Assignment: Chapter 8 Rewards for Robots

1. What is the primary method used by animal trainers?

1A. When the journalist Amy Sutherland was doing research for a book on exotic animal trainers, she learned that their primary method is preposterously simple: “reward behavior I like and ignore behavior I don’t.”

2. What is meant by the term “operant conditioning?”

2A. A learning method where behaviors are chosen from rewards/punishments.

3. TRUE/FALSE - Operant conditioning inspired an important machine-learning approach called reinforcement learning.

3A. True

4. TRUE/FALSE - Reinforcement learning requires labeled training examples.

4A. False

5. TRUE/FALSE - In reinforcement learning, an agent – the learning program – performs actions in an environment (usually a computer simulation) and occasionally receives rewards from the environment. These intermittent rewards are the only feedback the agent uses for learning.

5A. True

6. TRUE/FALSE - The technique of reinforcement learning is a relatively new addition to the AI toolbox.

6A. False

7. TRUE/FALSE - Reinforcement learning played a central role in the program that learned to beat the best humans at the complex game of Go in 2016.

7A. True

8. In just a few sentences, describe the “illustrative example” that MM used to communicate the basic concepts associated with reinforcement learning, in general, and the variant of reinforcement learning known as Q Learning, in particular.

8A. With the Rosie robot learns with Q learning over a series of episodes. Initially the only value in the Q-table is the end result of kicking the ball. Rosie performs random actions until she kicks the ball. The next episode starts where a value is assigned to the table just before she reached the goal. Rosie performs her actions again until she reaches the point where she is one step before the goal and uses the value in the table to perform the action that will lead her to the kick. This is repeated until she has learned from start to finish which actions she should perform the reach the ball and kick it.

9. TRUE/FALSE - The promise of reinforcement learning is that the agent can learn flexible strategies on its own simply by performing actions in the world and occasionally receiving rewards (that is, reinforcement) without humans having to manually write rules or directly teach the agent every possible circumstance.

9A. True

10. TRUE/FALSE - In general, the state of an agent in a reinforcement learning situation is the agent's perception of its current situation.

10A. True

11. TRUE/FALSE - A crucial notion in reinforcement learning is that of the value of performing a particular action in a given state.

11A. True

12. In reinforcement learning, what is the value of action A in state S?

12A. The value of action A in state S is a number reflecting the agent's current prediction of how much reward it will eventually obtain if, when in state S, it performs action A.

13. What is the "Q-table" in Q-learning?

13A. This table of states, actions, and values is called the Q-table.

14. Why the name "Q-learning?"

14A. The letter Q is used because the letter V (for value) was used for something else in the original paper on Q-learning.

15. The Q-learning manifestation of reinforcement learning is a process that iterates over “episodes” until the learning is accomplished. What is an episode in this learning technique?

15A. Reinforcement learning occurs by having Rosie take actions over a series of learning episodes, each of which consists of some number of iterations.

16. List a couple of issues, other than the “exploration versus exploitation balance” issue, that reinforcement-learning researchers face for complex tasks.

16A. For example, in real-world tasks the agent’s perception of its state is often uncertain, unlike Rosie’s perfect knowledge of how many steps she is from the ball. A real soccer-playing robot might have only a rough estimate of distance, or even some uncertainty about which light-colored, small object on the soccer field is actually the ball.

17. Deciding how much to explore new actions and how much to exploit (that is, stick with) tried-and-true actions is called the exploration versus exploitation balance. Achieving the right balance is a core issue for making reinforcement learning successful. What real world example does MM use to illustrate the exploration versus exploitation balance?

17A. When you go to a restaurant, do you always order the meal you’ve already tried and found to be good, or do you try something new, because the menu might contain an even better option?

18. MM identifies two “stumbling blocks” to using reinforcement learning in the real world. Please briefly describe each of these stumbling blocks.

18A. First, there’s the Q-table. In complex real-world tasks—think, for example, of a robot car learning to drive in a crowded city—it’s impossible to define a small set of “states” that could be listed in a table.

The second stumbling block is the difficulty, in the real world, of actually carrying out the learning process over many episodes, using a real robot. Even our “Rosie” example isn’t feasible. Imagine yourself initializing a new episode—walking out on the field to set up the robot and the ball—hundreds of times, not to mention waiting

around for the robot to perform its hundreds of actions per episode. You just wouldn’t have enough time.