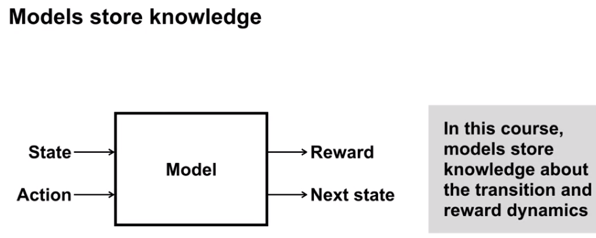
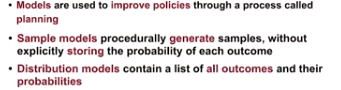
# WEEK5

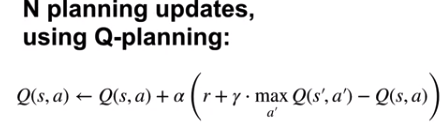
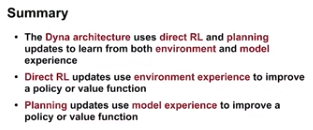
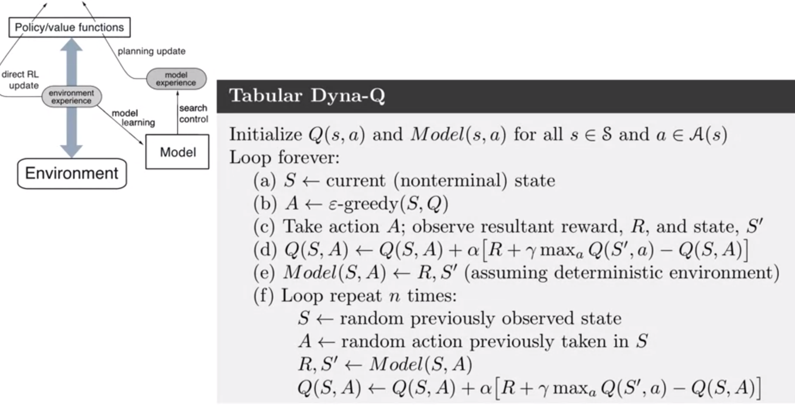
**Lesson 1: What is a model?**

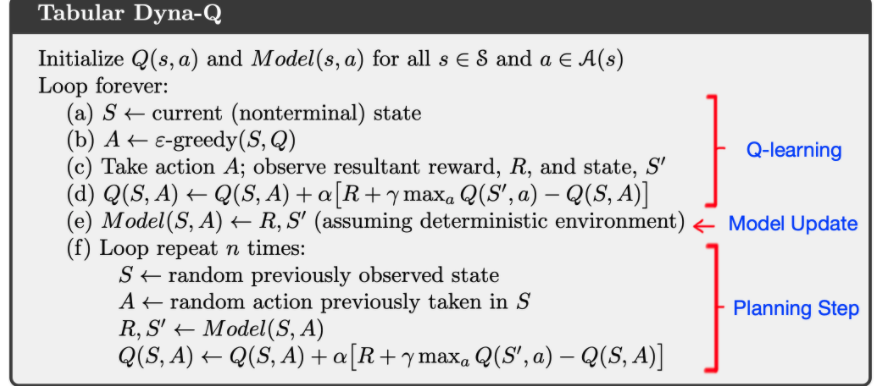
* Describe what a model is and how they can be used
* 
* Classify models as distribution models or sample models
* Identify when to use a distribution model or sample model
* 
* Describe the advantages and disadvantages of sample models and distribution models
  + Sample models require less memory
  + Distribution models can be used compute exact expected outcome.
  + Knowing exact expected outcome allows us assess risk
* Explain why sample models can be represented more compactly than distribution models

**Lesson 2: Planning**

* Explain how planning is used to improve policies
* Describe random-sample one-step tabular Q-planning

**Lesson 3: Dyna as a formalism for planning**

* Recognize that direct RL updates use experience from the environment to improve a policy or value function
* Recognize that planning updates use experience from a model to improve a policy or value function
* 
* Describe how both direct RL and planning updates can be combined through the Dyna architecture
* 
* Describe the Tabular Dyna-Q algorithm
* 



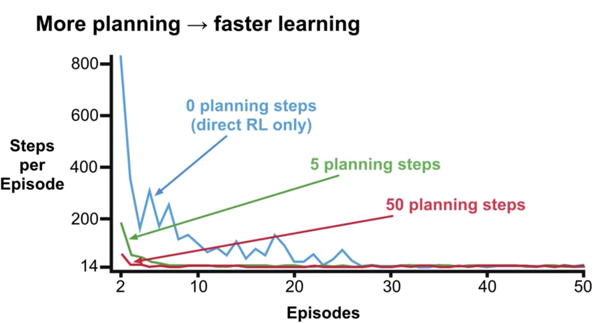
Dyna-Q involves four basic steps:

1. Action selection: given an observation, select an action to be performed (here, using the 𝜖ϵ-greedy method).
2. Direct RL: using the observed next state and reward, update the action values (here, using one-step tabular Q-learning).
3. Model learning: using the observed next state and reward, update the model (here, updating a table as the environment is assumed to be deterministic).
4. Planning: update the action values by generating 𝑛n simulated experiences using certain starting states and actions (here, using the random-sample one-step tabular Q-planning method). This is also known as the 'Indirect RL' step. The process of choosing the state and action to simulate an experience with is known as 'search control'.

Steps 1 and 2 are parts of the [tabular Q-learning algorithm](http://www.incompleteideas.net/book/RLbook2018.pdf#page=153) and are denoted by line numbers (a)â€“(d) in the pseudocode above. Step 3 is performed in line (e), and Step 4 in the block of lines (f).

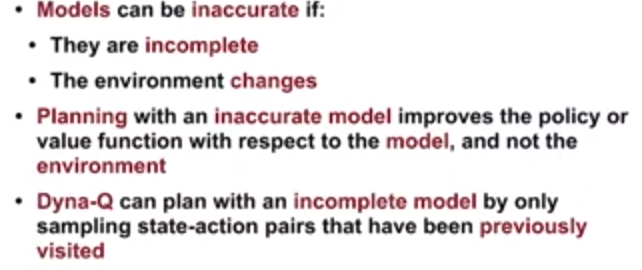
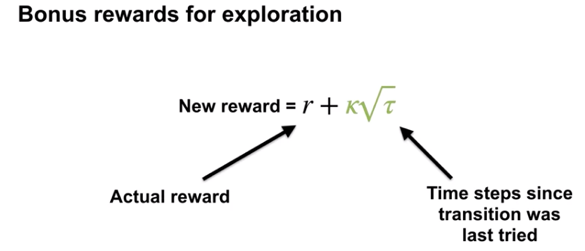
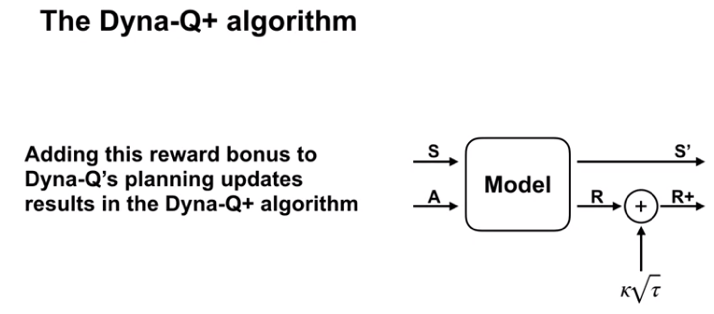
We highly recommend revising the Dyna videos in the course and the material in the RL textbook (in particular, [Section 8.2](http://www.incompleteideas.net/book/RLbook2018.pdf#page=183)).

* Identify the direct-RL and planning updates in Tabular Dyna-Q



* Identify the model learning and search control components of Tabular Dyna-Q
* Describe how learning from both direct and simulated experience impacts performance
  + Planning can dramatically speed up learnin
* Describe how simulated experience can be useful when the model is accurate

**Lesson 4: Dealing with inaccurate models**

* Identify ways in which models can be inaccurate
  + Transitions are different – incomplete model
  + Environment changes
* Explain the effects of planning with an inaccurate model
  + Model becomes outdated, makes it worse
* Describe how Dyna can plan successfully with a partially inaccurate model
  + 
* Explain how model inaccuracies produce another exploration-exploitation trade-off
  + 
  + 
* Describe how Dyna-Q+ proposes a way to address this trade-off