

Please answer the below questions succinctly:

1. (45 points) Delayed reward: experiments show that a delayed reward is less valued than an immediate reward.
 - a. (10 points) Suggest a (simple) experiment that will demonstrate this
 - b. (10 points) Suggest a (simple) experiment that will let you evaluate how much 4 pellets of food are worth when they are delayed by 5 seconds, 15 seconds or 30 seconds.
 - c. (10 points) In the temporal difference learning model we can model the decrease in value of the reinforcer by multiplying the reinforcer by a *discount factor* $0 \leq \gamma \leq 1$ for every timestep of delay. Thus, instead of defining the value at time t as a sum of future rewards

$$V(t) = r(t) + r(t+1) + r(t+2) + \dots = \sum_{i \geq t} r(i)$$
 we define it as the sum of *discounted* future rewards

$$V(t) = r(t) + \gamma r(t+1) + \gamma^2 r(t+2) + \dots = \sum_{i \geq t} \gamma^{i-t} r(i)$$
 write down the prediction error in this case, and what would be the learning rule (hint: follow the steps we took to develop the prediction error and learning rule for TD learning without discounting in class, but with the new definition of V)
 - d. (5 points) How, according to this model, does the value of the reinforcer decrease with time? Draw curves for a reinforcer of initial size 1, as a function of delay time, for discount factors 0.95, 0.8 and 0.5. These are called *exponential discounting functions*.
 - e. (5 points) Evidence shows that the discounting function that humans and animals use are not exponential, but rather hyperbolic, that is, the value of the reinforcer is divided by $(1+a*t)$ where t is the delay and $0 < a < 1$ is a constant. Draw curves for hyperbolic discounting of a reinforcer of initial size 1, as a function of delay, using several values of a .
 - f. (5 points) Compare exponential discounting and hyperbolic discounting qualitatively in terms of the shape of the function: in what ways are they similar, and, importantly, in what do they differ?
 - g. (Bonus - 10 points) Why can't we write a temporal difference learning rule for hyperbolic discounting?

2. (40 points) Sniffy the virtual rat: operant training in a Skinner box.

You can download a 2-week trial version of Sniffy at http://www.wadsworth.com/psychology_d/templates/student_resources/0534633609_sniffy2/sniffy/download.htm; I recommend going over the tutorial in Project 2 which is the one relevant for us: http://www.wadsworth.com/psychology_d/templates/student_resources/0534633609_sniffy2/sniffy/exercises/exer2.htm. Note that you will have a trial version and thus I believe you cannot save files.

 - a. (15 points) Train Sniffy to press the lever to get food with free operant training (use shaping). You can use anything you have learned in the course, instrumental or Pavlovian. Write down how long it took (in minutes) until full acquisition of the

association between leverpressing and food. You can also note any unusual events that took place during training, or strategies that you developed. (Note: the initial training can take a long time, 30 minutes or more. The better you are at training the faster it will go).

- b. (5 points) Extinguish the trained behavior (use the 'design operant conditioning experiment' option under the 'Experiment' menu). Note how long it took until you recorded 5 minutes without a lever press. You can use the option of "isolated Sniffy" at this point, to have the simulation run faster.
- c. (10 points) Retrain Sniffy (you don't have to start a new rat, you can use the same previously trained one) with partial reinforcement of 50% until full acquisition of the leverpress association (to train with 50% reinforcement use the option "variable responses" in the above menu). Think carefully about which techniques can shorten the training time, and try them. Write down the techniques you tried, whether they worked or not, and why do you think they did.
- d. (10 points) Extinguish the response and compare to the extinction in part b of this question. Was extinction after partial reinforcement training faster? Slower? What, in your opinion, is the effect of training with partial reinforcement on extinction and why?

3. (15 points) Unblocking:

- a. (5 points) In the blocking paradigm learning about one stimulus is blocked by another stimulus. Describe an experimental protocol under which you will get blocking. Don't forget to mention how you will test for the blocking effect, and against what control group.
- b. (5 points) In 'unblocking', in stage 1 a stimulus A is paired with a US (say, a food pellet), and in stage 2 a compound stimulus AB is paired with *more* US (eg., 2 food pellets). Explain why this procedure does not block learning to B. After the second stage of training would you expect the conditioned response to B to be higher, similar, or lower than that to A?
- c. (5 points) In 'downward unblocking' in stage 1 a stimulus A is paired with 2 food pellets, and in stage 2 a compound stimulus AB is paired with only one food pellet. What would you predict would be learned about B in this case?