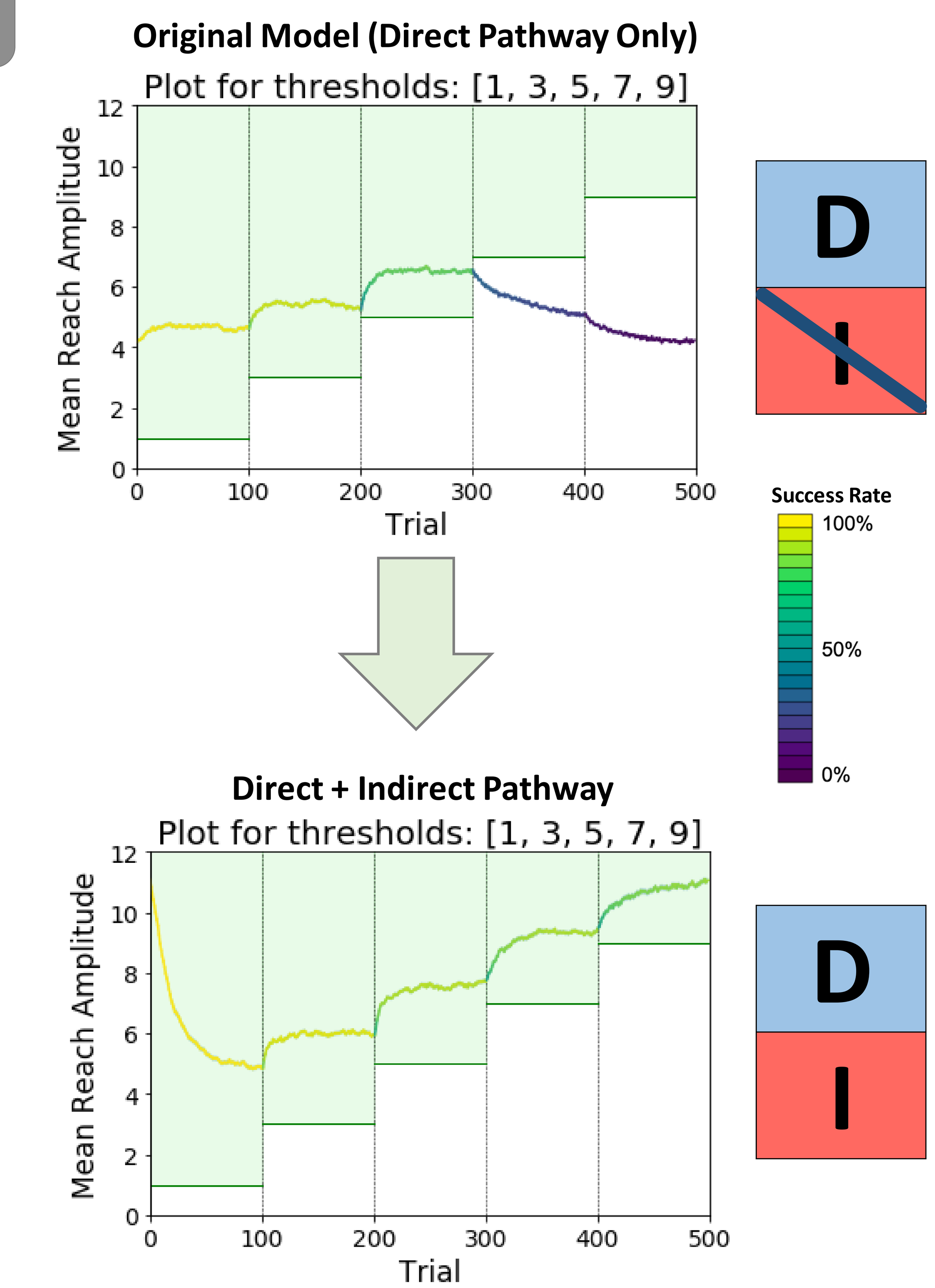
4

Using a simple reward threshold, performance fell off sharply at 3 standard deviations (SDs) above the set point, independently of the magnitude of the transition and the direction of the transition. Next the agent attempted to learn graded reward boundaries in place of the reward threshold. We noted that it acted more efficiently, learning to reach closer to target amplitudes that were within 1 SD of the set point. We also observed that the target amplitude, not the edge closest to the set point, determined whether the agent could learn the distribution. We tested this hypothesis by implementing asymmetrically graded reward bounds with different target amplitudes.



5

Implementing an indirect pathway into the original direct pathway-only model increased overall model performance.



6

In our exploration of our MeSH model, we found that a graded reward bounded paradigm had mean reaches closer to the target amplitudes than the original success threshold paradigm, making the graded reward bounded paradigm more efficient.

We also found that an implementation of the indirect pathway made the MeSH model more successful, especially at higher difficulties, than the original direct pathway-only model.