Temporally Abstract Partial Models


Khimya Khetarpal


Zafarali Ahmed


Gheorghe Comanici

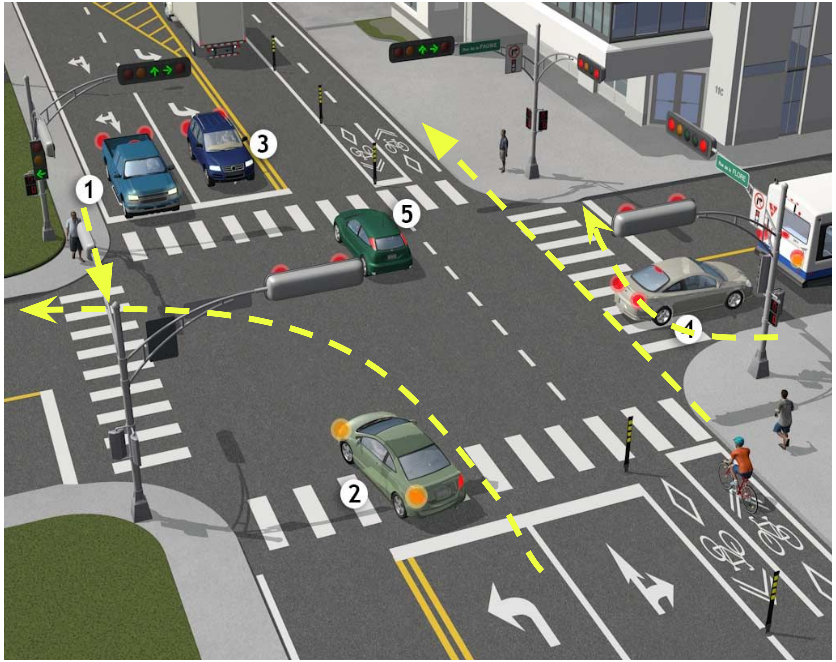

Doina Precup

McGill Mila DeepMind

“

Theories of embodied cognition and perception suggest that humans are able to represent the world knowledge in the form of *internal models* across *different time scales*.

Pezzulo & Cisek, 2016



Motivation

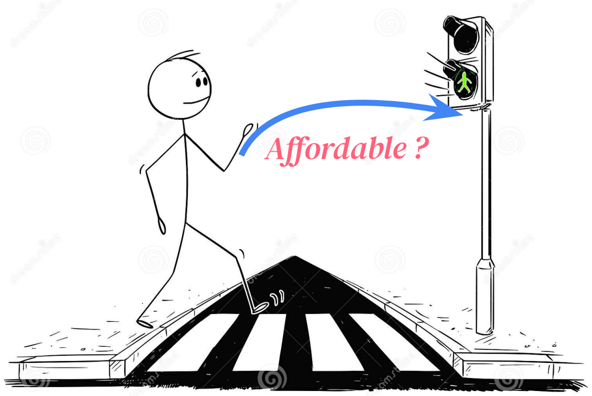
Building internal models across different time scales would allow

- Faster Learning
- Efficient Planning
- Ability to make predictions across different scales

Existing literature considers

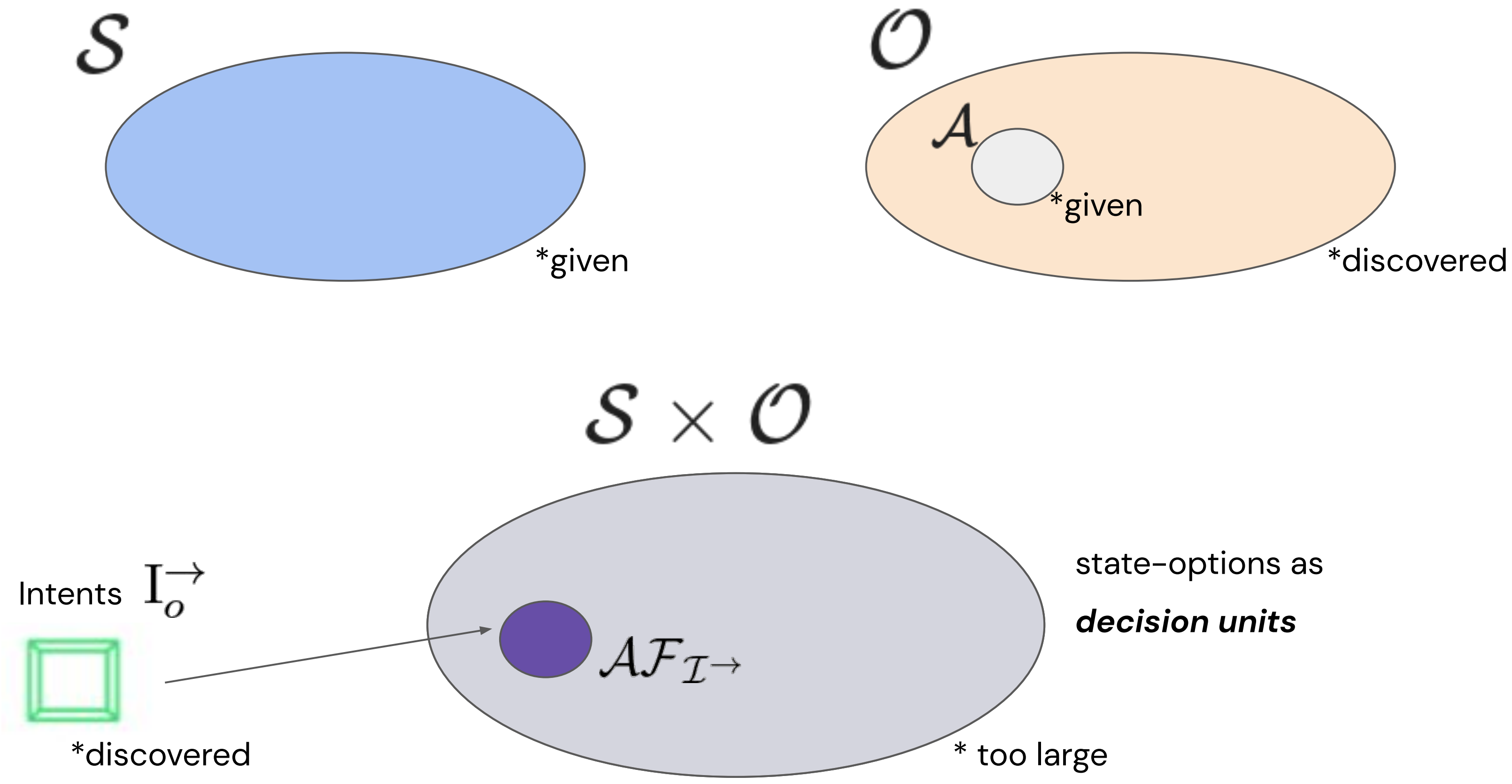
- Single-step model learning which is challenging – accumulates error!
- Model based RL where models are built over entire state-action space – intractable!
- Learning & planning with options that apply everywhere – no spatial specialization!

Key Contributions



- We extend option models to account for *affordances*.
- We establish *a theoretical understanding* of the trade-offs associated with using options vs. actions jointly with affordances.
- *Empirically demonstrate* end-to-end learning of affordances and partial option models in a function approximation setting.

Key Concepts



$\mathcal{S}$
*given
$\mathcal{O}$
$\mathcal{A}$
*given
*discovered
$\mathcal{S} \times \mathcal{O}$
Intents $\mathrm{I}_o^{\rightarrow}$
*discovered
$\mathcal{AF}_{\mathcal{I}^{\rightarrow}}$
state-options as
decision units
* too large

Temporally Extended Intent $\mathrm{I}_o^{\rightarrow}$

- Describes the **intended result** of executing ***option o*** *in state s*

- The associated intent model is denoted by

$$I_o^{\rightarrow}(s,\tau) = P_I(\tau|s,o)$$

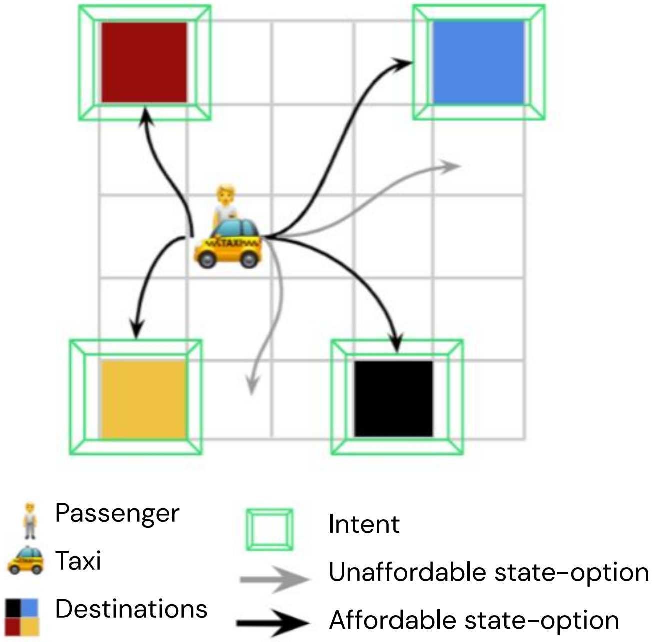
- Intent is satisfied to a degree $\zeta_{s,o}$ if and only if:

$$d(P_I(\tau|s,o), P(\tau|s,o)) \leq \zeta_{s,o}$$

Metric between probability distributions

Trajectory starting in s and following option o

Degree of satisfiability



Affordances for Temporal Abstractions

- Consider a set of intents $\mathcal{I}^{\rightarrow} = \cup_{o \in \mathcal{O}} \mathrm{I}_o^{\rightarrow}$

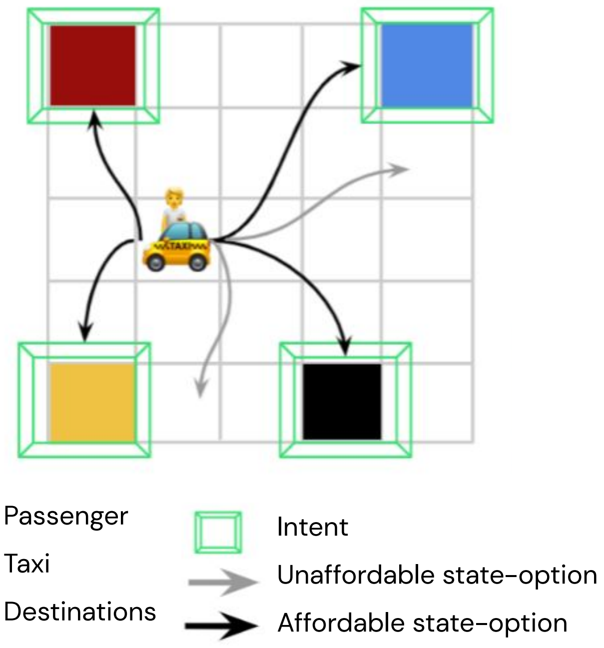
Option ***affordances*** are defined as a relation

$$\mathcal{AF}_{\mathcal{I}^{\rightarrow}} \subseteq \mathcal{S} \times \mathcal{O}$$

such that $\forall (s, o) \in \mathcal{AF}_{\mathcal{I}^{\rightarrow}}, \mathrm{I}_o^{\rightarrow}$ is ***satisfied***

at (s, o) to a degree $\zeta_{s,o} \leq \zeta^{\mathcal{I}^{\rightarrow}}$

Global degree of satisfiability



Theoretical Analysis

Planning Loss

- ***Certainty-equivalence (CE) planning loss:*** act according to the policy that is optimal with respect to the *estimated* model.

Let us consider the following:

$$\left\| \underbrace{V_M^*}_{\text{Value of the } \textbf{\textit{true optimal}} \text{ policy.}} - \underbrace{V_M^{\pi^*_{\hat{M}_{\mathcal{A}\mathcal{F}_{\mathcal{I}}}}}}_{\text{Value of the } \textbf{\textit{CE control}} \text{ policy.}} \right\|_\infty$$

Both are evaluated in the *true* model **M**.

Planning Loss Bound: *temporal abstraction*

$$\frac{2R_{max}}{(1-\gamma)^2} \times \left(\sqrt{\frac{1}{2n}\log\frac{2|S||A||\Pi_{S\times A}|}{\delta}}\right) \Rightarrow \frac{2R^{\mathcal{O}}_{max}}{(1-\bar{\gamma})^2}\left(\sqrt{\frac{1}{2n}\log\frac{\boxed{2|S||\mathcal{O}||\Pi_{S\times\mathcal{O}}|}}{\delta}}\right)$$

Jiang et al. 2015

Too big!

Planning Loss Bound: *temporal abstraction + affordances*

Trade-Offs

$$\frac{2R^{\mathcal{O}}_{max}}{(1-\bar{\gamma})^2}\left(\sqrt{\frac{1}{2n}\log\frac{\boxed{2|S||\mathcal{O}||\Pi_{S\times\mathcal{O}}|}}{\delta}}\right) \Rightarrow \frac{2R^{\mathcal{O}}_{max}}{(1-\bar{\gamma})^2}\left(\boxed{2\bar{\gamma}\zeta^{\mathcal{I}^{\rightarrow}}}+\sqrt{\frac{1}{2n}\log\frac{\boxed{2|\mathcal{AF}_{\mathcal{I}^{\rightarrow}}||\Pi_{\mathcal{I}^{\rightarrow}}|}}{\delta}}\right)$$

Intent Approximation

Dependence on Affordances

Faster planning across different timescales, though at the cost of potential approximation bias.

Sample Complexity

- To build the transition model, transitions are estimated by sampling the simulator, with the number of calls to this simulator referred to as the ***sample complexity***.

- Modelling one-time step dynamics would require samples in the order of magnitude of the size of the state-action space!

- ***Solution***: construct temporally abstract partial models

- Sample complexity of obtaining an $\mathcal{E}$ - estimation of the optimal action-value function given only access to a generative model.

Sample Complexity: *temporal abstraction*

$$\mathcal{O}\Big(\frac{|\mathcal{S}||\mathcal{A}|}{(1-\gamma)^4\varepsilon^2}\Big) \Rightarrow \mathcal{O}\Big(\frac{\boxed{|\mathcal{S}||\mathcal{O}|}}{(1-\bar{\gamma})^4\varepsilon^2}\Big)$$

Sample Complexity: *temporal abstraction + affordances*

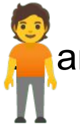
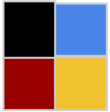
$$\mathcal{O}\left(\frac{|\mathcal{S}||\mathcal{A}|}{(1-\gamma)^4\varepsilon^2}\right) \Rightarrow \mathcal{O}\left(\frac{\boxed{|\mathcal{S}||\mathcal{O}|}}{(1-\bar{\gamma})^4\varepsilon^2}\right) \Rightarrow \mathcal{O}\left(\frac{\boxed{|\mathcal{AF}_{\mathcal{I}\rightarrow}|}}{(1-\bar{\gamma})^4\varepsilon^2}\right)$$

📌 The ability to understand abstract action opportunities resulting in improved sampled efficiency.

Empirical Analysis

The Taxi Domain

Task

…. must pick up  and drop to 

Rewards

- Correct drop off **+20 reward**.
- Wrong drop off **-10 reward**.
- *-1 reward* per step.

Experimental Setup

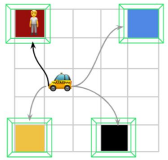
Pre-specified

Option Policies $\pi_o(a|s)$ — *Taxi centric: Go to x,y, Drop at x,y, Pickup at x,y*

Intents — *Passenger centric, Drop @ any destination*

Affordances — *Everything [37500] , Pickup+Drop [25000] , Pickup+Drop@Goal [4000]*

Rollouts $\xrightarrow{S,O,R,S'}$ **Function approximation** $\xrightarrow{M_\phi}$ **Dynamic Programming** $\xrightarrow{\pi(o|s)}$ **Evaluation**



*Learn Option Model using **Fixed** Set of Affordances*

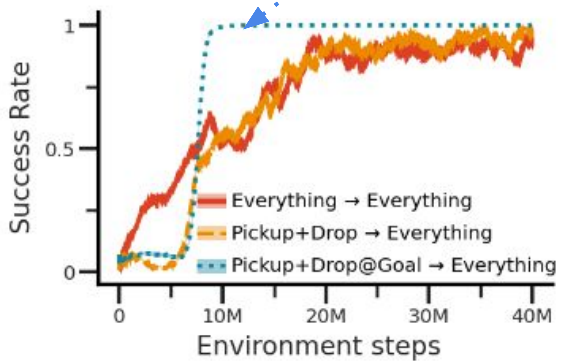
Planning over options using Option model and affordances

Data collection and learning

Planning

When are intents & affordances are most useful?

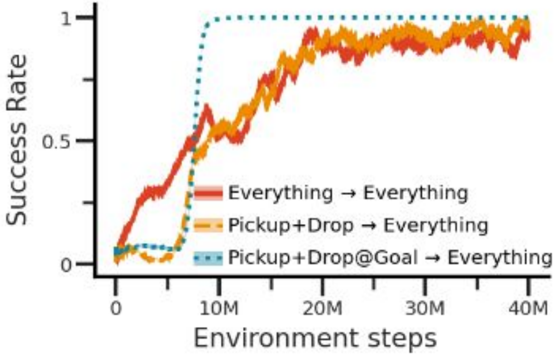
Affordances improve model learning even in the absence of them during planning => useful partial option model



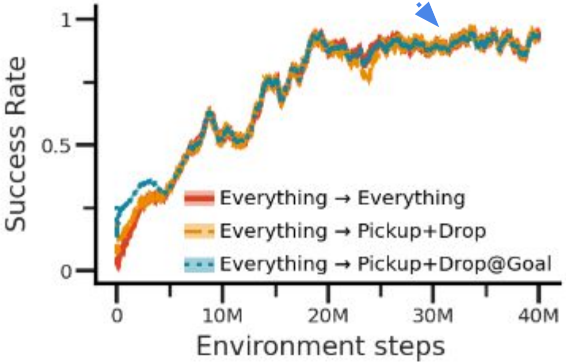
(a) Data collection and model learning with affordances.

When are intents & affordances are most useful?

Affordances did not impact planning as the underlying model is the same.



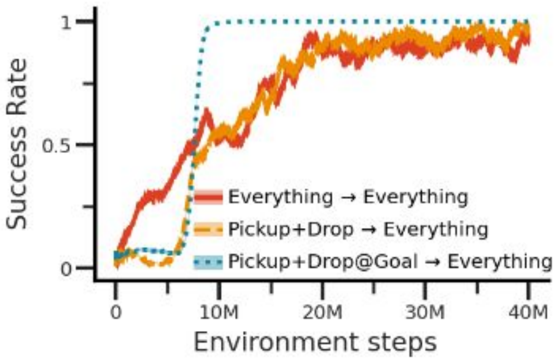
(a) Data collection and model learning with affordances.



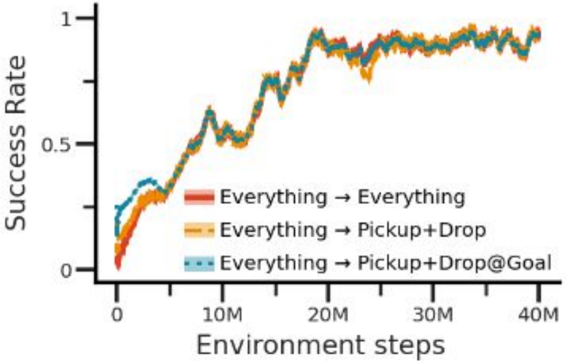
(b) Planning with affordances.

When are intents & affordances are most useful?

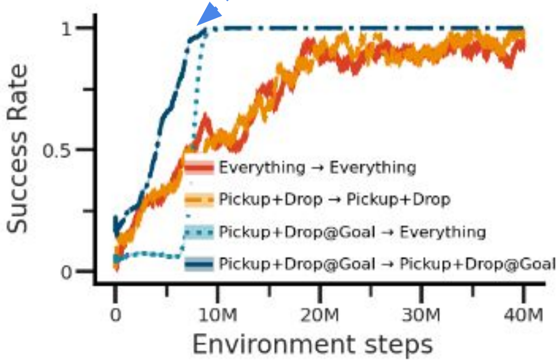
Affordances informing both model learning and planning result in best performance.



(a) Data collection and model learning with affordances.



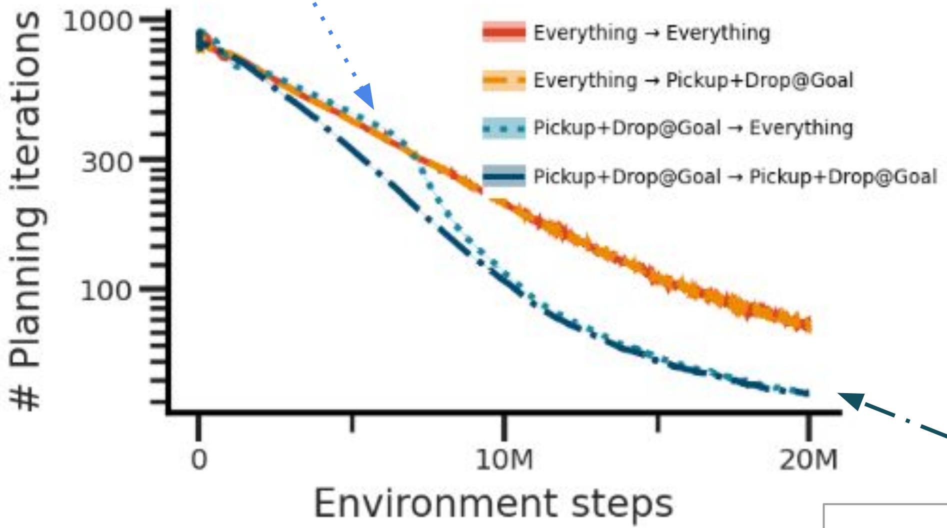
(b) Planning with affordances.



(c) Data collection, model learning and planning with affordances.

Impact on sample efficiency

Learning a partial option model requires much fewer samples as opposed to learning a full model.

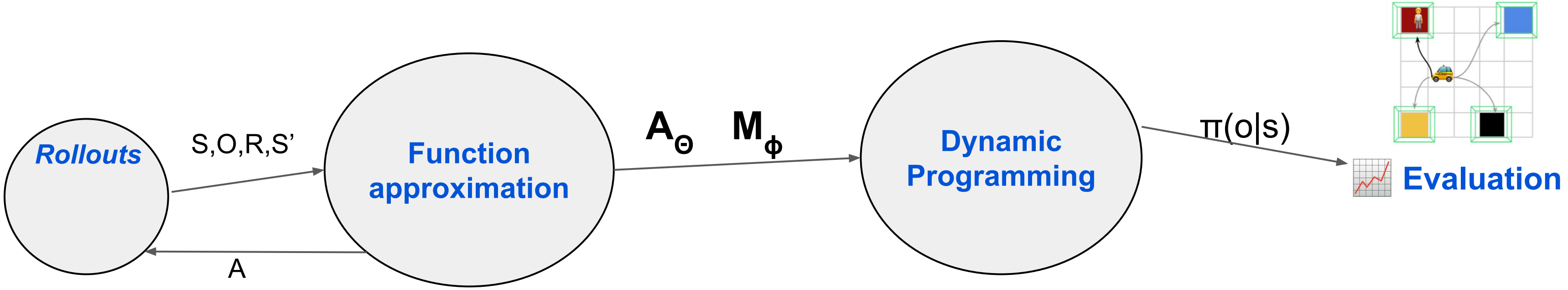


Using affordances during model learning and planning decreases planning iterations.

Experimental Setup - Learned affordances

Pre-specified

Option Policies $\pi_o(a\|s)$	*Taxi centric: Go to x,y, Drop at x,y, Pickup at x,y*
Intents	*Passenger centric: Drop @ any destination*

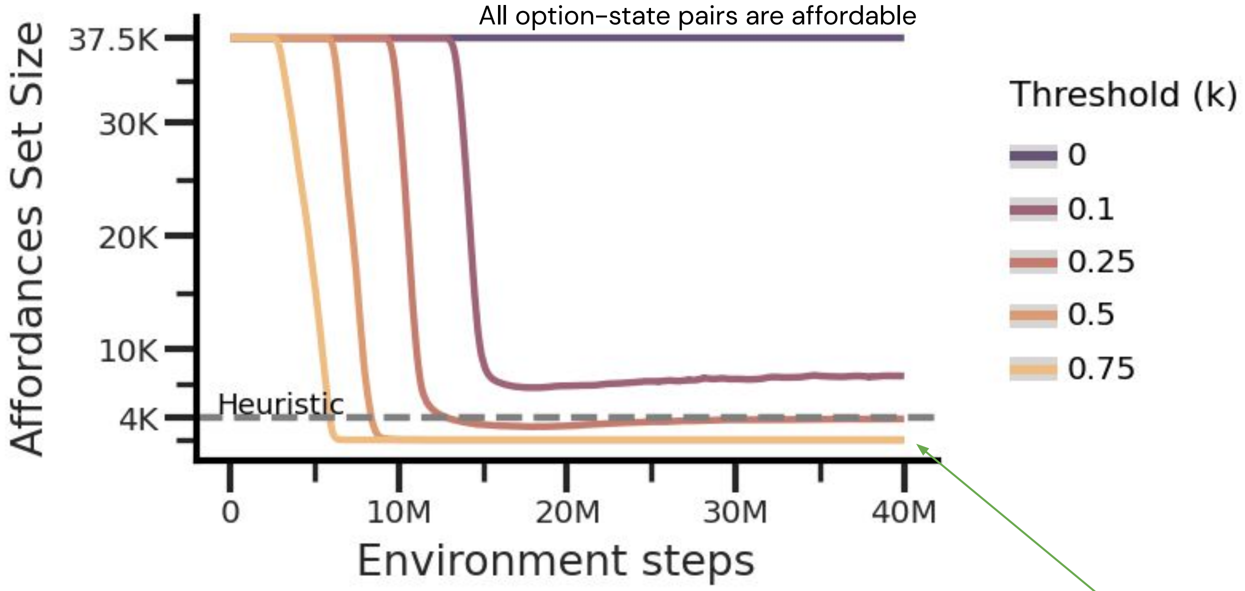


*Learn Option Model **and*** ***Learned Affordances***

Planning over options using Option model and affordances

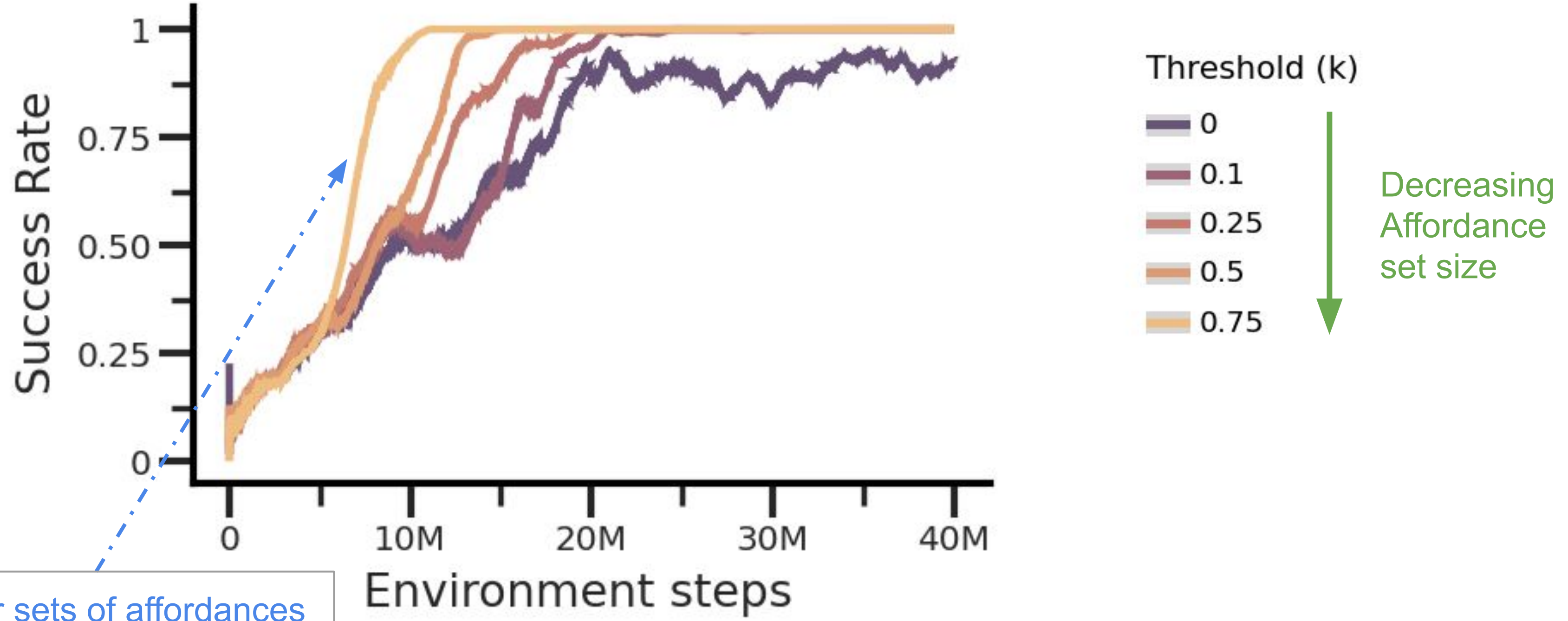
We can learn affordance sets online!

Controls the size of the affordance set



The best learned affordance sets are smaller than what we could come up as heuristics!

The impact of the affordance set size on performance



Smaller sets of affordances
= quicker learning

Conclusion

We presented notions of intents and affordances that can be used together with options.

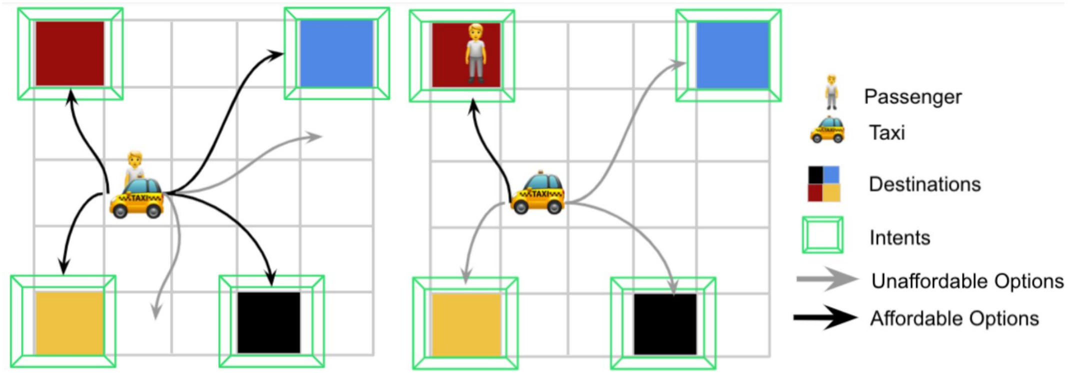
- [Theoretically] Modelling temporally extended dynamics for only relevant parts leads to
  - Faster planning across different timescales
  - Improved sampled complexity in learning such models
- [Empirically] Learning affordances online for model learning and planning results in
  - Improvements in performance in downstream task
  - Drastically reduced state-option space

Future Work

- *Discovery* of options as well as intents
- Study the emergence of *new* affordances at the boundary of the agent-environment interaction in the presence of non-stationarity.
- Relate our work to cognitive science models of *intentional* options

tl;dr Temporally Abstract Partial Models

Proposed temporally abstract partial options models via the notion of affordances, with theoretical guarantees and empirical analysis demonstrating improvement in *final performance* and *sample efficiency*.



To appear @NeurIPS 2021

Contact: khimya.khetarpal@mail.mcgill.ca