Agent
state
s_t
reward
r_t
action
a_t
r_{t+1}
s_{t+1}
Environment