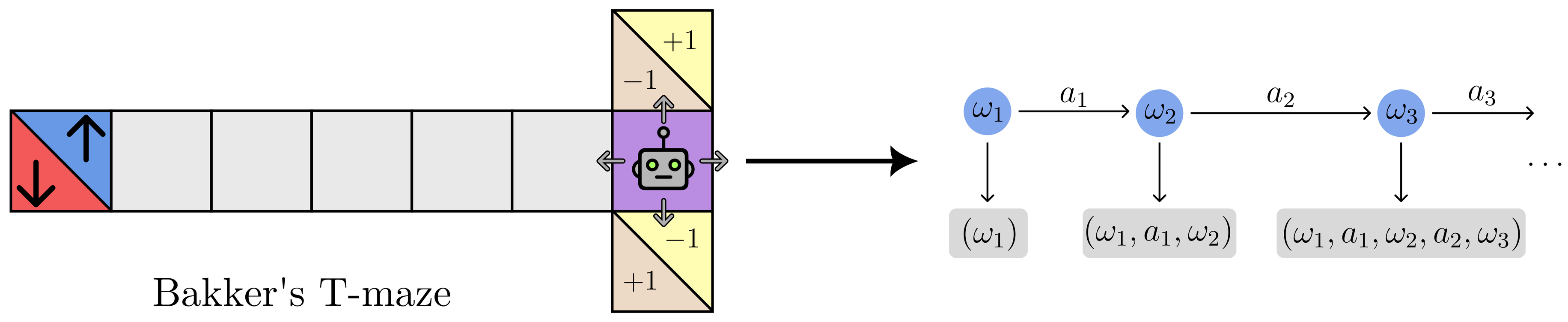
Viewing memory as a state abstraction over trajectories helps classify POMDPs

Memory as State Abstraction over Trajectories

Aaron Kirtland,* Alexander Ivanov,* Cameron Allen, Michael L. Littman, George Konidaris
*Equal Contribution

1

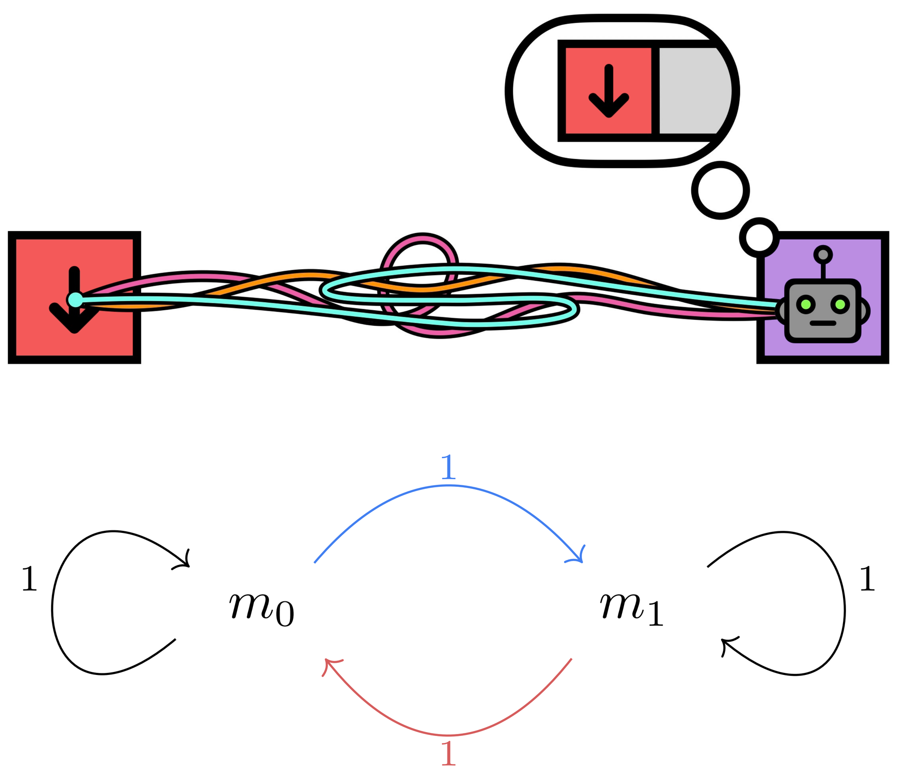
Remembering the entire trajectory turns any POMDP into a trajectory MDP.

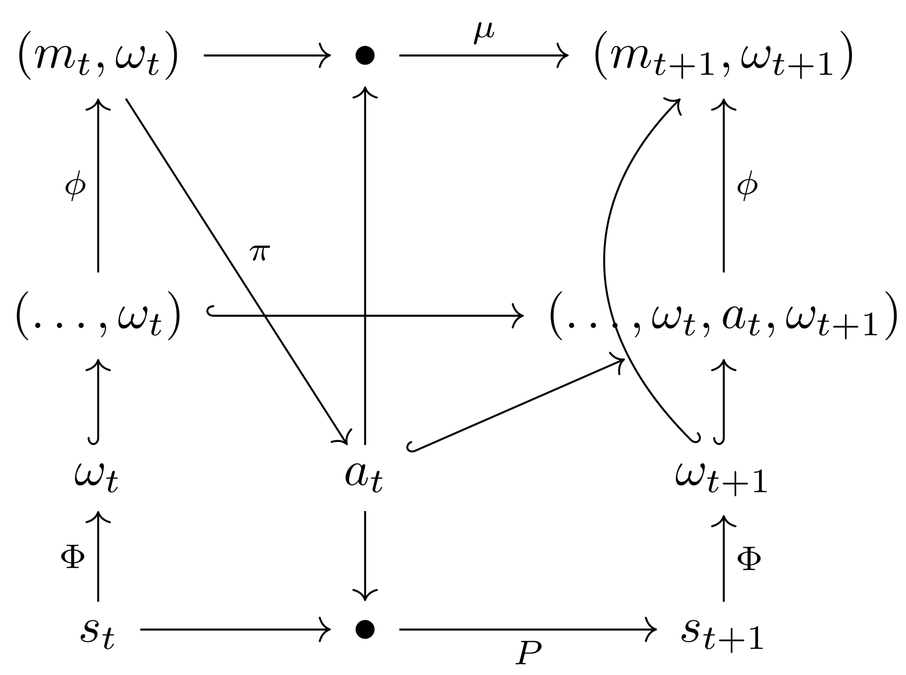


But no one wants to use the entire history, so we use memory instead.

2

Memory functions induce abstractions over trajectories.





Target	Exact State Abstractions	Approximate State Abstractions
model	$\forall s^{(1)}, s^{(2)}.\forall \bar{s}.\forall a.\phi(s^{(1)}) = \phi(s^{(2)}) \Rightarrow$ $R(s^{(1)}, a) = R(s^{(2)}, a)$ $\sum_{s' \in \phi^{-1}(\bar{s})} P(s' \mid s^{(1)}, a) = \sum_{s' \in \phi^{-1}(\bar{s})} P(s' \mid s^{(2)}, a)$	$\exists f_P : \phi(T) \times A \to \Delta\Omega$ $\lVert [f_P]_\phi - P_o \rVert_1 < \epsilon_P$ $\exists f_R : \phi(T) \times A \to \mathbb{R}$ $\lVert [f_R]_\phi - R \rVert_\infty < \epsilon_R$
Q^*	$\forall s, s'.\forall a.\phi(s) = \phi(s') \Rightarrow$ $Q^*(s, a) = Q^*(s', a)$	$\exists f : \phi(S) \times A \to \mathbb{R}$ $\lVert [f]_\phi - Q^*_M \rVert_\infty \le \epsilon_{Q^*}$
π^*	$\forall s, s'.\exists a^*.\phi(s) = \phi(s') \Rightarrow$ $Q^*(s, a^*) = \max_a Q^*(s, a)$ $\max_a Q^*(s', a) = Q^*(s', a^*)$	$\exists \pi : \phi(S) \to \Delta A$ $\left\lVert V_M^{[\pi]_\phi} - V_M^* \right\rVert_\infty \le \epsilon_{\pi^*}$

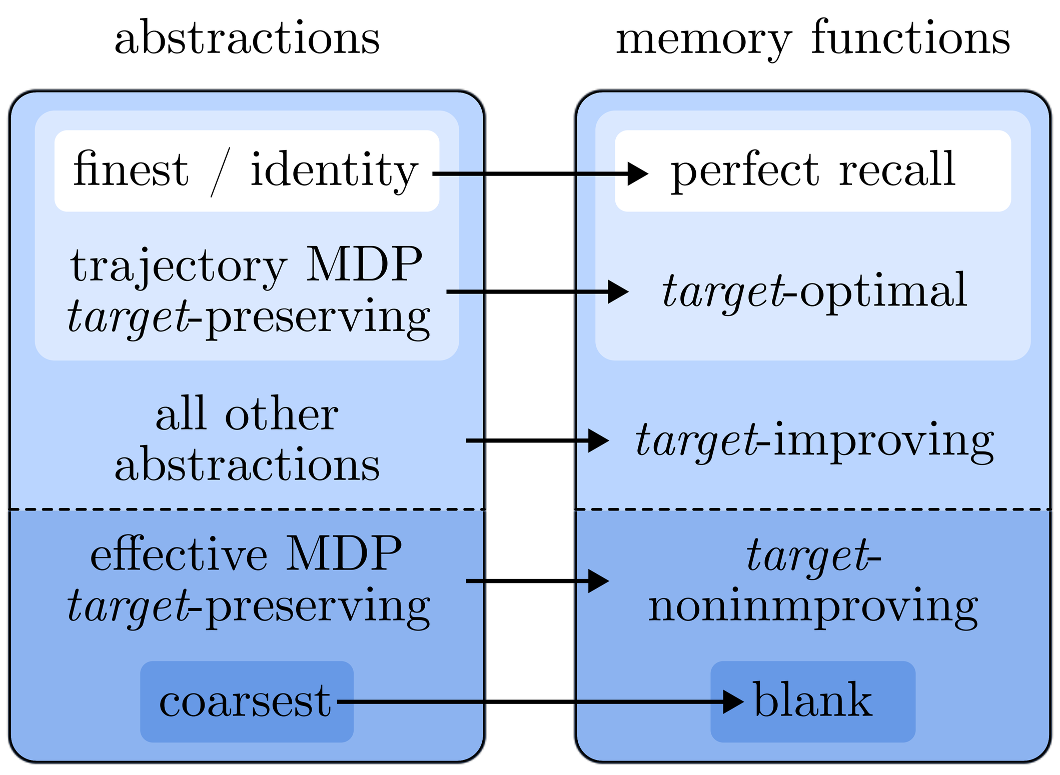
Type	Optimal	Improving
model	$\exists f_P : \phi(T) \times A \to \Delta\Omega. \lVert [f_P]_\phi - P_o \rVert_1 \le \epsilon_P$ $\exists f_R : \phi(T) \times A \to \mathbb{R}. \lVert [f_R]_\phi - R \rVert_\infty \le \epsilon_R$	$\forall f_P : \phi(T) \times A \to \Delta\Omega. \left\lVert [f_P]_\phi - \hat{P}_o \right\rVert_1 > \epsilon_P$ $\forall f_R : \phi(T) \times A \to \mathbb{R}. \left\lVert [f_R]_\phi - \hat{R} \right\rVert_\infty > \epsilon_R$
Q^*	$\exists f : \phi(T) \times A \to \mathbb{R}. \lVert [f]_\phi - Q^*_M \rVert_\infty \le \epsilon_{Q^*}$	$\forall f : \phi(T) \times A \to \mathbb{R}. \left\lVert [f]_\phi - Q^*_{\hat{M}} \right\rVert_\infty > \epsilon_{Q^*}$
π^*	$\exists \pi : \phi(T) \to \Delta A. \left\lVert V_M^{[\pi]_\phi} - V_M^* \right\rVert_\infty \le \epsilon_{\pi^*}$	$\forall \pi : \phi(T) \to \Delta A. \left\lVert V_{\hat{M}}^{[\pi]_\phi} - V_{\hat{M}}^* \right\rVert_\infty > \epsilon_{\pi^*}$

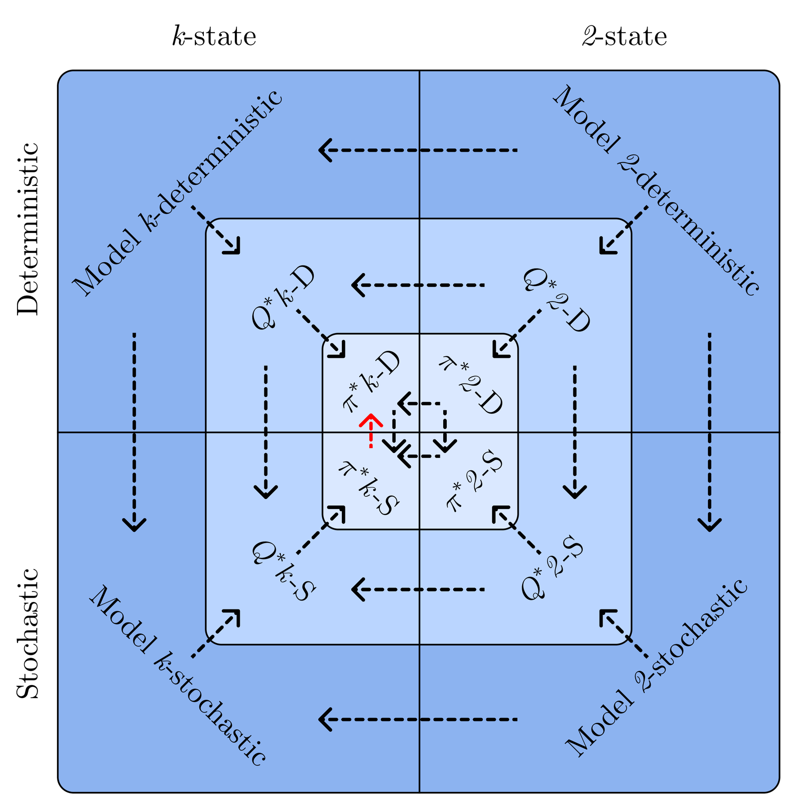
3

We can use state abstraction to define classes of POMDPs based on:

- Preservation target (π^*, Q^*, model)
- Stochasticity (deterministic, stochastic)
- Number of states (2, k)
- Quality (optimal, improving)

The state abstraction hierarchy holds!





4

Two additional findings:

With bounded reward, deterministic memory can approach stochastic memory

Stochastic memory may be strictly more powerful than deterministic memory