Emmanuel Velazquez
Exercise #8

Question 7.1

After running the model and observing the learning process, the layers reflected the probability through the undermining dopamine bursts or dips that come with different actions. In the MatrixGo Activation, the layer seemed to increase when the probabilities of reward were higher. This is because the dopamine burst reinforces the unit activation for the action which as a result causes the Go activation to show strong activation. For the MatrixNoGo Activation, it was evident that the activation decreased for actions that were associated with lower probabilities. This is because the dopamine dips reinforce the NoGo unit activation for which these actions are discouraged. This shows that the model learns to avoid these actions due to their association with negative outcomes. In PFCoutD Activation, it was illustrated that the actions that the model deemed rewarding, showed higher ActAvg activation due to the fact that this layer acts as a sort of decision maker. This resulted in a pattern where more rewarding actions are usually preferred. These models emerge because the model was created in a way that is supposed to simulate how real systems learn from rewards and punishments.

Question 7.2

After simulating the reduction in dopamine by setting BurstDaGain to 0.25 there was a significant change in learning and action selection. In the MatrixGo layer the ActAvg showed a decreased level of activation because of the reduced dopamine bursts. Due to the dopamine bursts being reduced, there is a weaker reinforcement of actions that are associated with positive outcomes. This impaired Go learning and the model ultimately becomes less effective at selecting items that lead to rewards.There wasn't a significant decrease in the ActAvg due to the fact that this simulation affected the Go pathway more by reducing the impact of the dopamine bursts. With reduced dopamine burst gain, the actions need to have a more rewarding outcome in order for the model to select them. This led the model to become more avoidant as it began to prefer actions that were less likely to result in a negative outcome. This reflects PD well as patients who are not on medication tend to show a tendency to learn more from avoiding losses rather than seeking gains.

Question 7.3

In the MatrixGo Activation with BurstDaGain restored to 1, the pathway increased compared to the simulation of patients without the medication. This change reflected the model's ability to reinforce Go learning due to the restored dopamine bursts. Actions that are assimilated with positive outcomes are accompanied by stronger probability which aligns with what we know about how patients with PD on medication tend to learn. For MatrixNoGo activation, setting DipDaGain to 0.25 mimics the effects of D2 agonists which block the dopamine dips. This causes the MatrixNoGo pathway's activation for actions associated with negative outcomes to

decrease. This is because the D2 agonists keep the D2 receptors activated which prevent the NoGO units from being overly excited by negative outcomes. This causes the model to become insensitive to learning from negative outcomes, ultimately decreasing NoGo activation as a whole. For the PFCoutD activation the layers ActAvg showed higher activation levels than in the simulation of patients who did not use medication. This means the model is more likely to select actions that are associated with positive outcomes.

Question 7.5

At the point where the reward was supposed to occur but did not, the RewPred layer predicts a reward based on the conditioned presentation. However, because the actual reward was absent the judgment between the expected and actual reward became negative. This is reflected in the TD layer with a value of -0.5. This represents a dip in the dopamine signal. This happened because the temporal difference learning mechanism relied on the difference between predicted rewards and the actual outcome. The model initially learned that the presence of the CS predicts a reward which leads it to adjust the weights to anticipate this reward. Then, when the reward is not received a signal is fired stating that the expectation was wrong. This negative TD value causes learning to digress which weakens the association between CS and the expected reward.

Question 7.6

Over the next several epochs of extinction training, the TD error signal plotted in the graph shows a negative spike at the time where the reward was previously received. This is a good representation of the network's surprise at not receiving the expected reward. As time goes on and the network begins to learn that the CS no longer predicts the rewards, this negative value is going to eventually go to 0. This is because the RewRed layer adjusts its predictions based on the lack of reward following the CS. The stimulus will cease to evoke an expectation of reward. This can be seen through the absence of the positive activation in the RewPred layer in response to the CS as the network begins to train. The TD error signals will also show that the network doesn’t expect a reward anymore in which the association between the CS and the reward has been extinguished. This means that the network has achieved a state where the reward predictions reflect the idea that the CS does not lead to a reward which illustrates the idea that the TD updated predictions based on changing environmental contingencies.