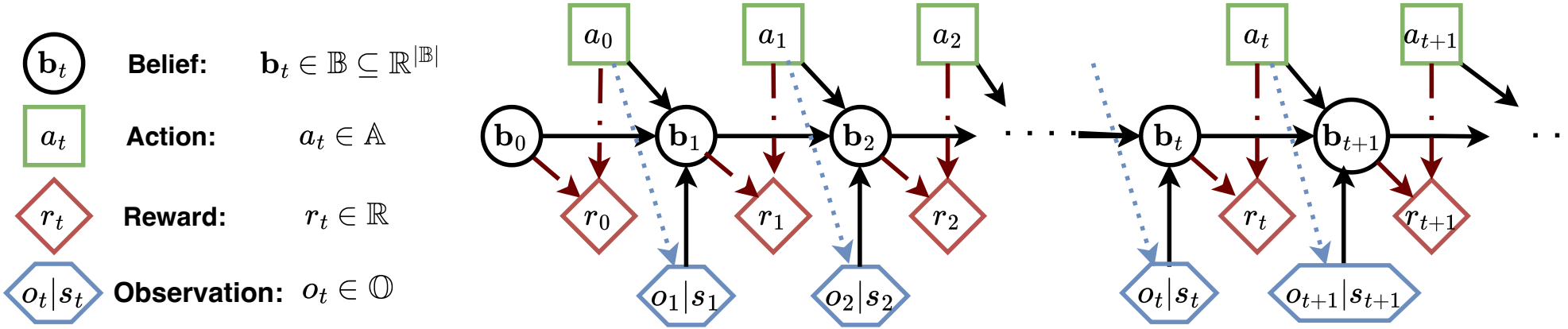

Agent

$\mathbf{b}_t = Prob.\ Density(\mathbb{S})$ at time step t
$b_{t+1}(s_{t+1}) = P(s_{t+1}|o_{t+1}, a_t, \mathbf{b}_t) =$ -

Belief Update $P(\mathbf{b}_{t+1}|\mathbf{b}_t, a_t) = \int_{o_{t+1} \in \mathbb{O}} P(o_{t+1}|\mathbf{b}_t, a_t) do_{t+1}$

Policy $a_{t+1} \sim \pi(a_{t+1}|\mathbf{b}_{t+1})$

Observation Model $o_{t+1} \sim P(o_{t+1}|\mathbf{b}_t, a_t)$

Reward Model $r_t = r(\mathbf{b}_t, a_t) = \int_{s_t \in \mathbb{S}} r(s_t, a_t) b_t(s_t) ds_t$

Environment

True State update $s_{t+1} \sim P(s_{t+1}|s_t, a_t)$

o_{t+1} Observation Model $o_{t+1} \sim P(o_{t+1}|s_t, a_t)$

r_t Reward Model $r_t = r(s_t, a_t)$