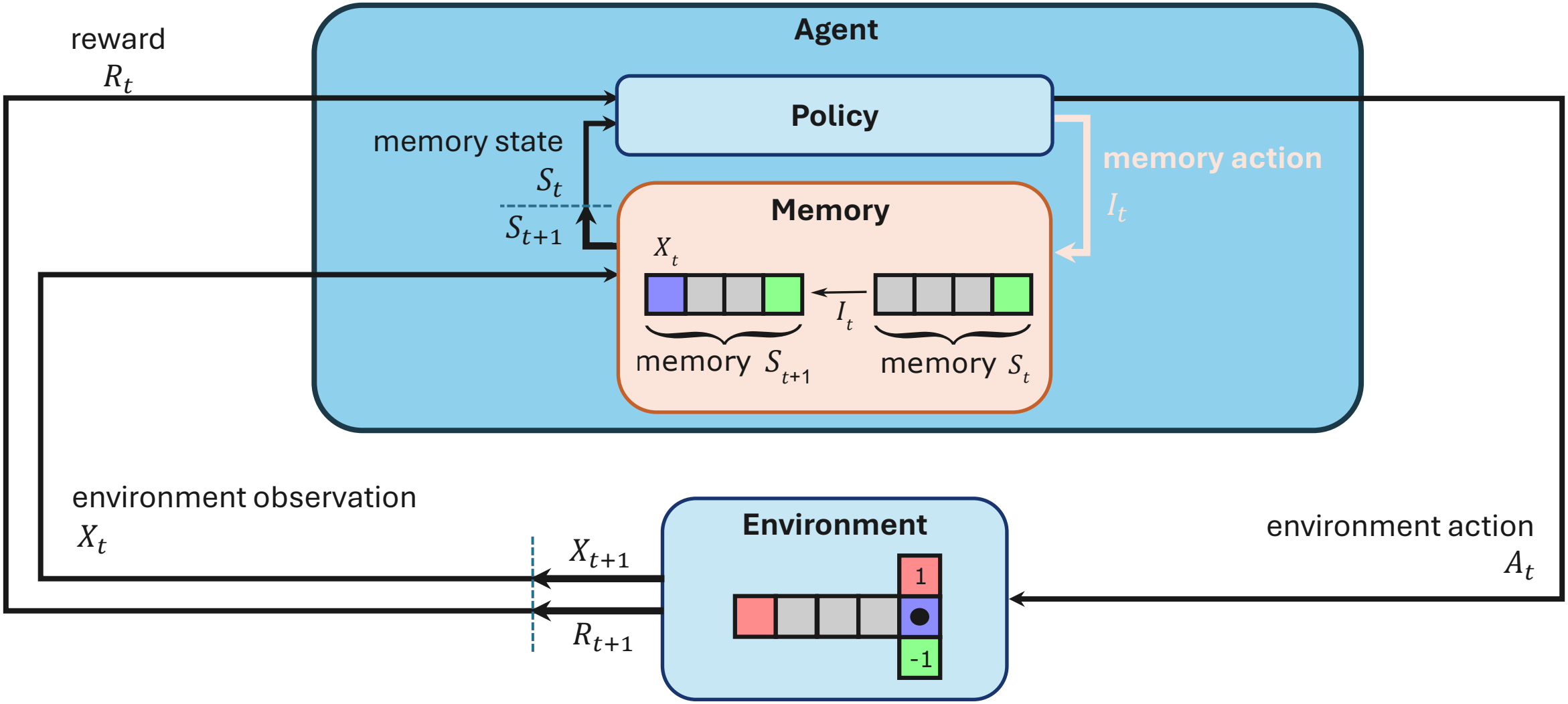
reward
R_t
Agent
Policy
memory state
S_t
S_{t+1}
memory action
I_t
Memory
X_t
I_t
memory S_{t+1}
memory S_t
environment observation
X_t
X_{t+1}
R_{t+1}
Environment
1
-1
environment action
A_t