WEEK 5 – REINFORCEMENT LEARNING

REINFORCEMENT LEARNING

We want to learn an ideal set of behaviors to perform a certain task

Example: studying

If one study strategy == D-, we will not try that again, we will keep searching for a different one

If a second study strategy == A+, we will use this strategy as our ideal **policy**

KEY TERMS

- Environment – the variables that are observable/relevant to the task
- State – how the environment looks at a certain location/point
- Actions(s) – something we can do
- Reward functions – how we evaluate our actions
- Policy –determines which actions we will take given a certain state (we want to learn this)

DEEP REINFORCEMENT LEARNING

Uses a neural network to learn the ideal policy

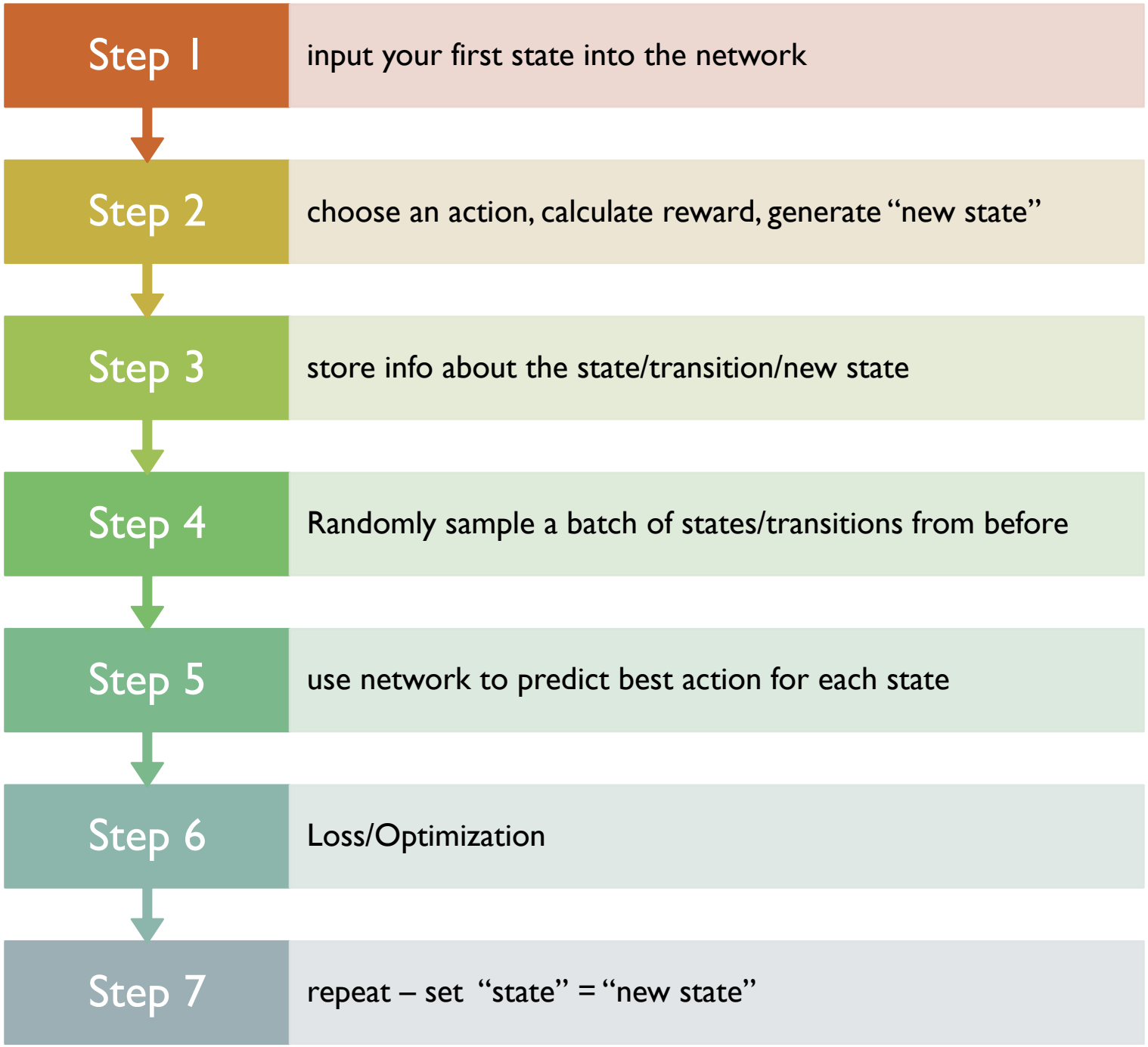
State is input into the network

Neural network predicts best action given the state

Action is taken (environment changes)

New state is input into the network - repeat

HOW DOES THIS WORK?



RANDOMNESS

Active reinforcement learning – exploration!

Sometimes, we can select random actions and use those to get our reward

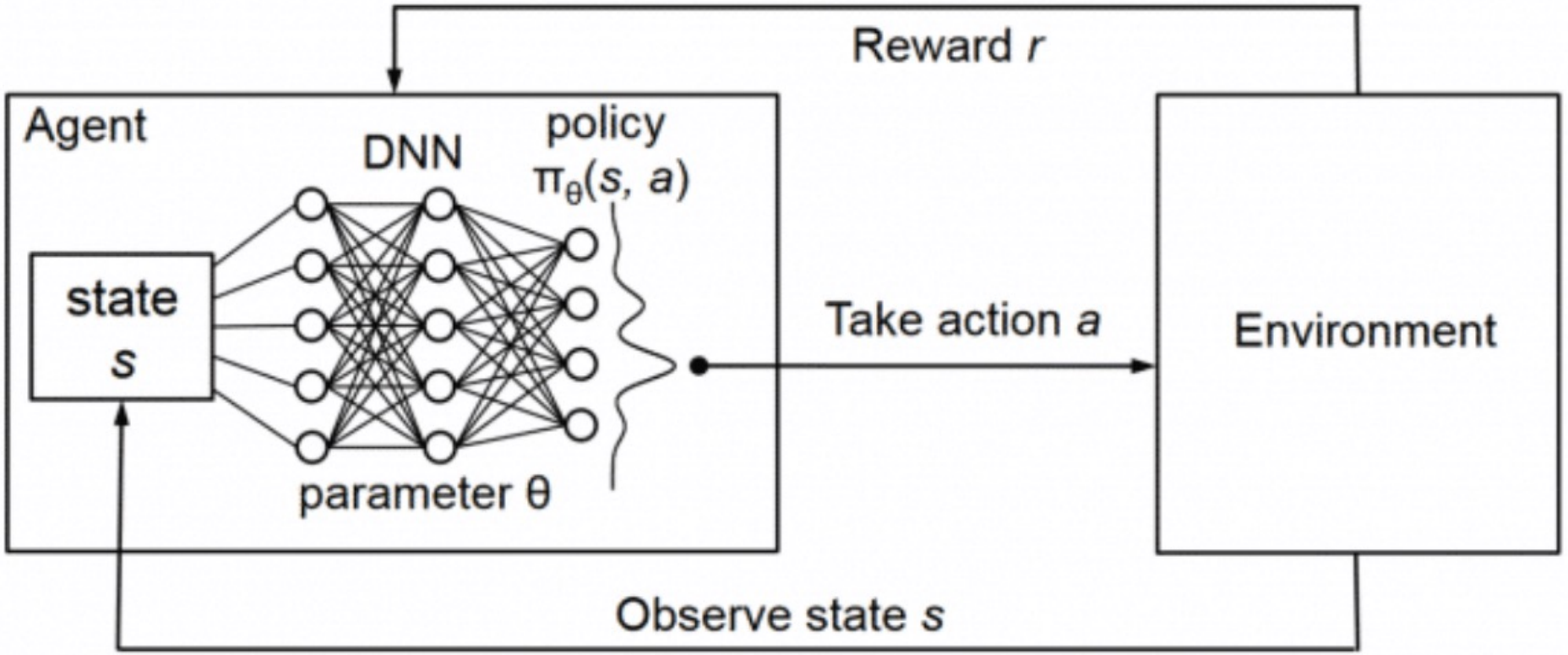
Ensure we do not converge to something mediocre (We keep trying new things!)

FAMILIAR STUFF

Feature-weight combos used to predict the best action

Loss function – tells us how we did

Optimization function – fixes weights for better policies


Reward r
Agent
DNN
policy
$\pi_\theta(s, a)$
state
s
Take action a
Environment
parameter θ
Observe state s

Flappy Bird: A Case Study