Lecture 10.3 Exercises

1.

a. Your misguided classmate argues that all instrumental learning is just the strengthening, through reinforcement, of stimulus-response associations (“habits”).
Name (or briefly describe) an experimental result that proves her wrong; explain (in one sentence) why it does.

b. Next, your misguided classmate argues that all instrumental learning occurs due to the acquisition of flexible cognitive maps or internal models, and never through the reinforcement of stimulus-response associations.
Name (or briefly describe) an experimental result that proves her wrong; explain (in one sentence) why it does.

c. Your misguided classmate blocks all dopaminergic activity in a hungry rat and finds that the rat is nonetheless still able to learn to press a lever for food. Your misguided classmate publishes an article in a prominent scholarly journal claiming that this experiment proves that dopamine must not be involved in instrumental learning.
If dopamine really does support stimulus-response instrumental learning, what alternative interpretation of this result is suggested by your knowledge of instrumental conditioning? What additional experiment should your misguided classmate do to test this interpretation?

2.

a. Explain in English, the intuition for why the temporal difference rule explains second-order conditioning.

b. Let’s work through an example of the temporal difference rule getting this right.

First, we train $A \rightarrow R$. Assume that the events in each trial consist of a series of 5 states, s_1: A , s_2: pause 1, s_3: pause 2, s_4: reward, s_5: end of trial, and that reward $r_t = 0$ everywhere but at s_4, where $r_t = 1$.
Assume the animal maintains a table of values, $V(s_i)$, one for each state and all initialized at zero.
These are updated at each timestep t according to $\text{new}\ V(s_t) \leftarrow \text{old}\ V(s_t) + \alpha\delta_t$, where
$\delta_t = r_t + \gamma V(s_{t+1}) - V(s_t)$ and the learning rate $\alpha = \frac{1}{4}$ and $\gamma = 1$. (Remember that in this model, t denotes timesteps within a trial, rather than trials.)
Run this learning rule for 3 trials, writing the predicted values and prediction errors at each timestep and trial. For each trial this will involve filling in the blanks in a table like this:

Trial 1					
timestep	t=1	t=2	t=3	t=4	t=5

state	s_1	s_2	s_3	s_4	s_5
r_t	0	0	0	1	0
old $V(s_t)$	0	0	0	0	always 0
$V(s_{t+1})$	0 (from t=2 ↗)	0	0	0	n/a
δ_t	0	0	0	1	n/a
updated $V(s_t)$	0	0	0	¼	always 0

Make two more tables of this sort, for the next two trials.

c. Assuming you ran this same sort of trial indefinitely, until the value predictions reached asymptote, what are the final values $V(s_t)$ for each state?

d. Finally, consider running a single trial of $B \rightarrow A$, with no reward. Assume the trial consists of a series of three states s_1: B, s_2: A, s_3: end of trial. Also assume that that $V(s_1)$, the predicted value of B is initially zero, and $V(s_2)$, the value of A is the asymptotic value following training (from c, above)

Run this learning rule for 1 trial, writing the values and prediction errors at each timestep. To do so, fill in the following table.

timestep	t=1	t=2	t=3
state	s_1	s_2	s_3
r_t	0	0	0
old $V(s_t)$	0		0
$V(s_{t+1})$			n/a
δ_t			n/a
updated $V(s_t)$			Always zero

1.

a. You receive a phone call. A shadowy voice says “Next time you are having a beer, I am going to steal it when the glass is still half full.” You believe the threat to be credible. What would you expect your dopamine neurons to do when you hear this message? Explain why, in terms of the prediction error equation.

b. Following this phone call, you go out and have a beer. It is, indeed, stolen when it is half full. What would you expect your dopamine neurons to do? Explain why, in terms of the prediction error equation.

c. Suppose dopamine neurons signaled prediction error according to a Rescorla-Wagner theory (i.e., as the difference between immediately expected and immediately received primary reward) rather than via the TD equation, which includes future value. How would you have answered question (a) above, in this case?