State S_t

reward R_t

Agent

action A_t

R_{t+1}

S_{t+1}

Environment