Markov Decision Process

**Summary**

**Reinforcement learning** is about learning from interaction with the environment. The objective is to learn how to behave in order to achieve a goal. In this session, you learnt the following:

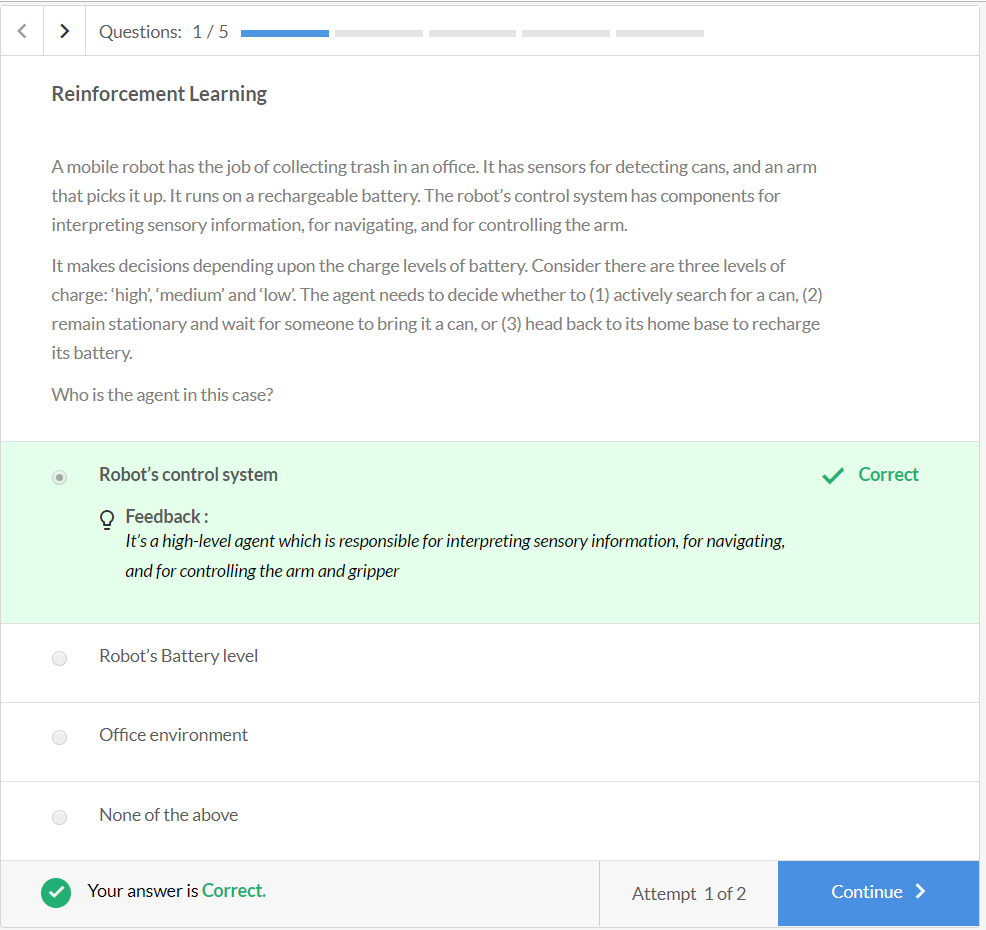
* In a reinforcement learning problem, **an agent** learns how to behave in an **environment**  
  by taking **actions** and **seeing the consequences** - **rewards and change in state.**
* The **control objective** of the agent is to **learn a policy to accumulate maximum rewards**over a period of time.
* The entire reinforcement learning problem is based on the **Markov assumption**: the current state contains all relevant information to take the future action.

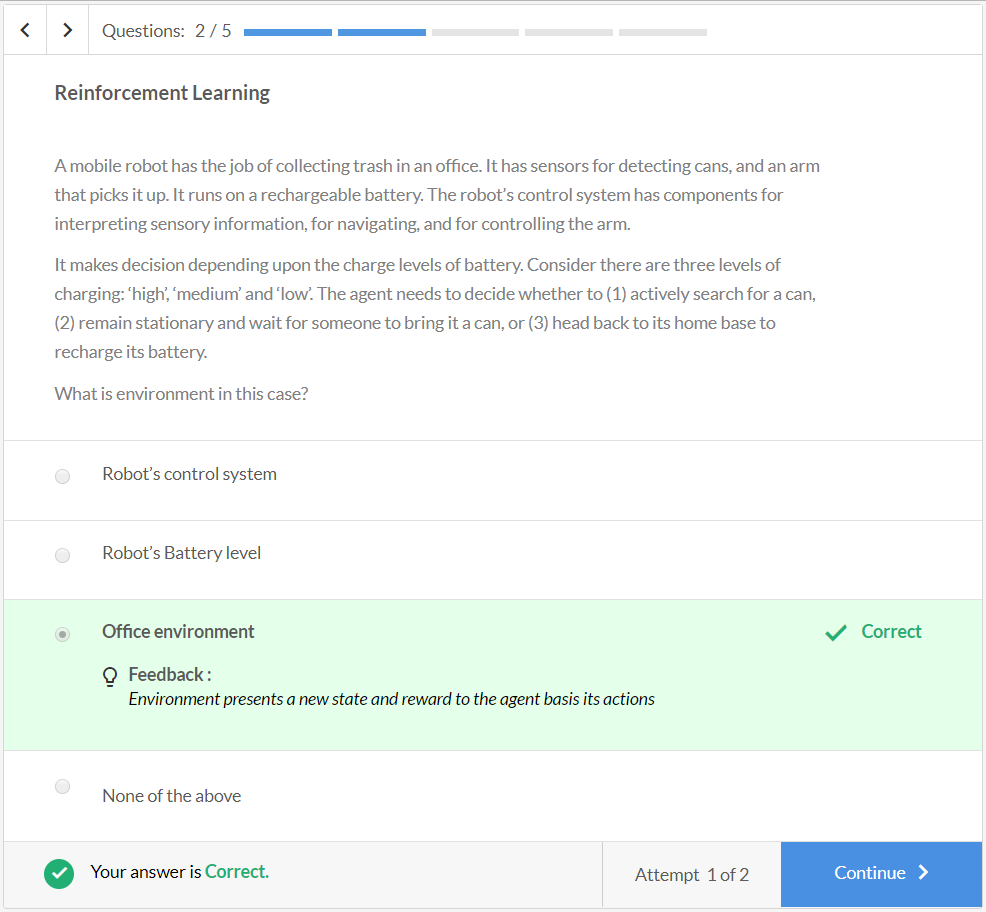
You also learnt about deterministic and stochastic policy. And how the stochastic policy gives a window for **exploration.**

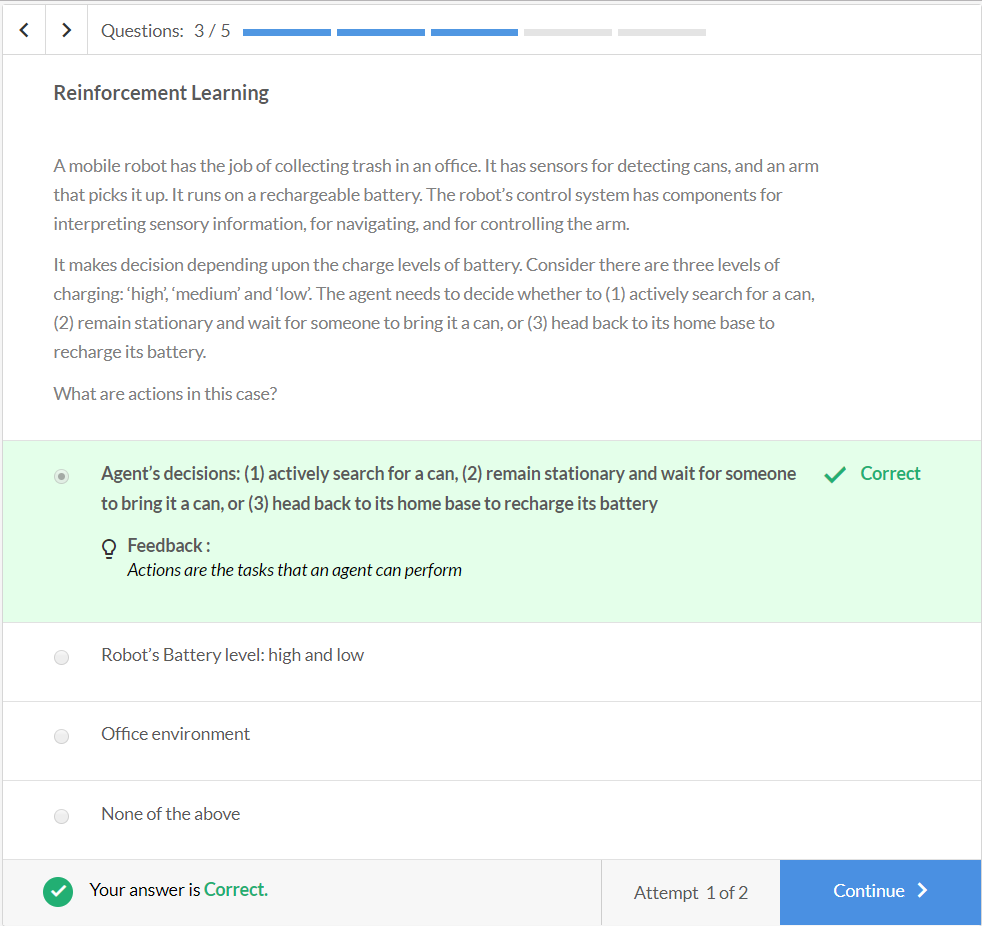
Then, you learnt about**value function** and **action-value (q-value)**functions. These are the expected return from that state, or state–action pair, given that the agent uses the policy.

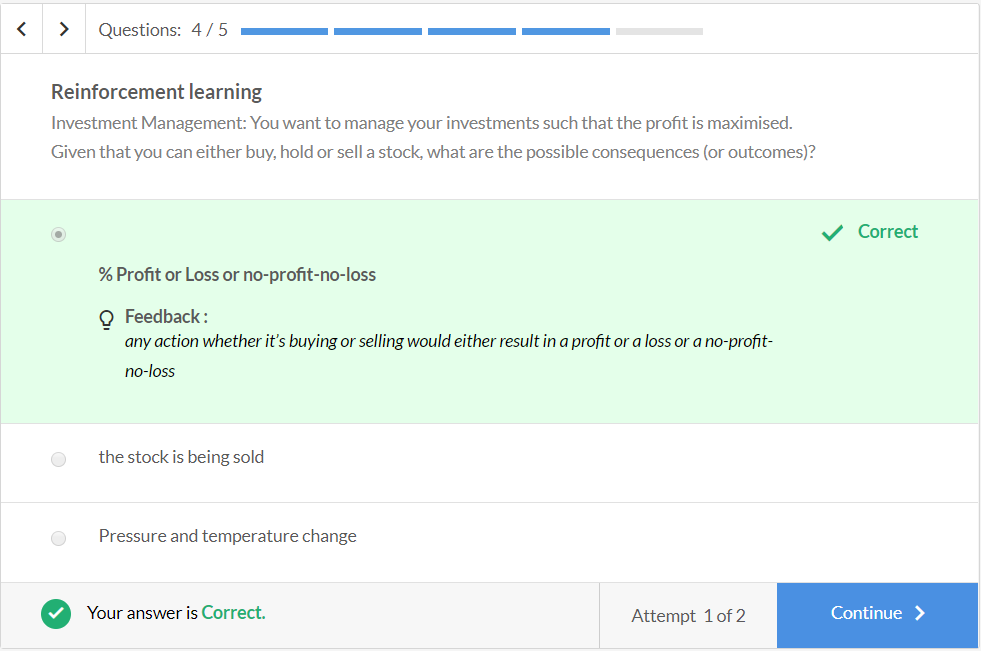
A policy for which value functions are optimal is an **optimal policy**. While the optimal value functions for states and state–action pairs are unique for a given MDP, though there can be many optimal policies.

Then, you learn about the **model of the environment.**In problems, where the agent has a complete knowledge about the environment’s dynamics we call them as **model-based problems.** In the case when a complete model of the environment is not available, we call those problems as **model-free.**

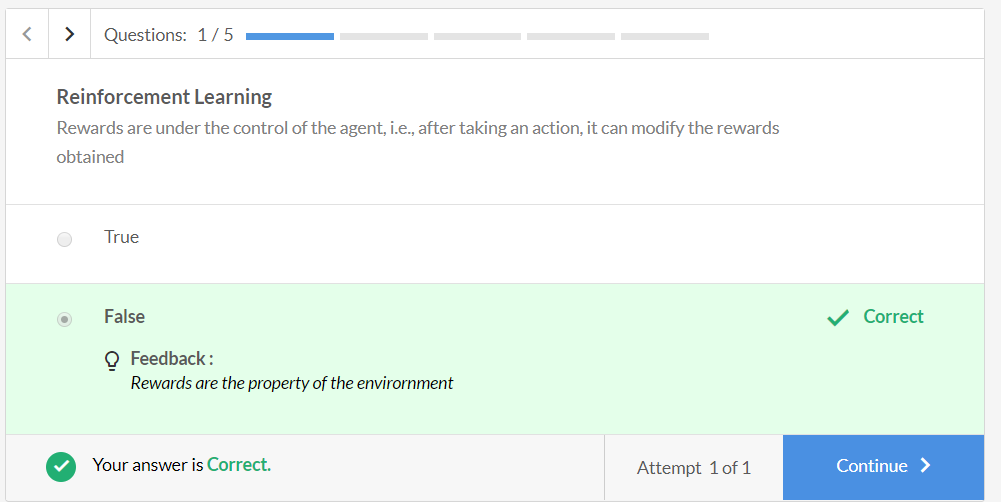


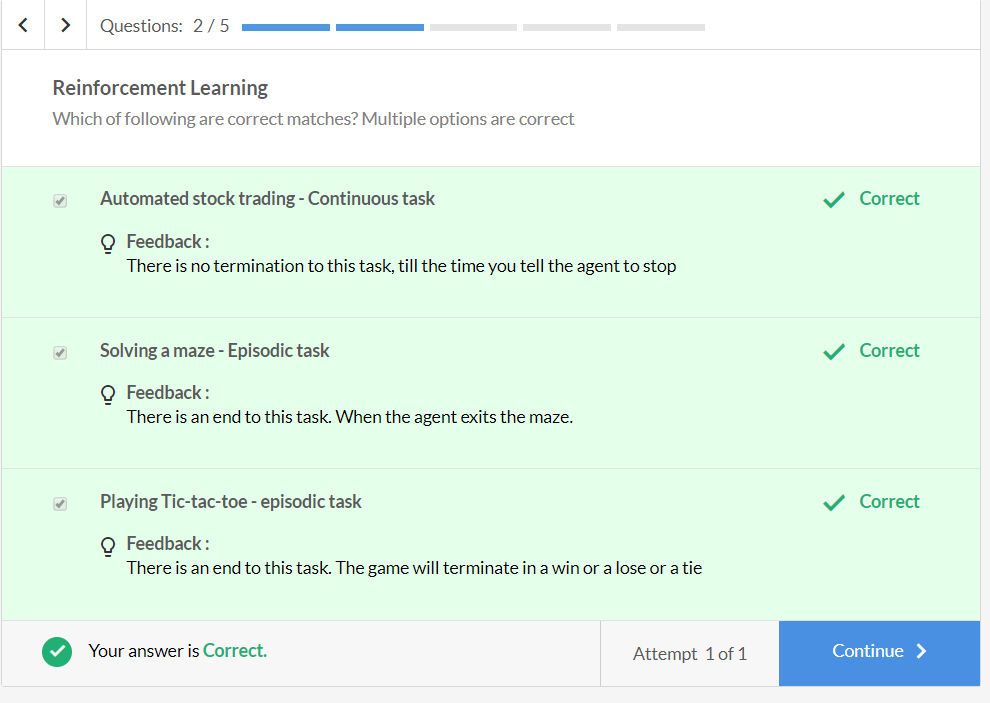


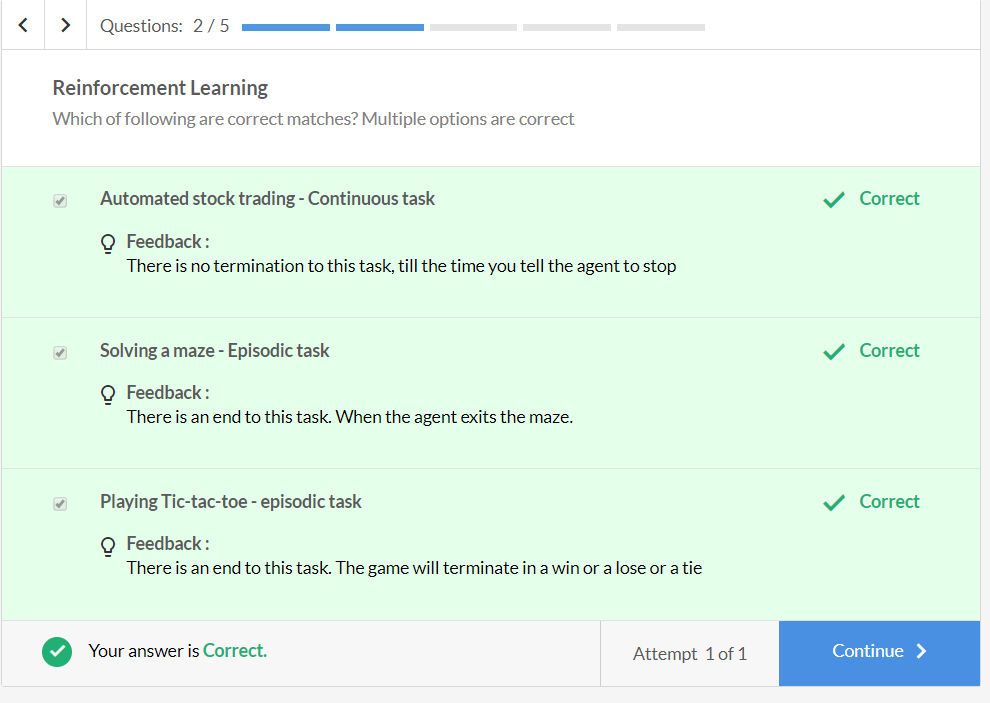


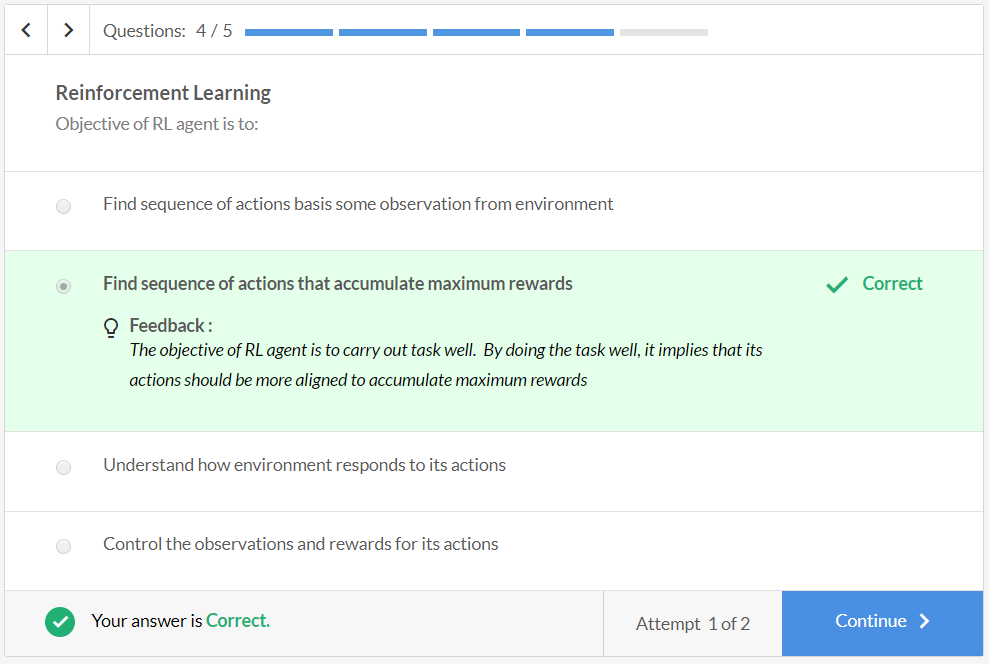






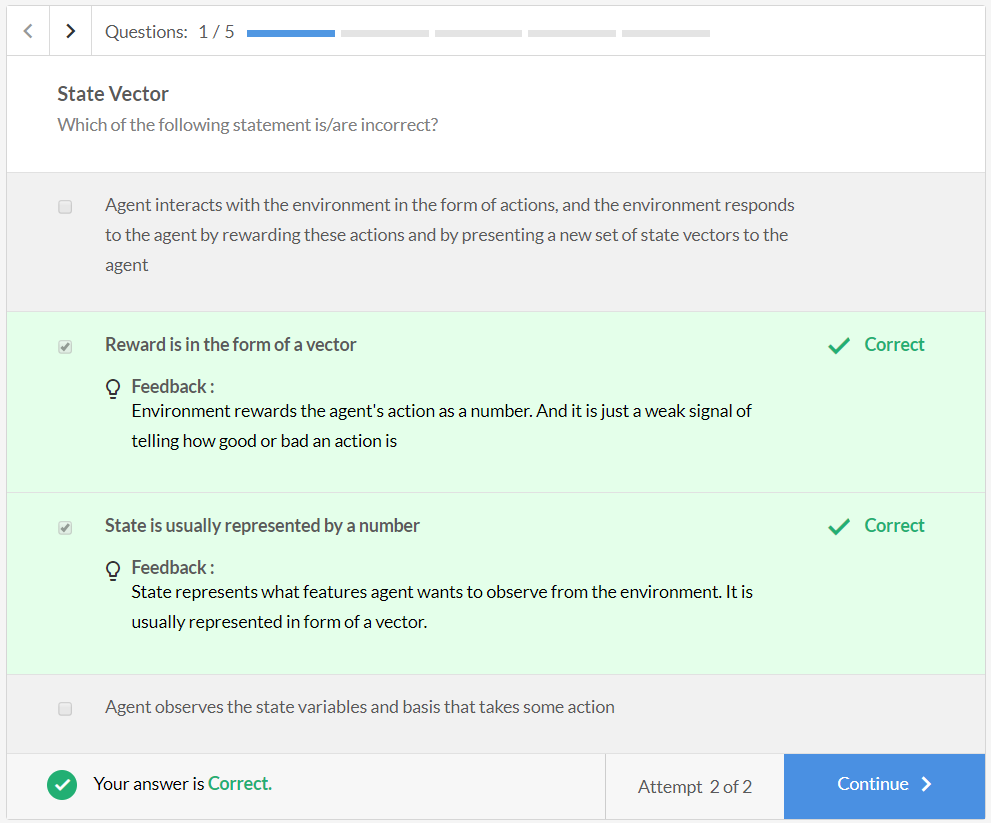


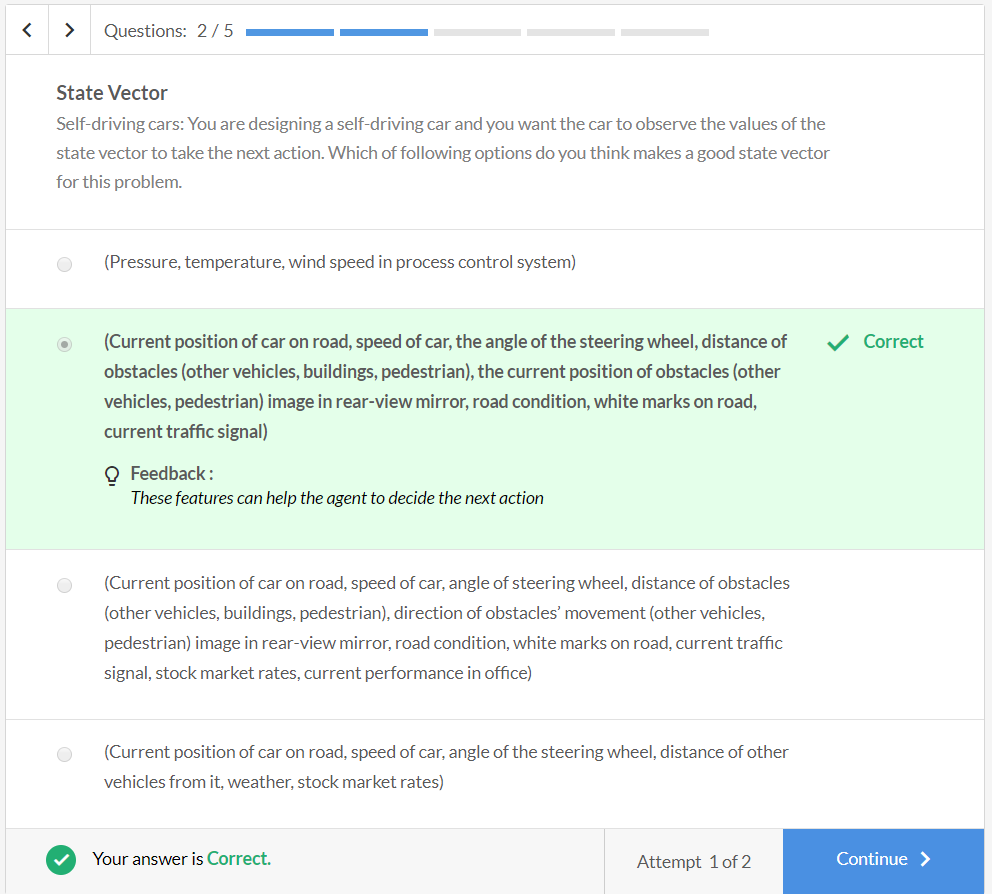


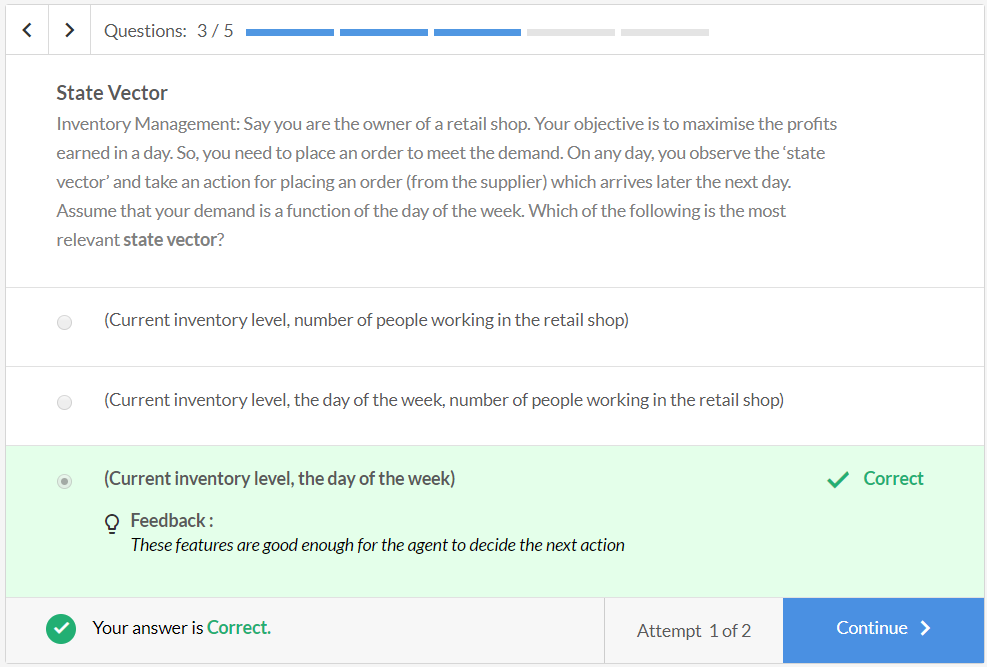


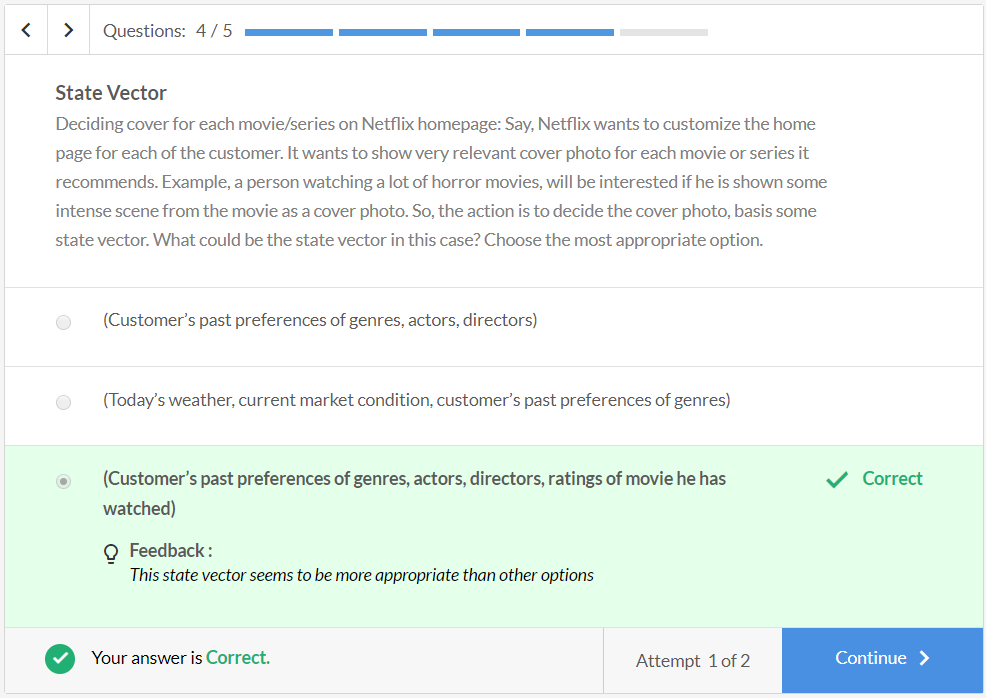


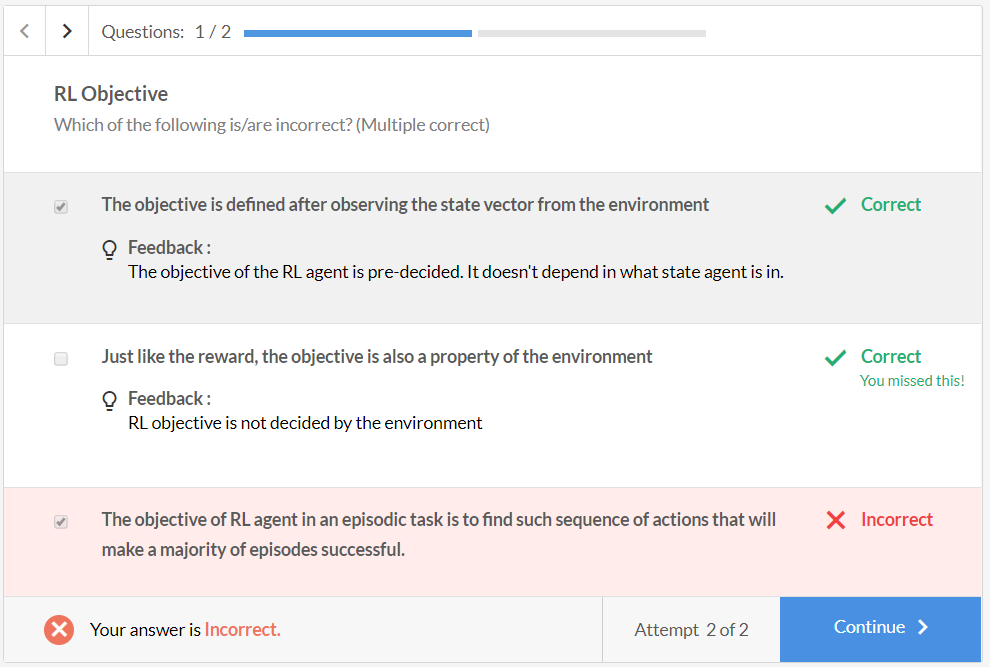
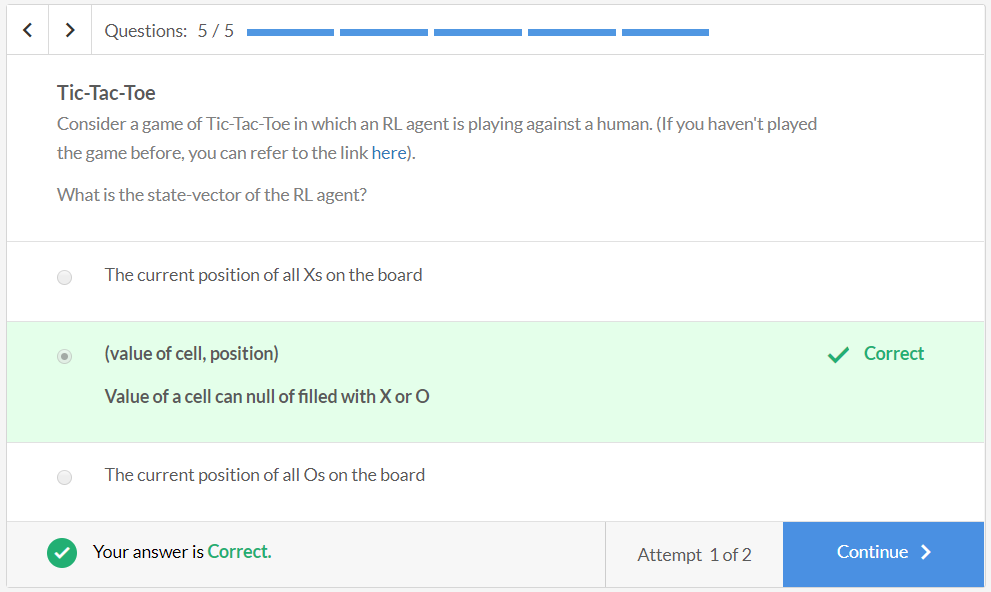


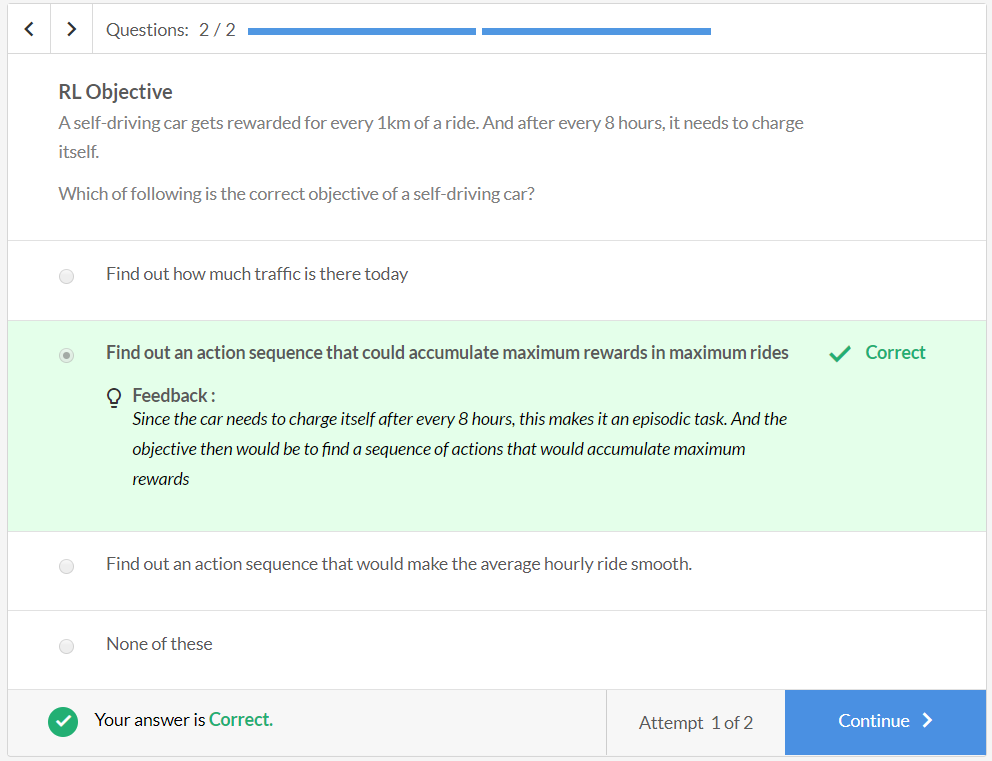


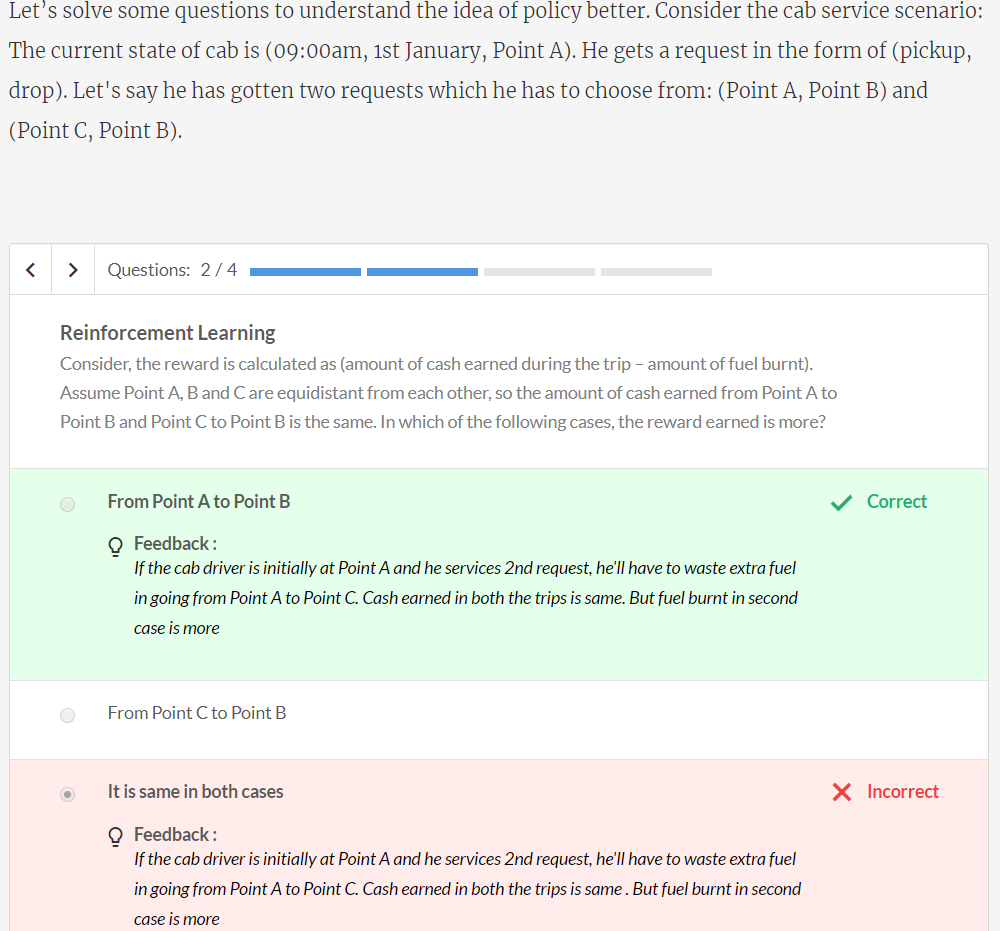
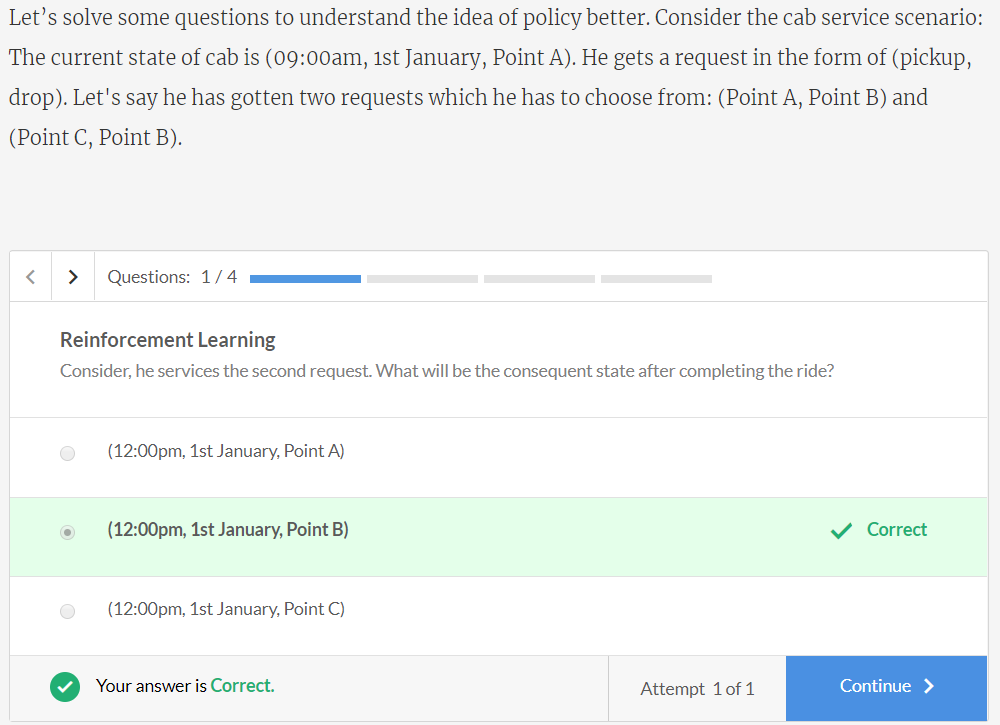


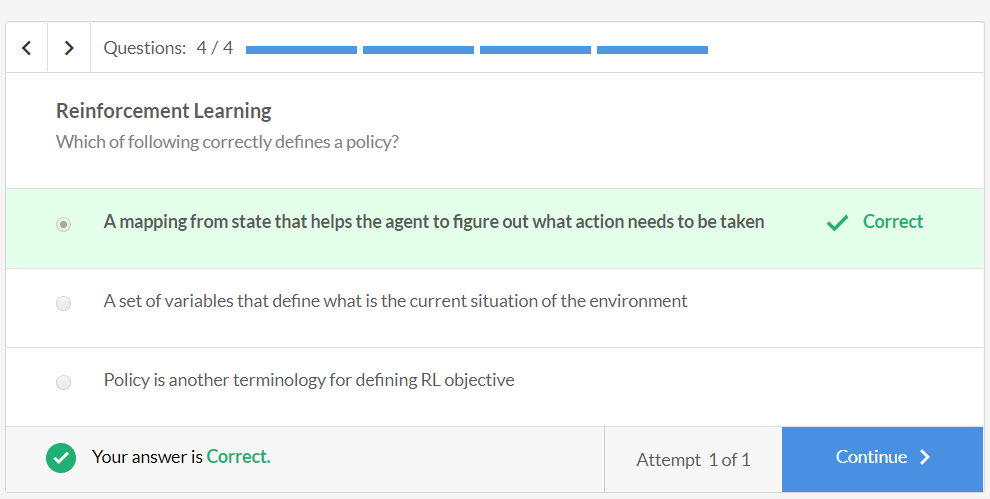
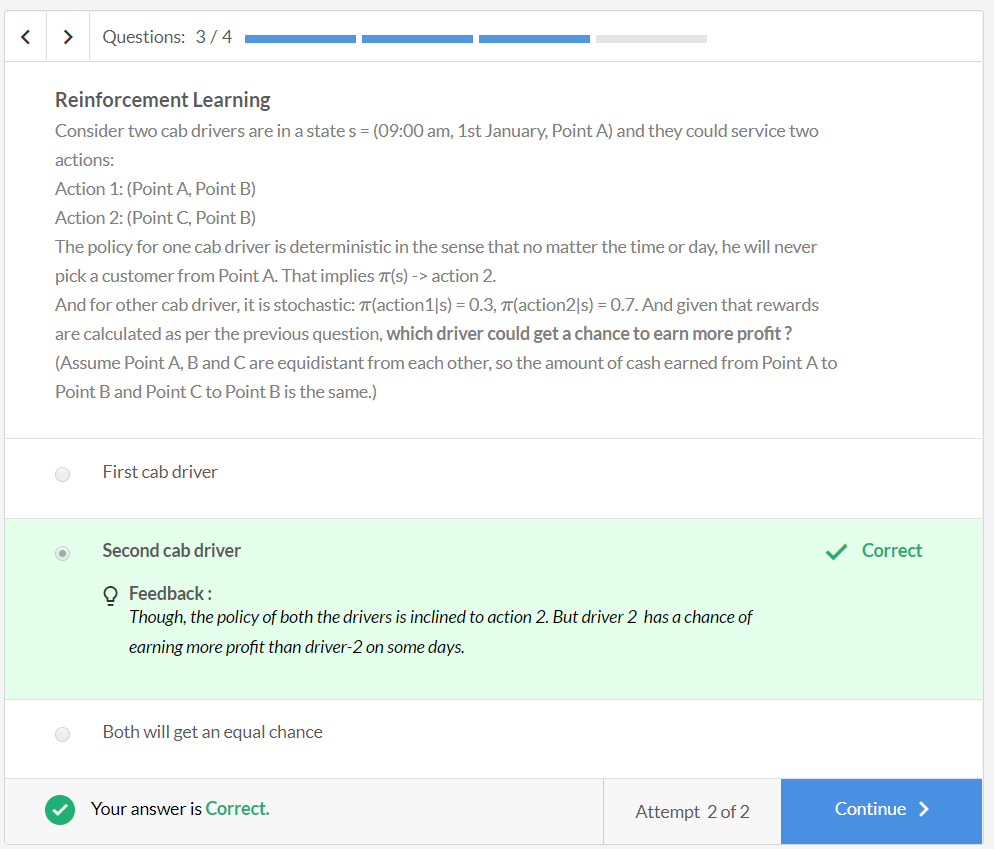


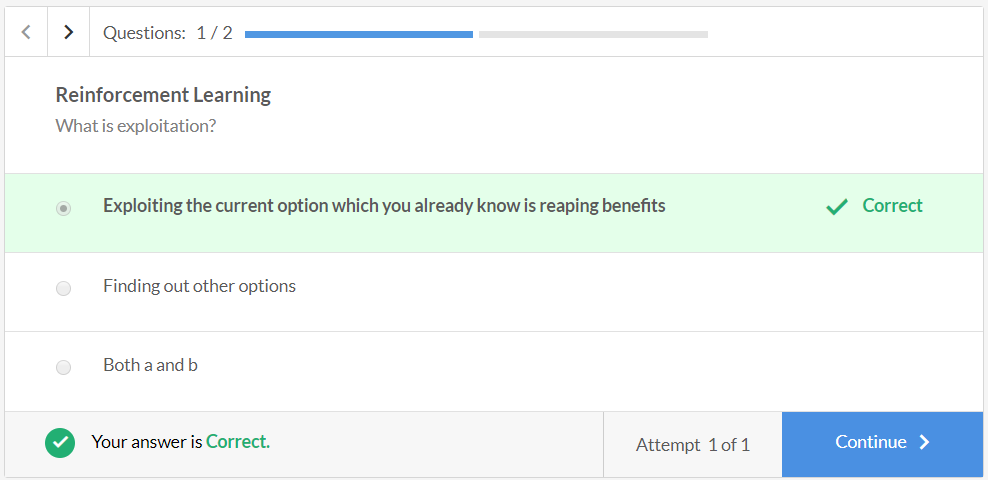


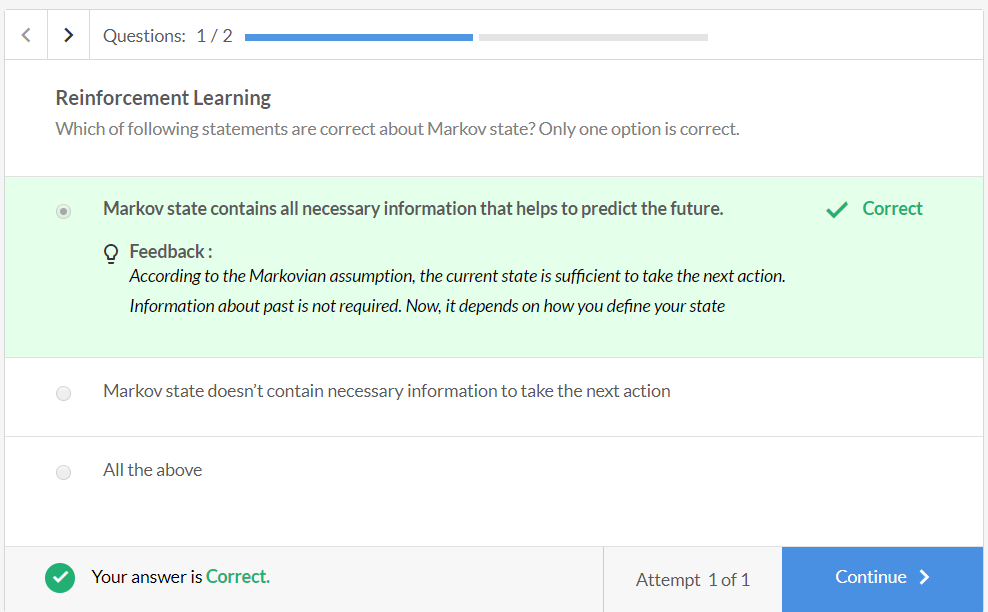


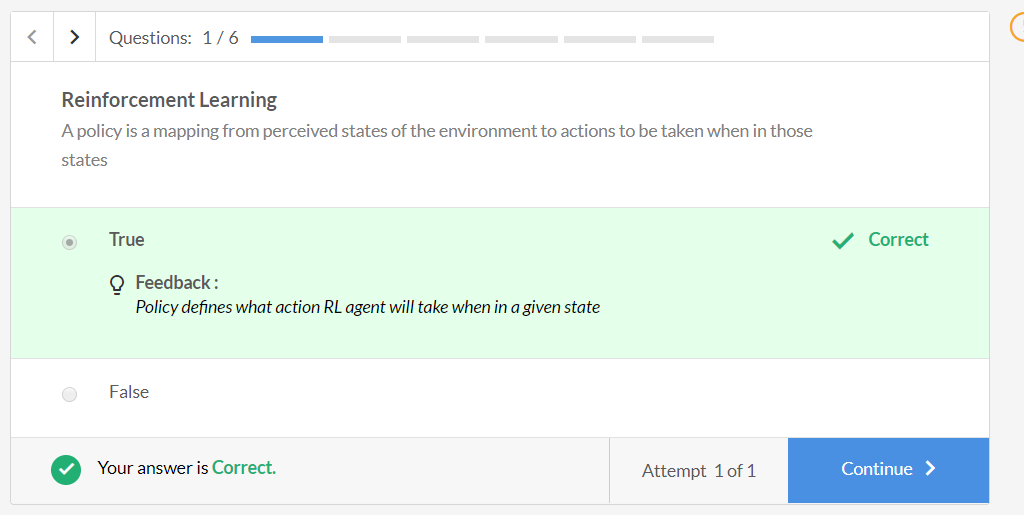


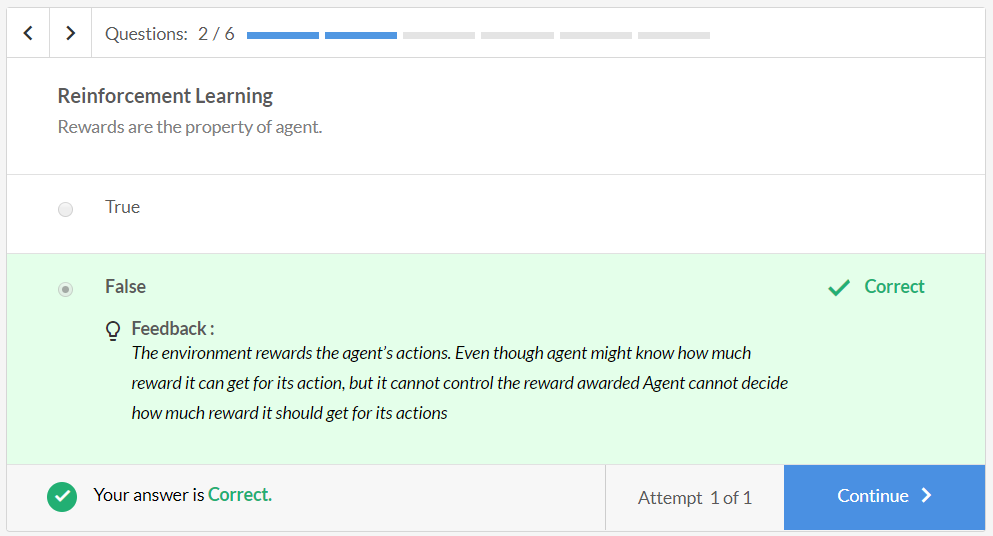


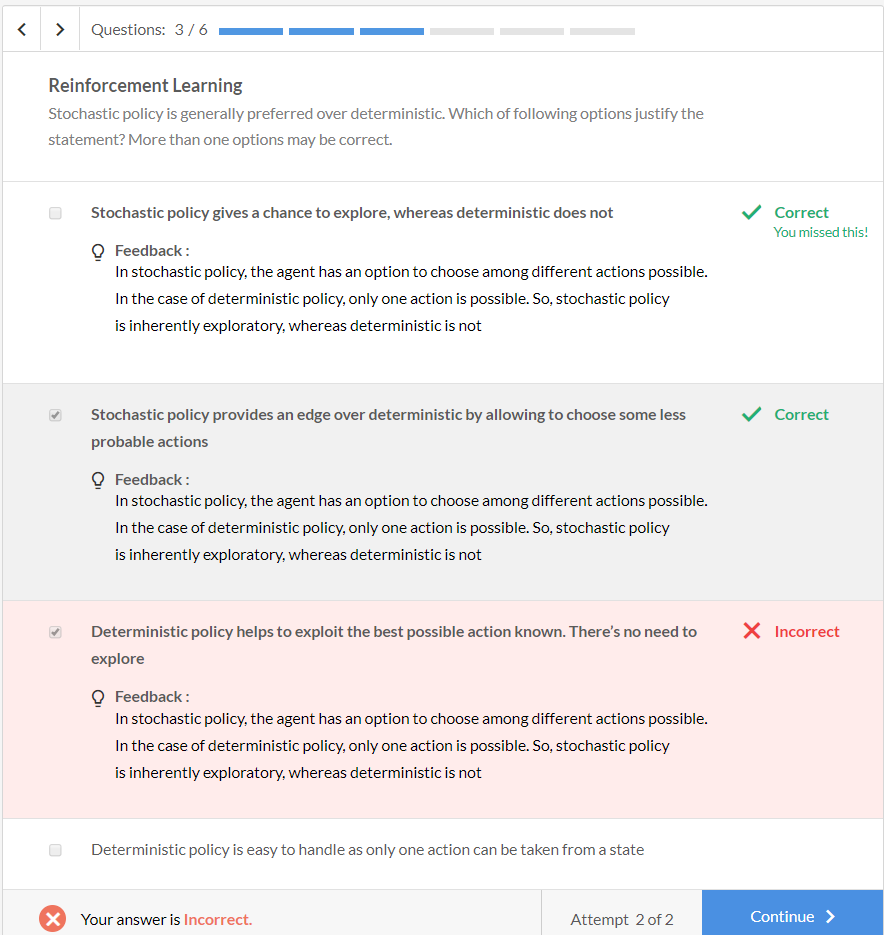


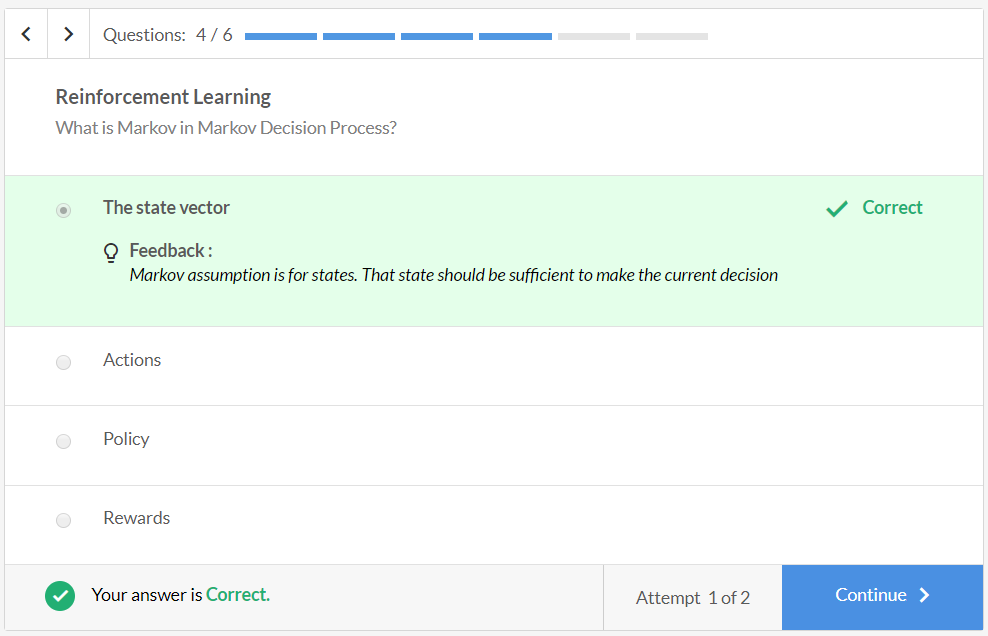


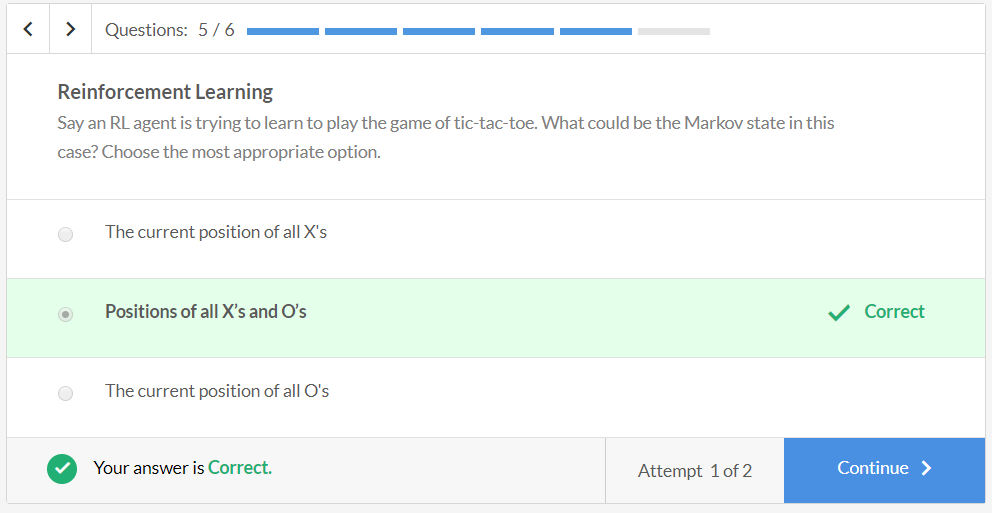


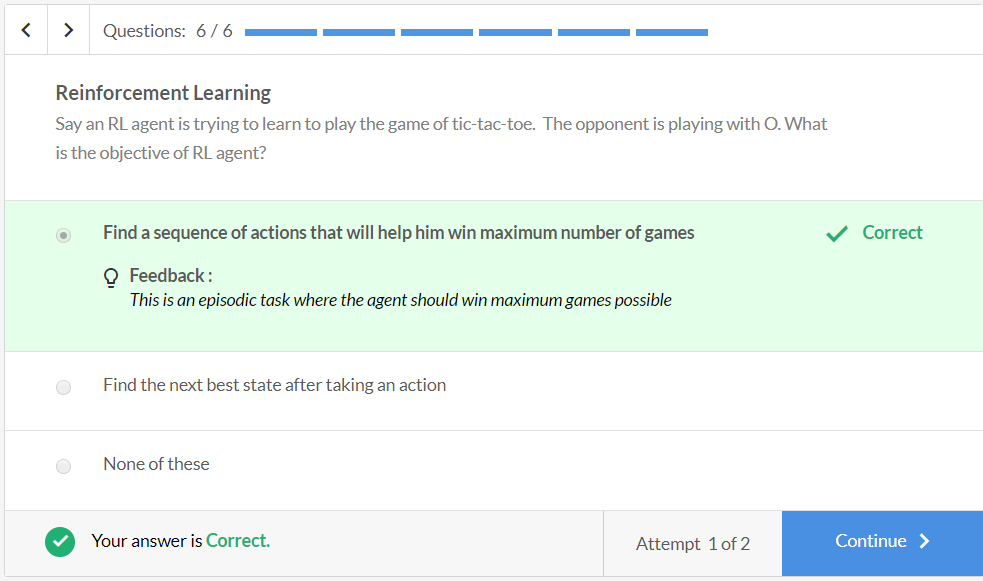


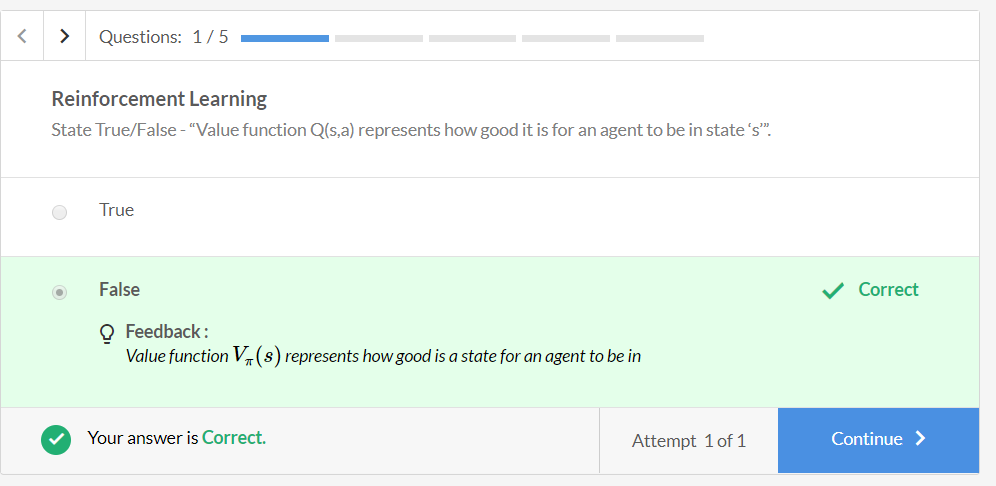


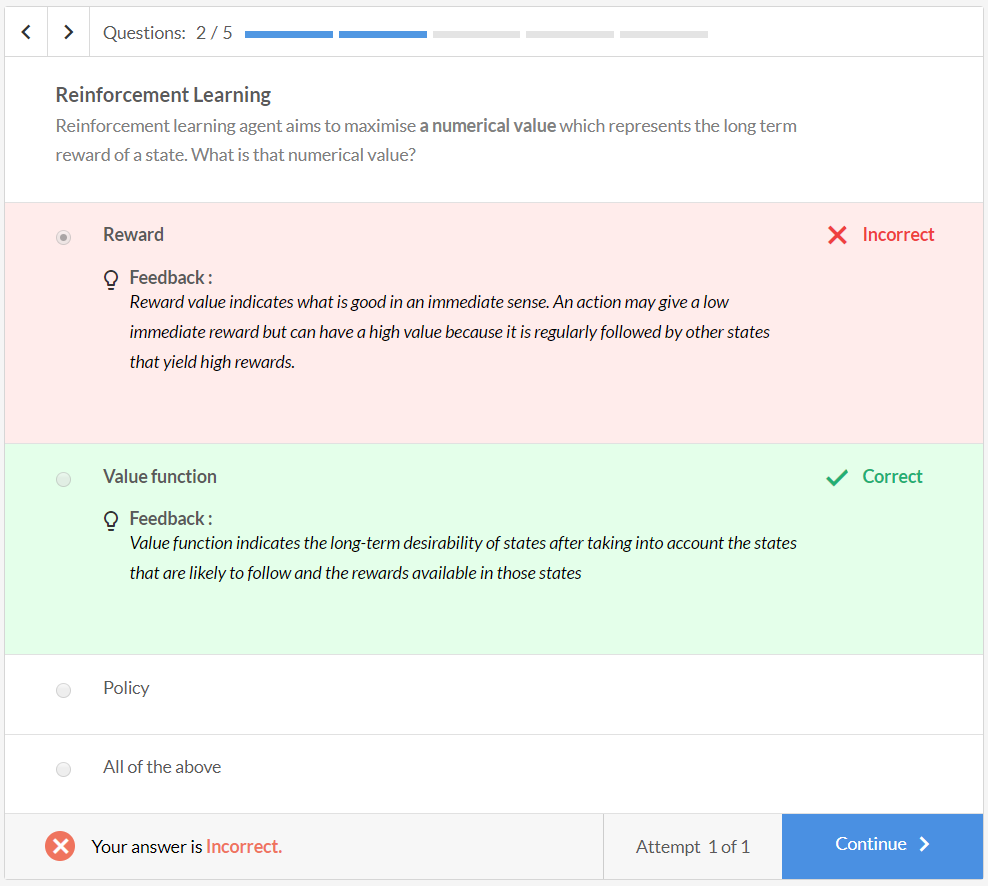


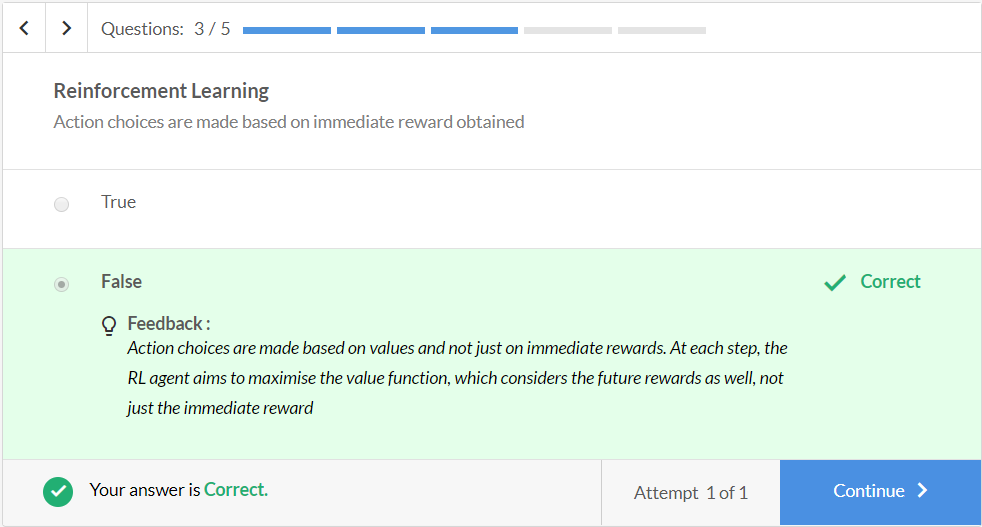


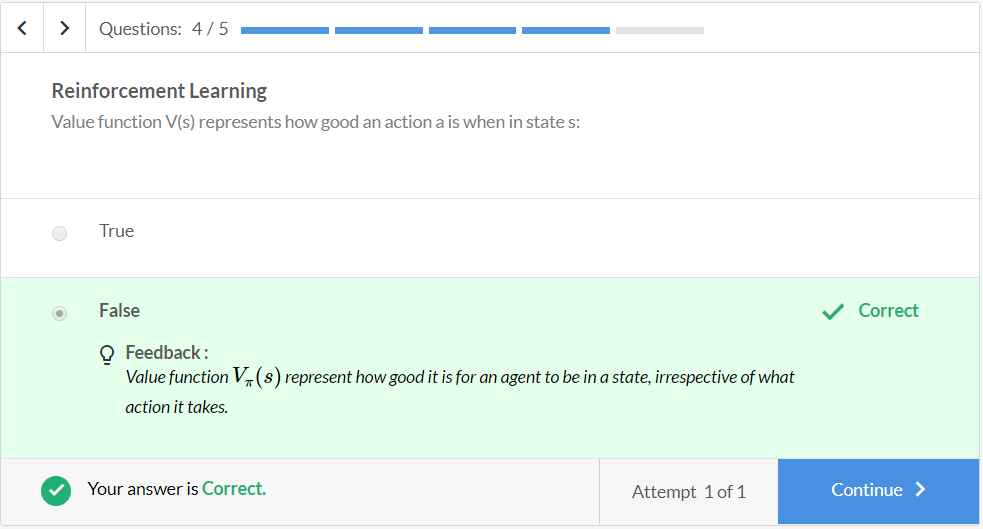


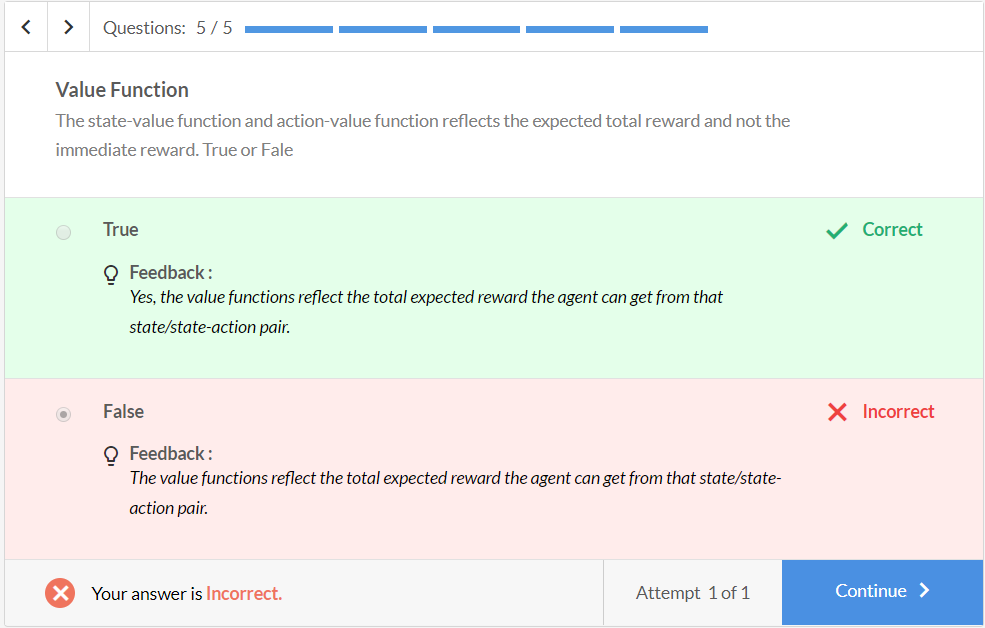


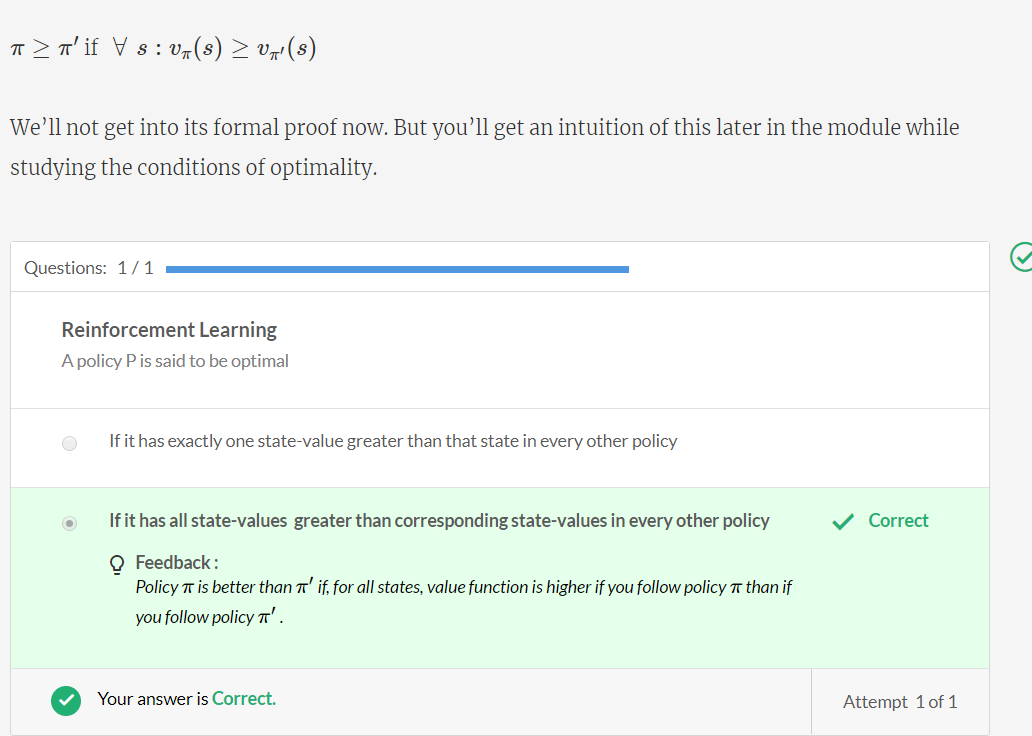




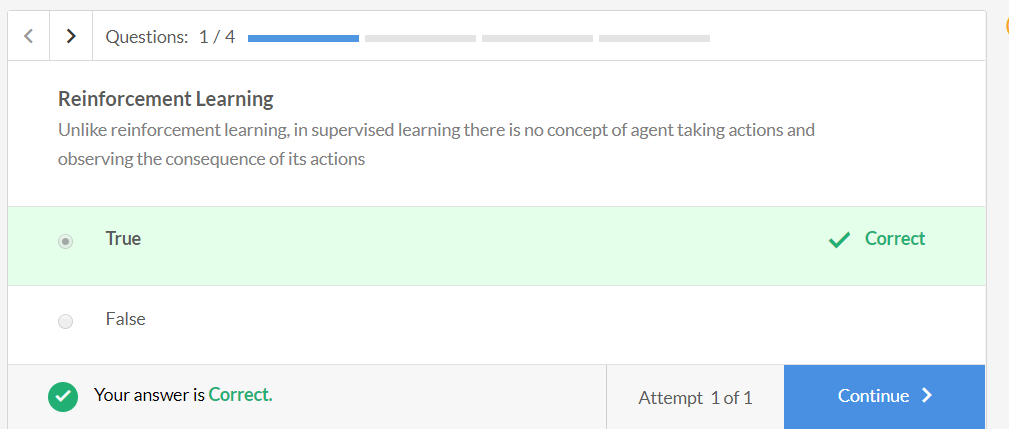


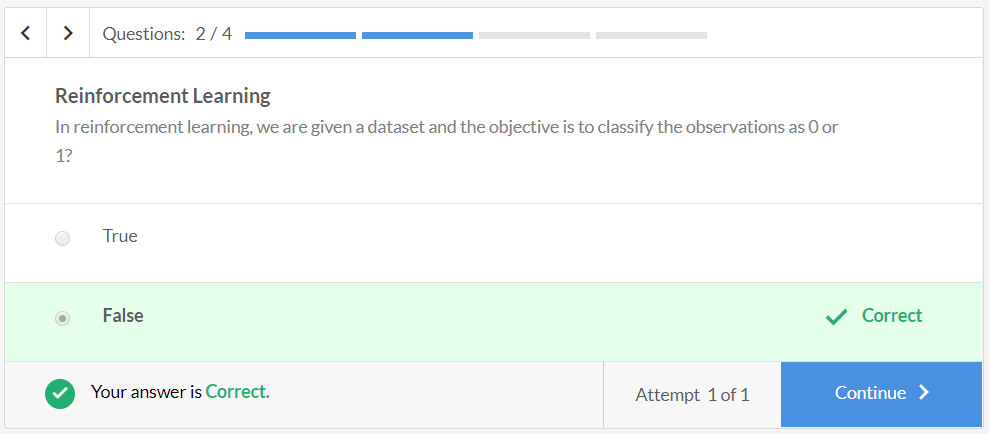




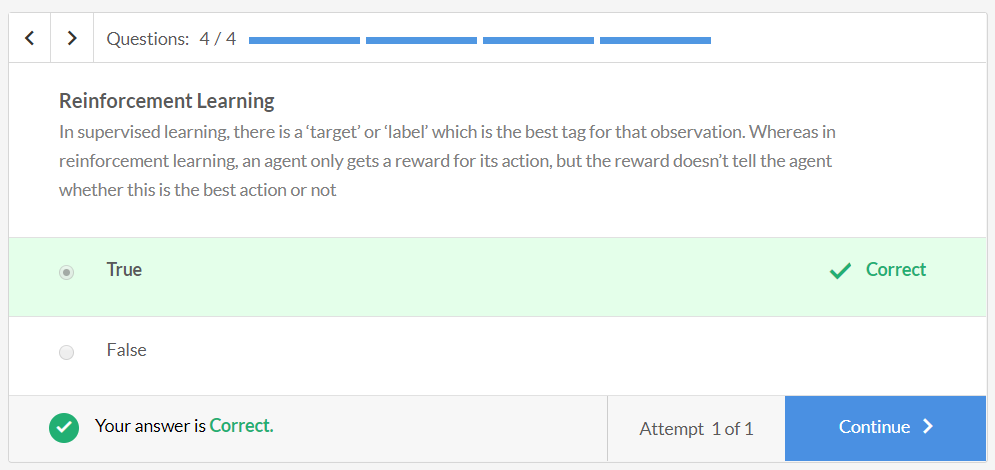


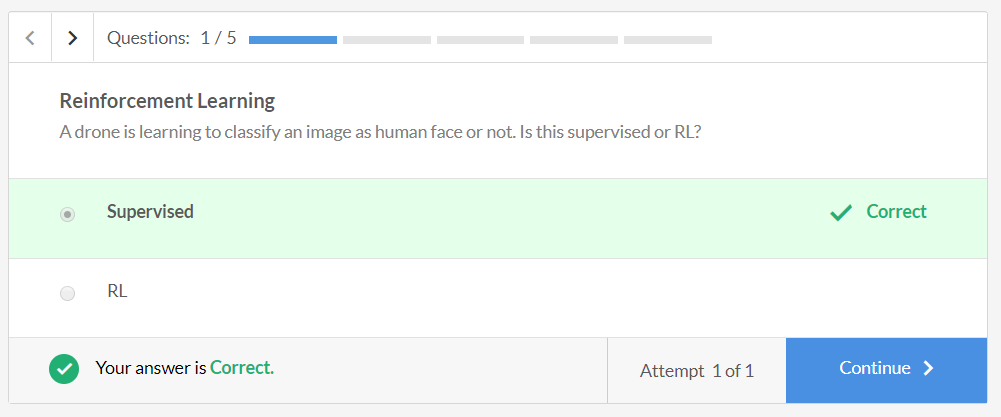


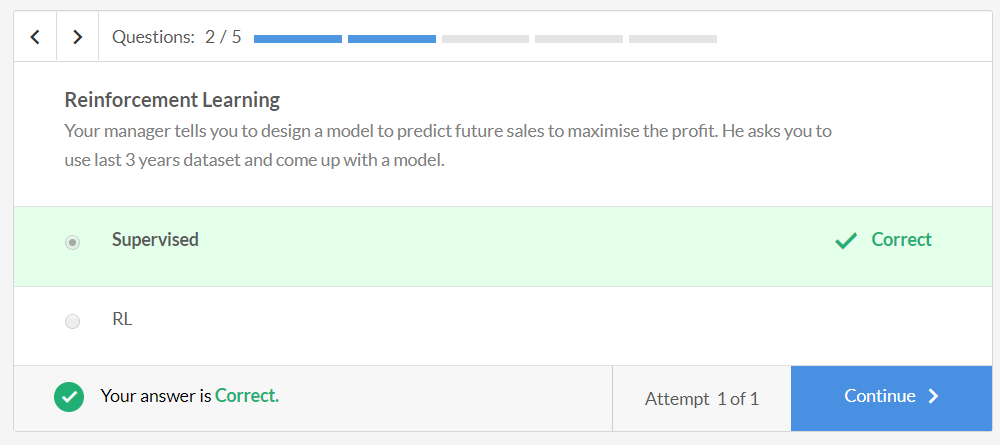


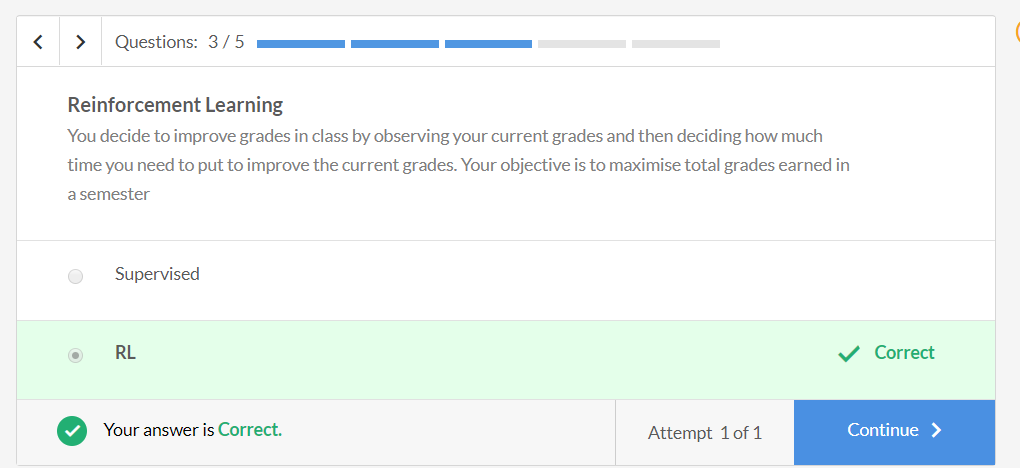


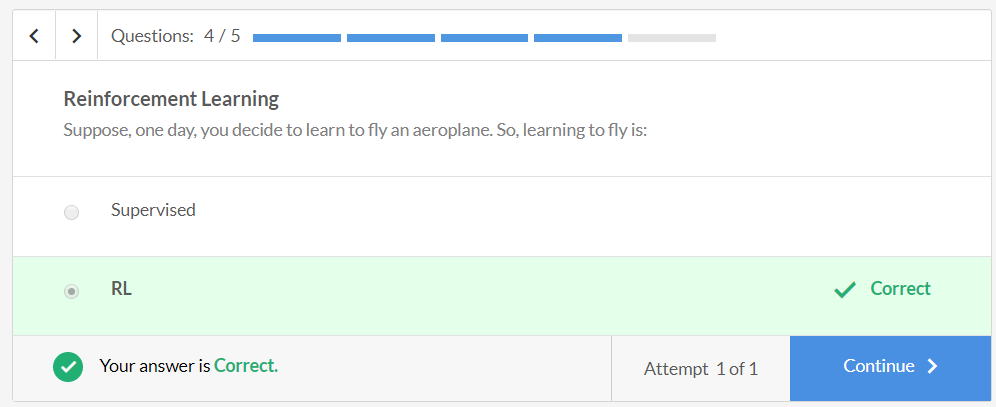


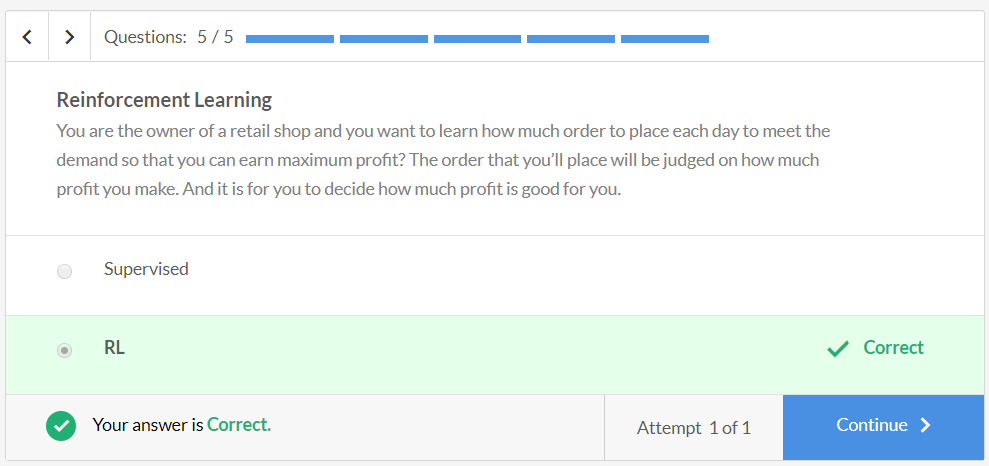


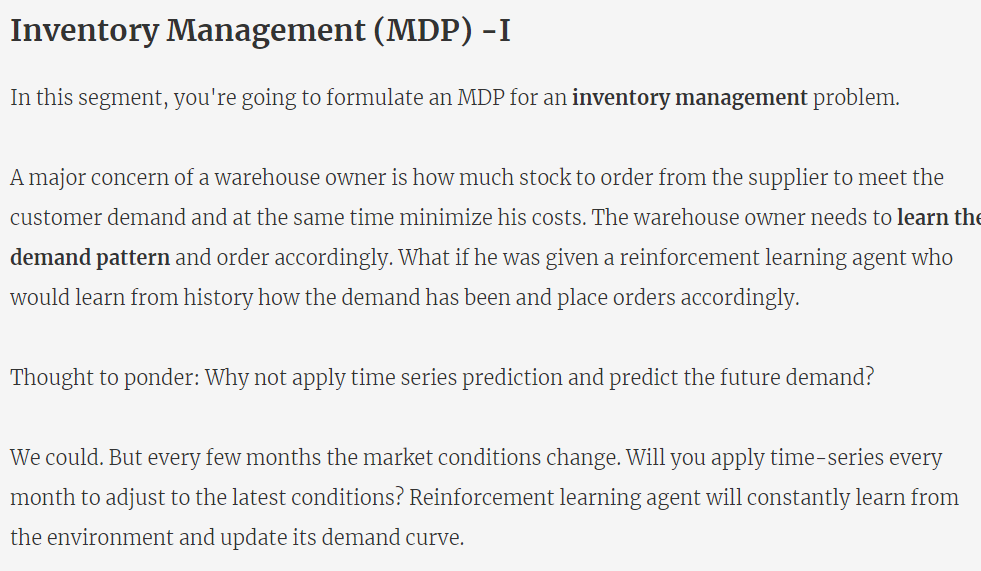


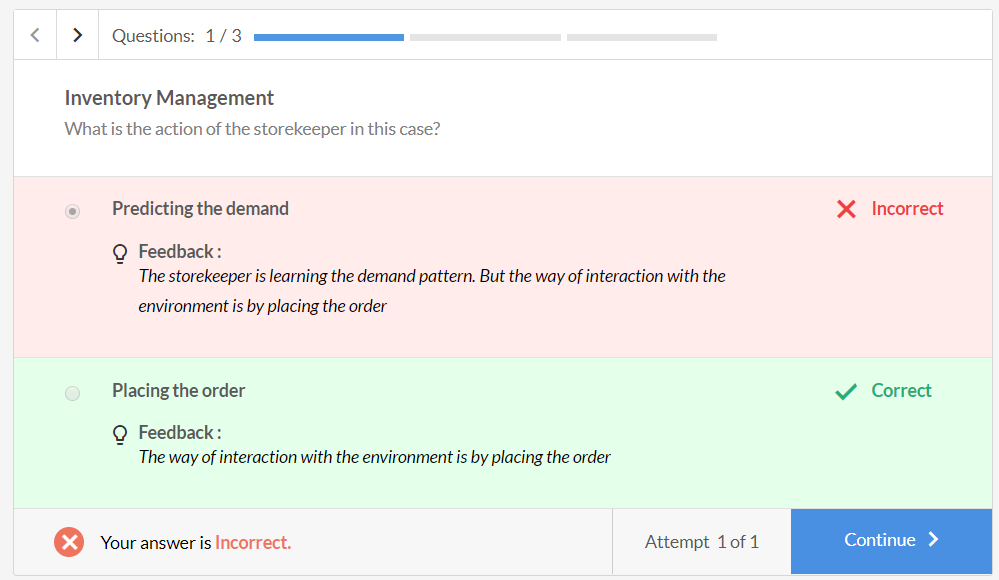


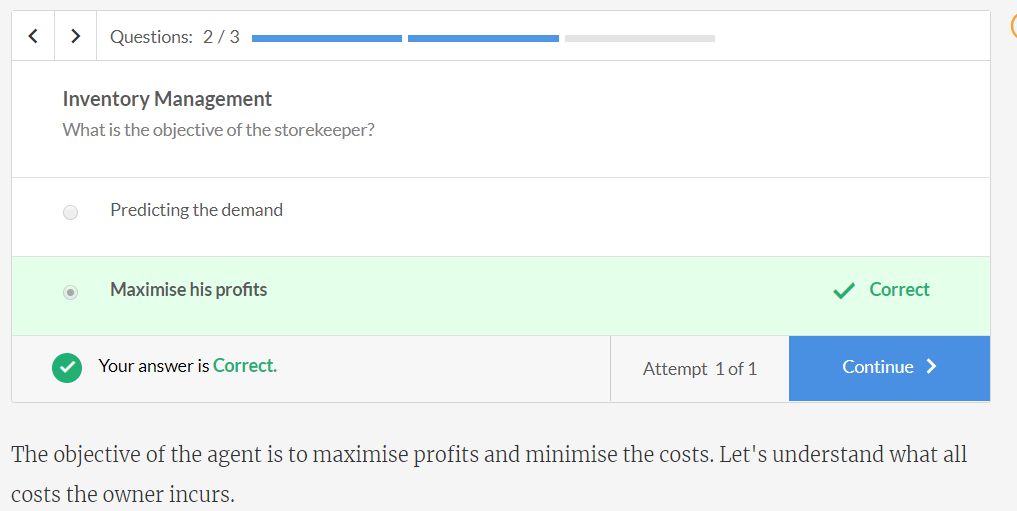


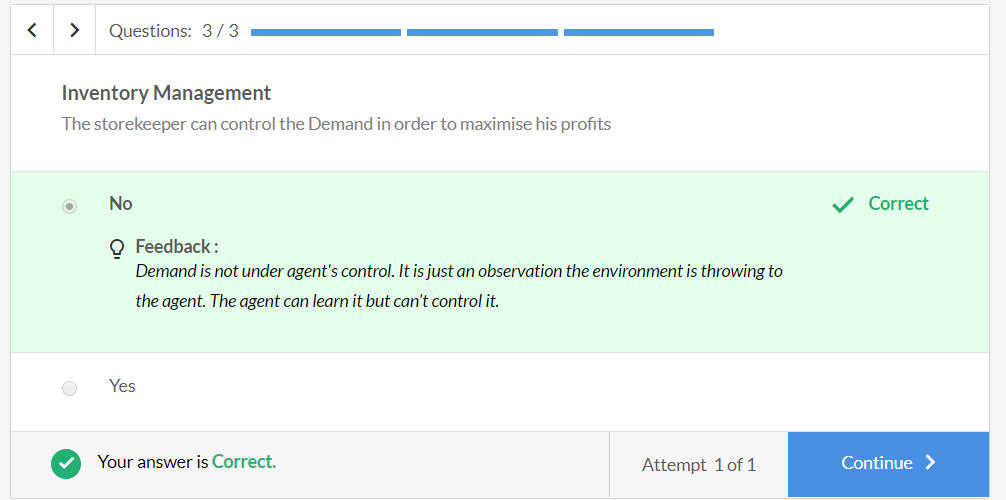


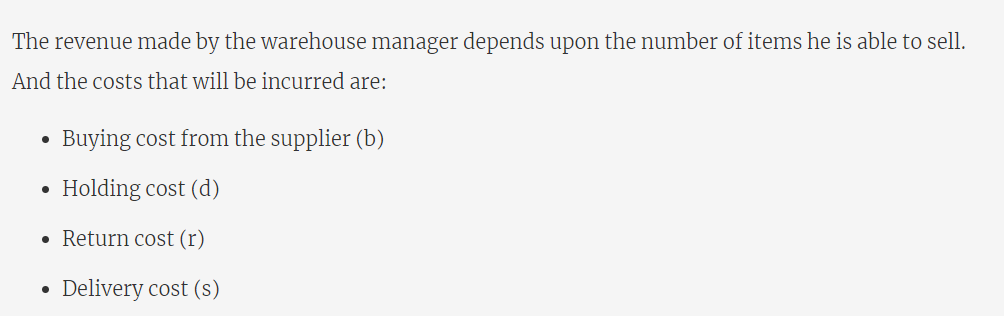


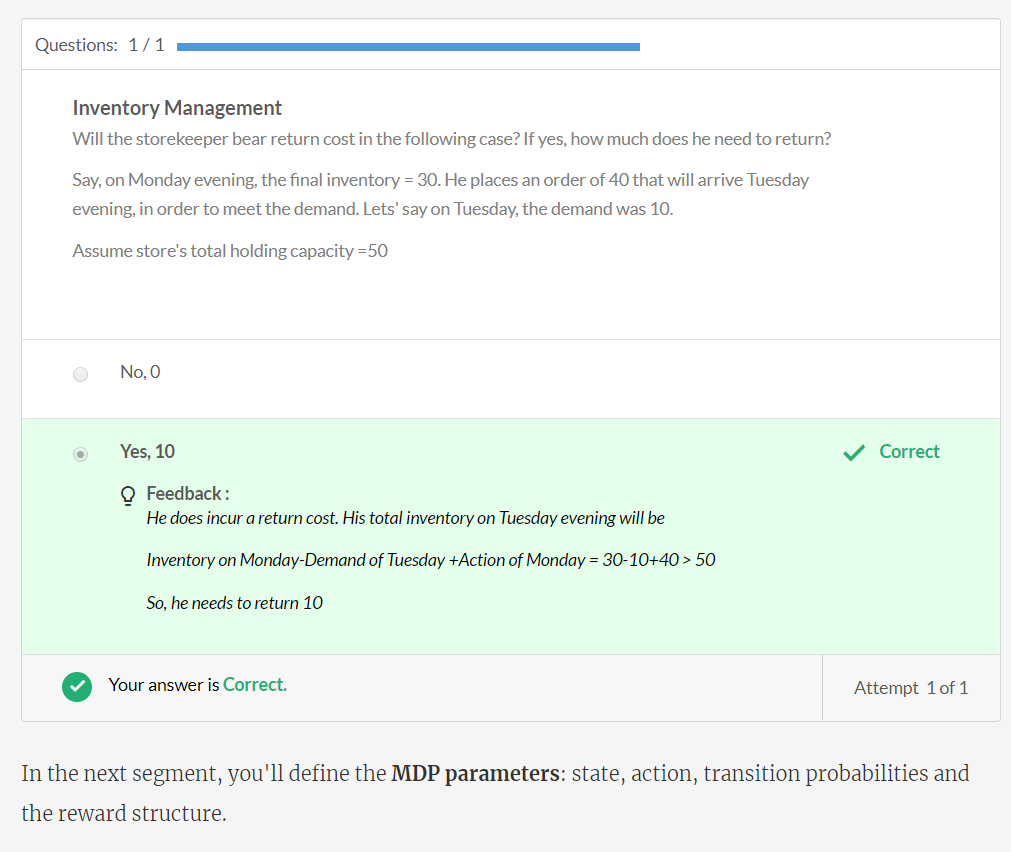


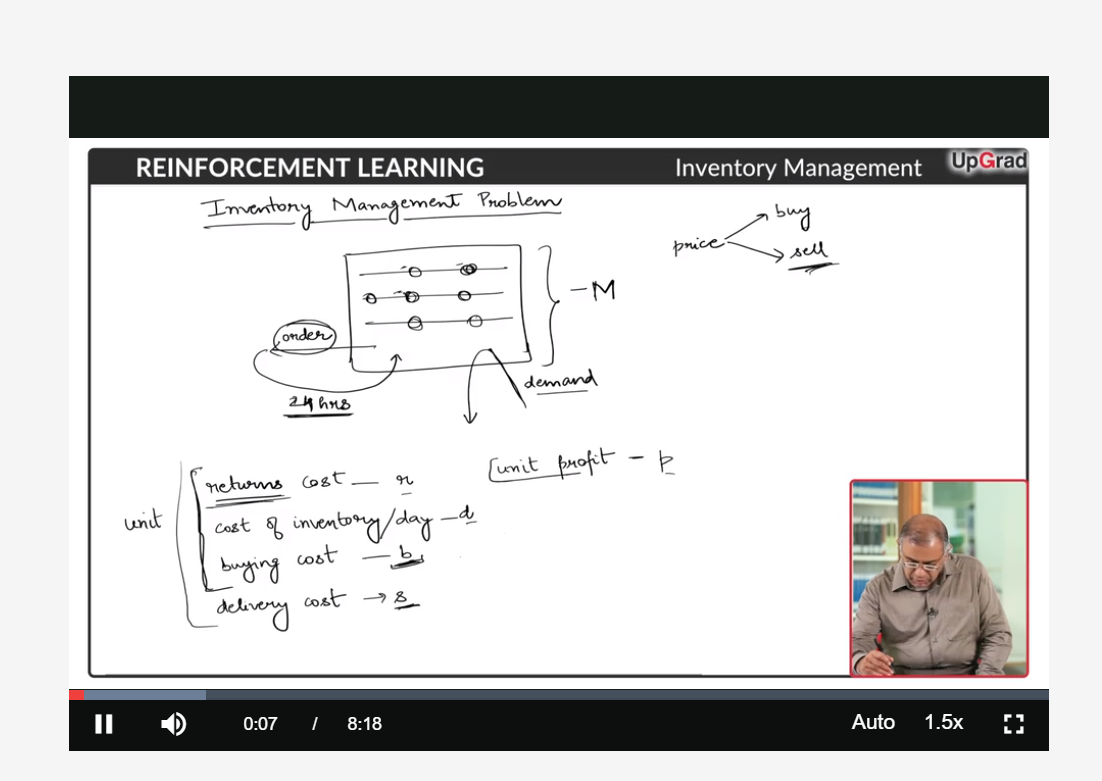


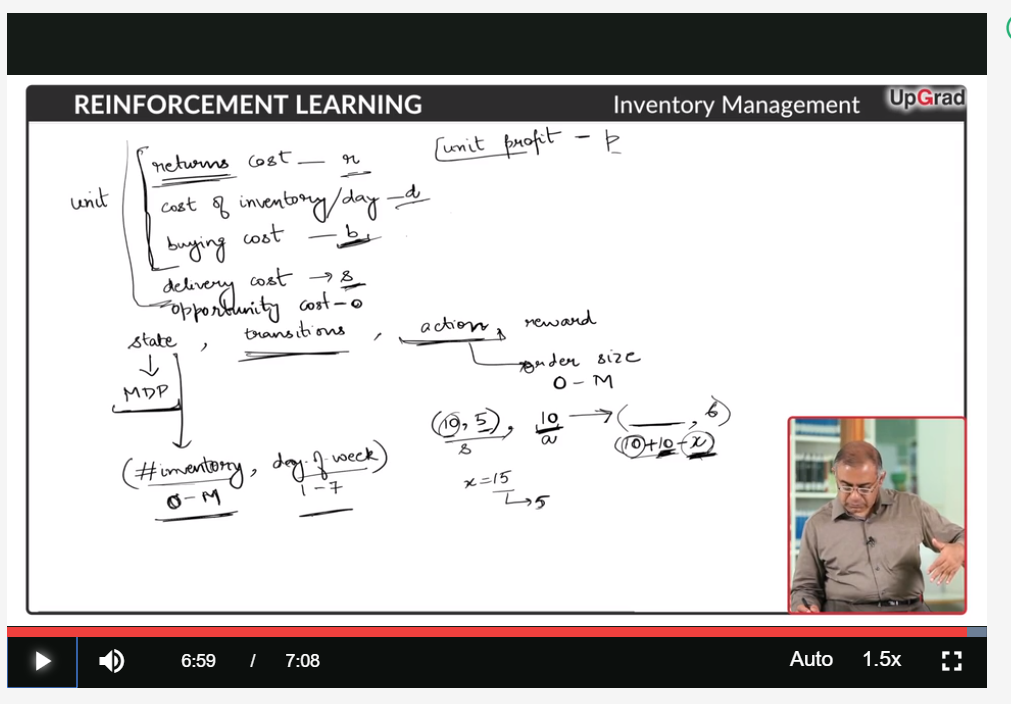


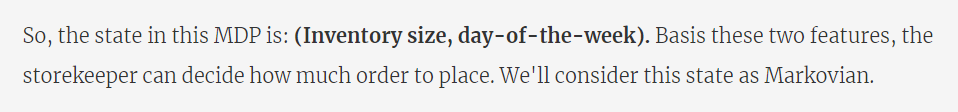


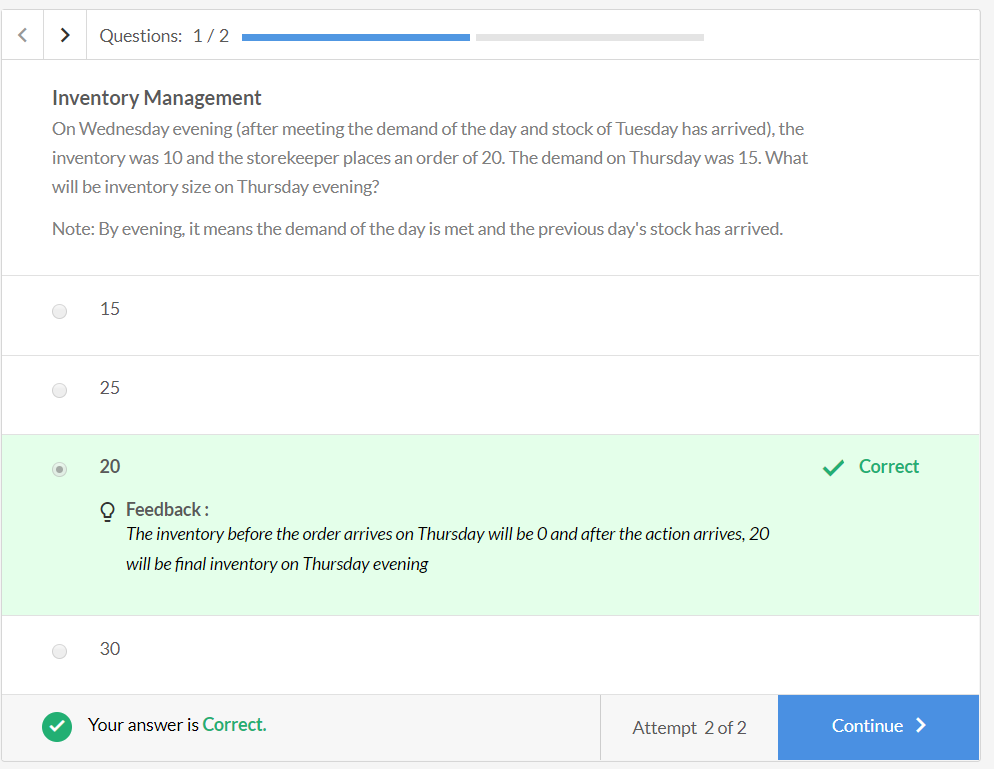


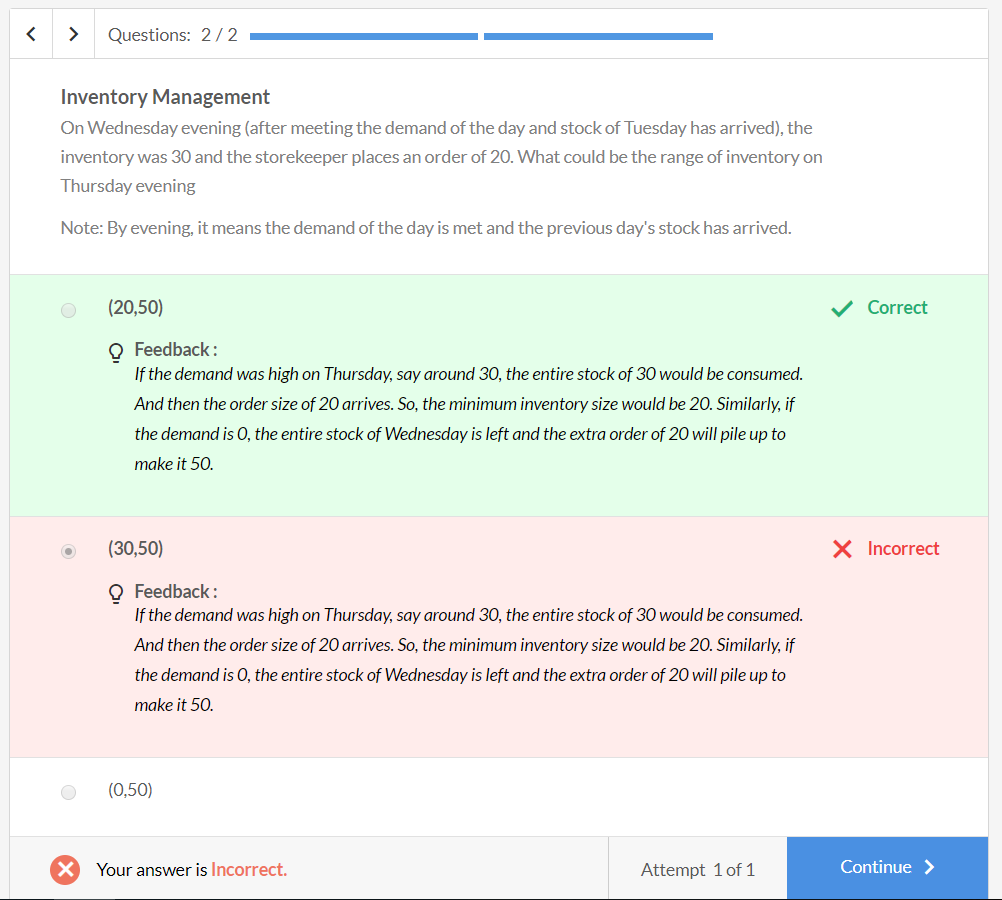


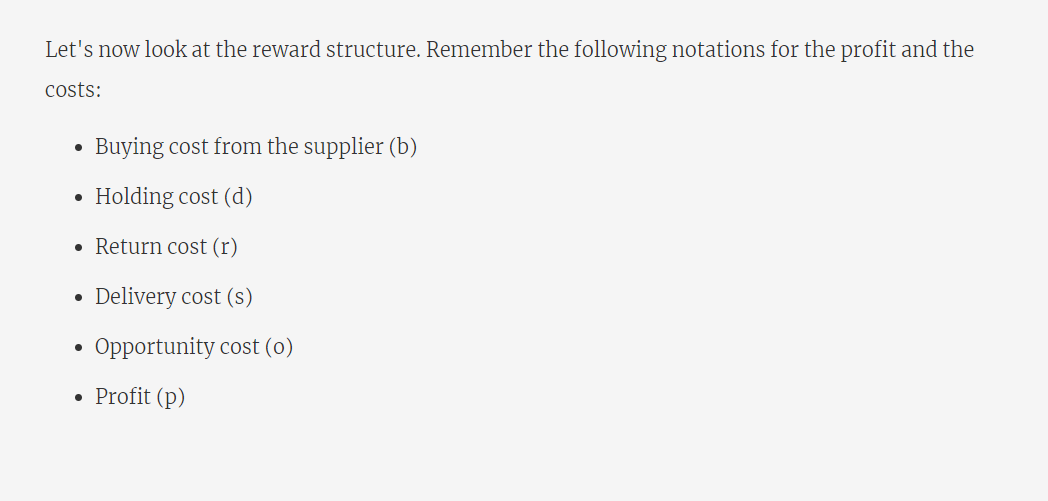


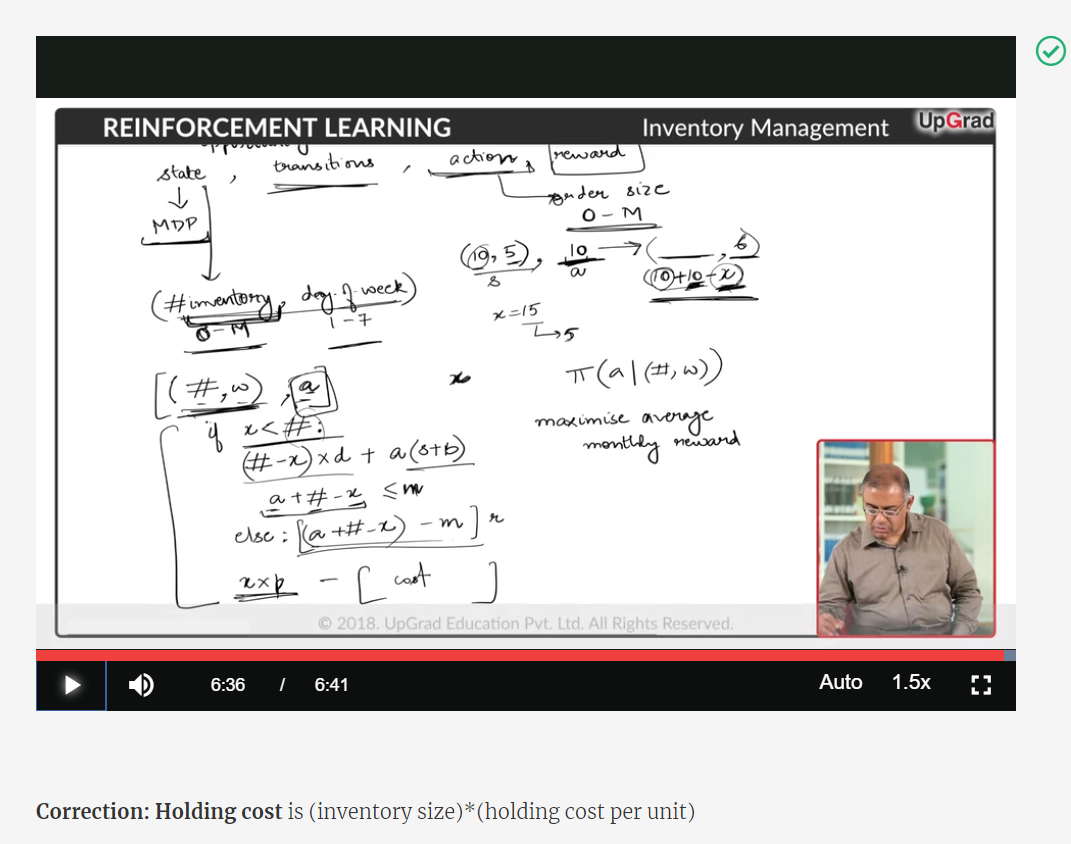


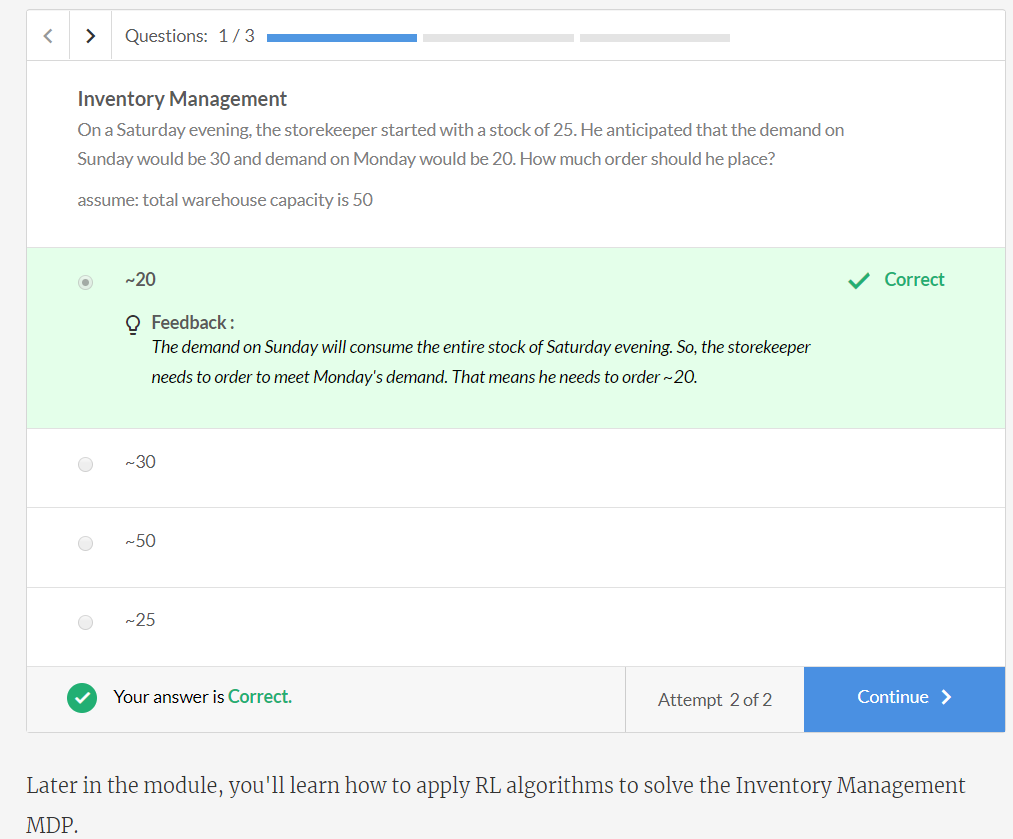


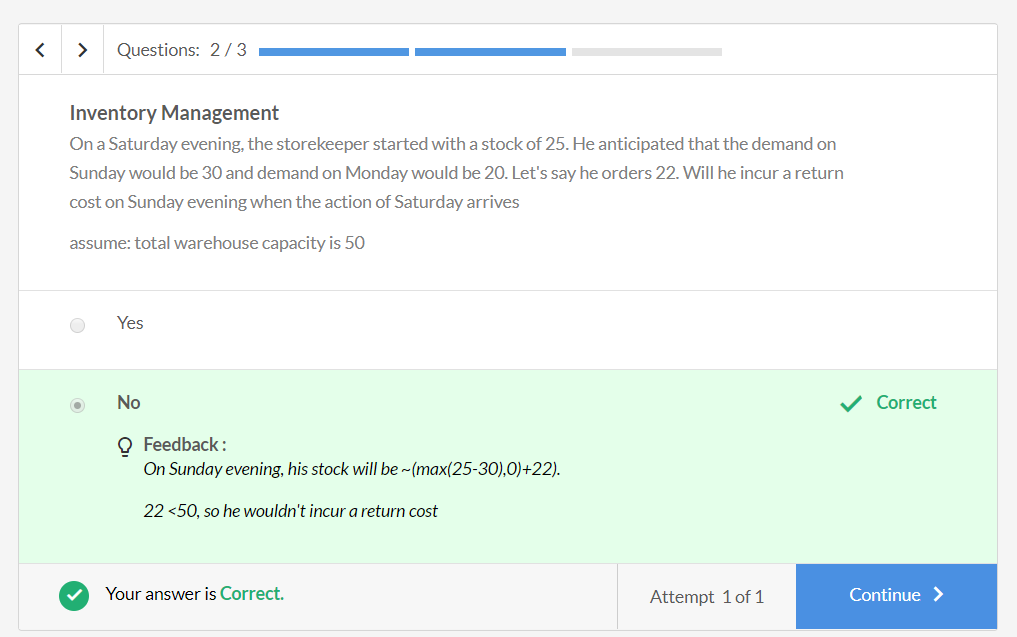


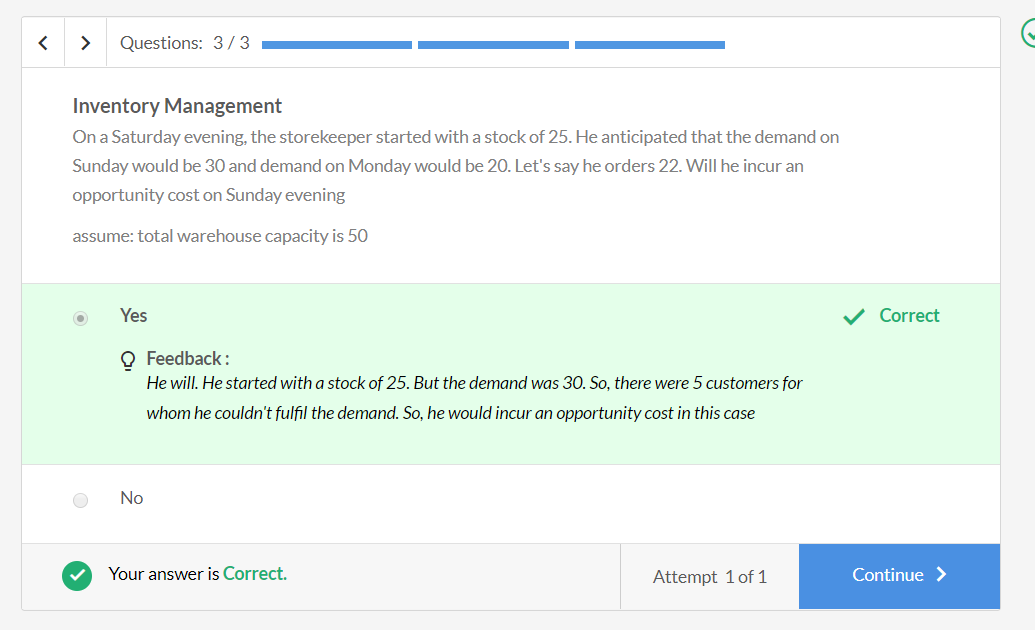


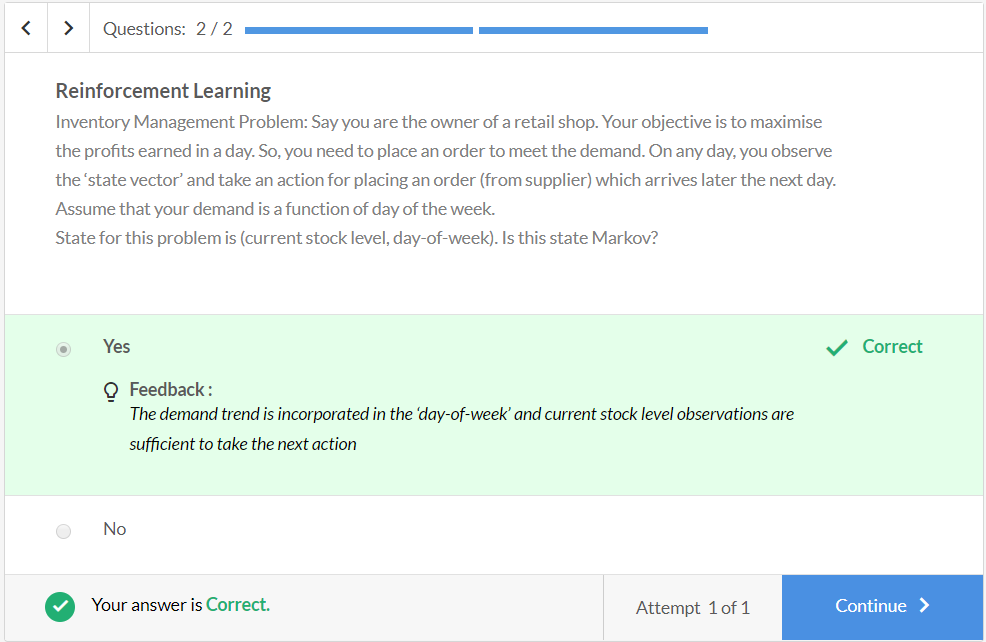












Graded Questions



