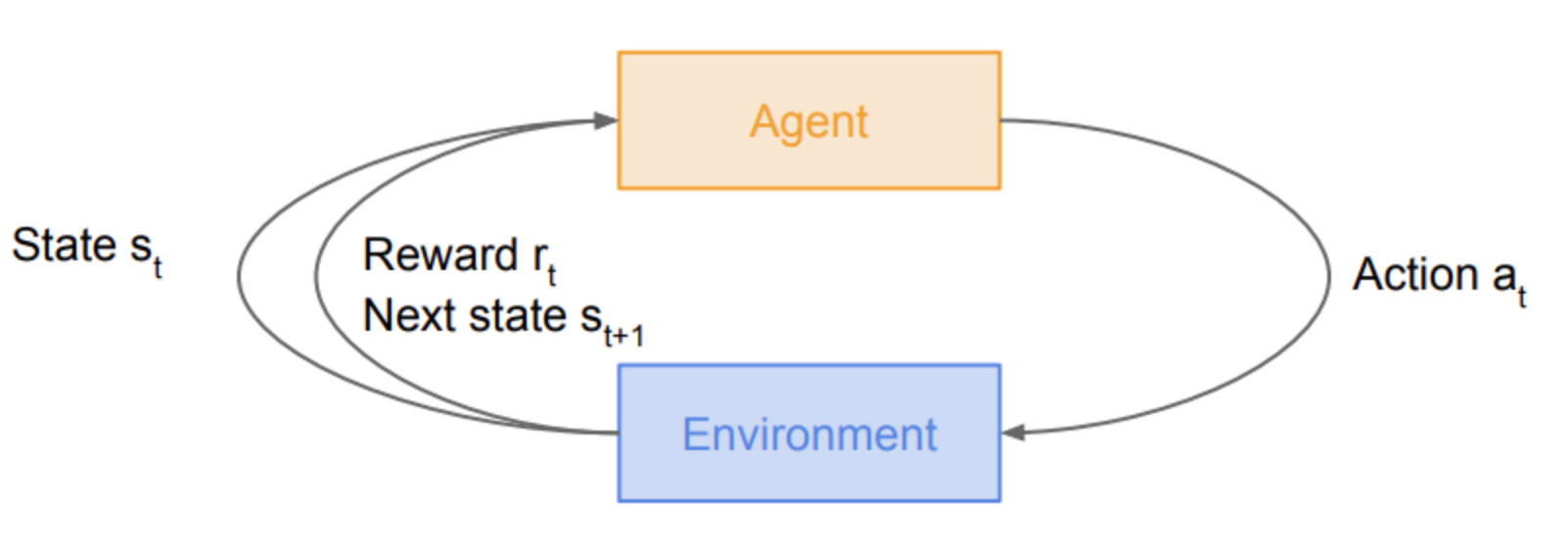
**14 - Deep Reinforcement Learning**



강화학습은 env와 agent 간의 상호작용에서 학습이 이루어짐

텍스트이(가) 표시된 사진

자동 생성된 설명

markov property?

: 미래의 state가 현재의 state에만 dependent하다 + also reward Markov

(원래는 과거의 것도 영향있지만, 과거 상관없이 현재의 것만)

텍스트이(가) 표시된 사진

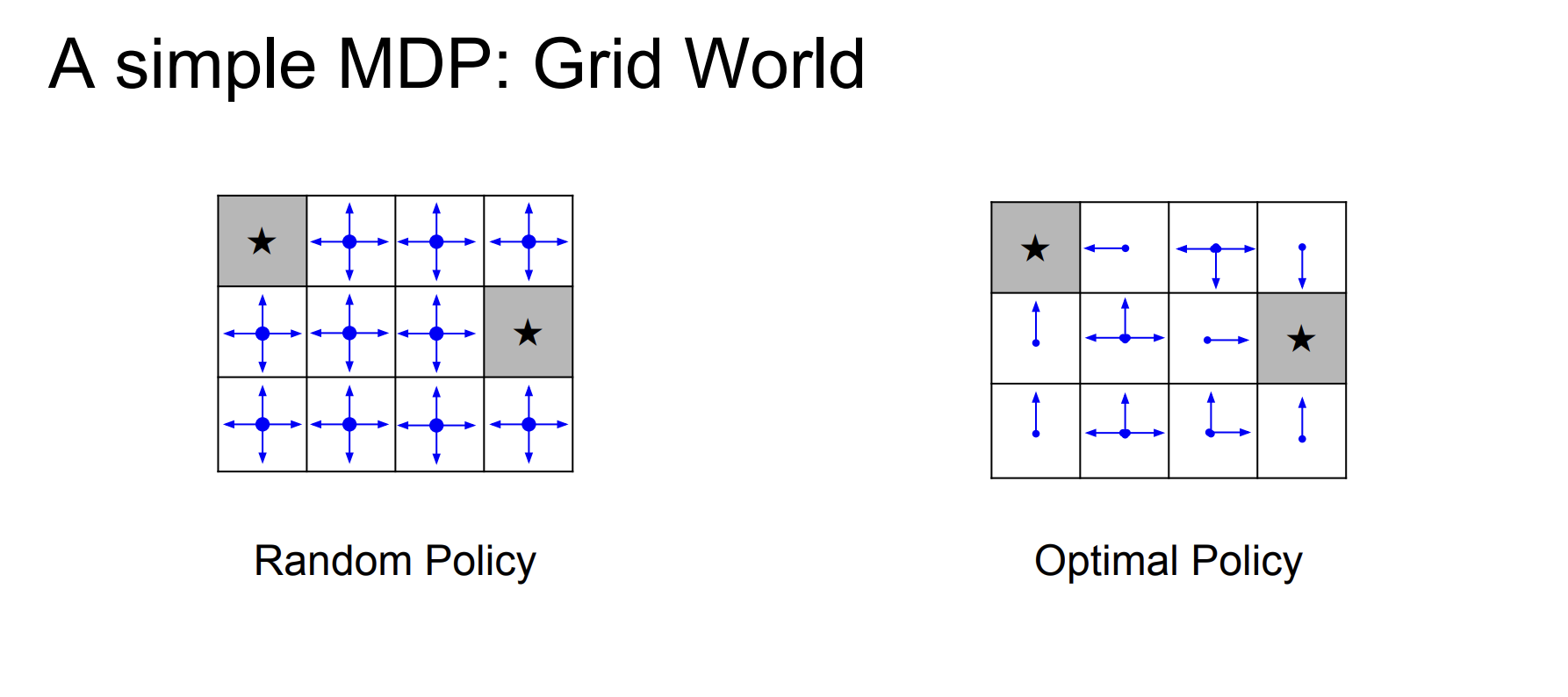
자동 생성된 설명

(1) 각 time step마다 Agent는 Environment로부터 state $s\_t$를 관측함

(2) 관측의 결과로부터 action $a\_t$ 취함

(3) action을 취한 대가로 reward $r\_t$를 받음

(4) state가 다음으로 바뀜 $s\_{t+1}$



텍스트이(가) 표시된 사진

자동 생성된 설명

<Q-learning>

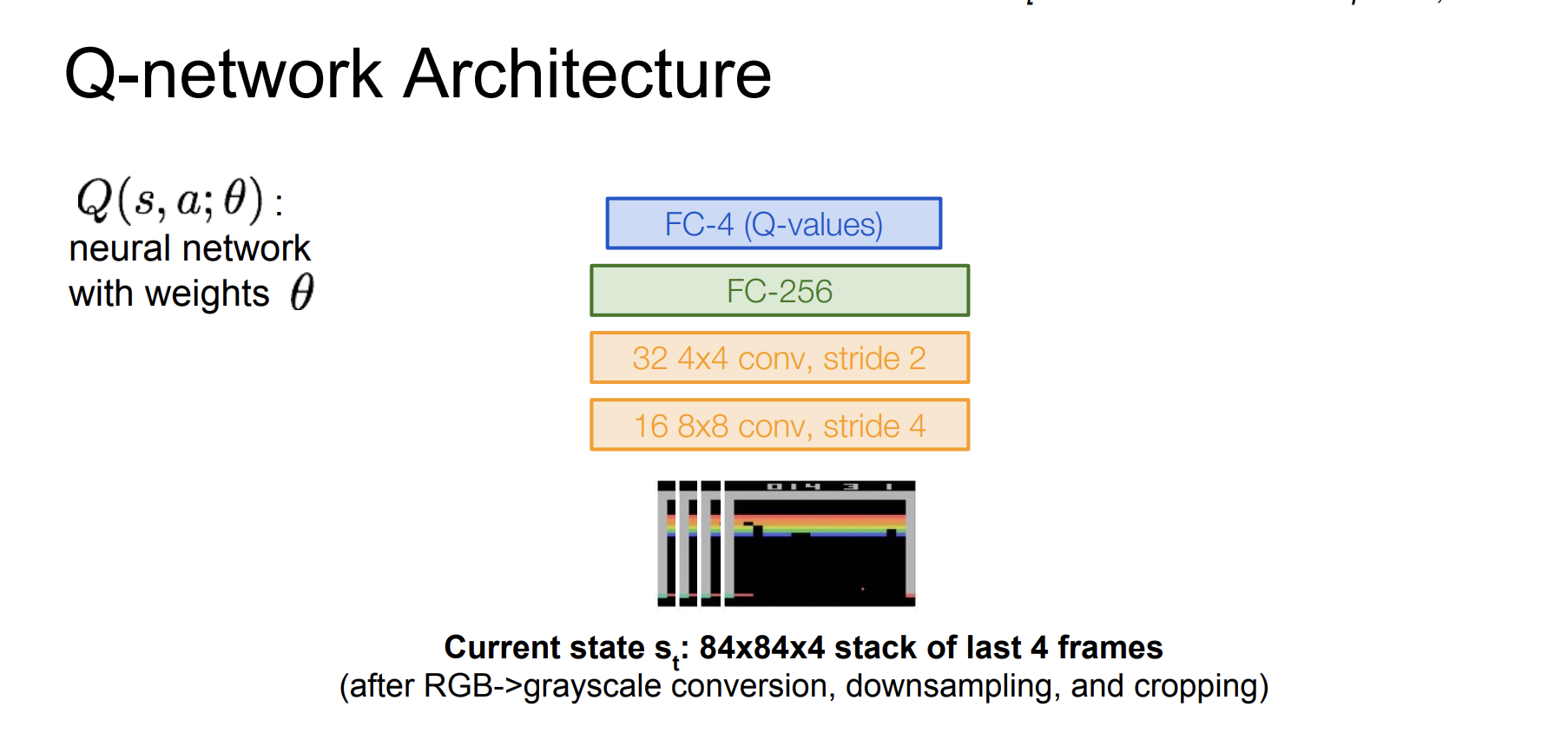
만일 $P(s\_t|s\_{t-1},a\_{t\_1})$를 모른다면?

예를 들어 P(A|B,up)처럼 B에서 up하면 A라는 것을 현실에선 알 수 없음

⇒ Q값을 도입하고, 이러한 Q값을 찾는 것을 Q Learning이라 함

텍스트이(가) 표시된 사진

자동 생성된 설명

텍스트이(가) 표시된 사진

자동 생성된 설명