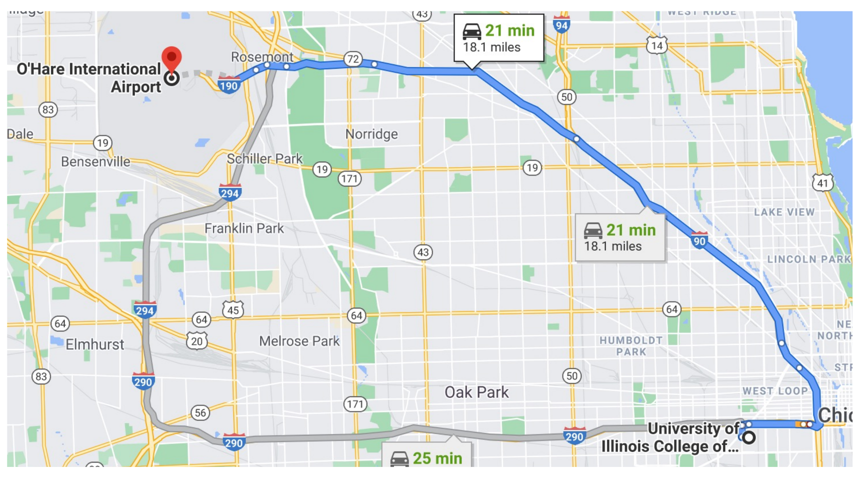
Dynamic probabilistic logic models for effective task-specific abstractions in RL

STARLinG LAB
UT DALLAS
Harsha Kokel, Sriraam Natarajan, Balaraman Ravindran, Prasad Tadepalli
IIT Madras
Oregon State University

Motivation

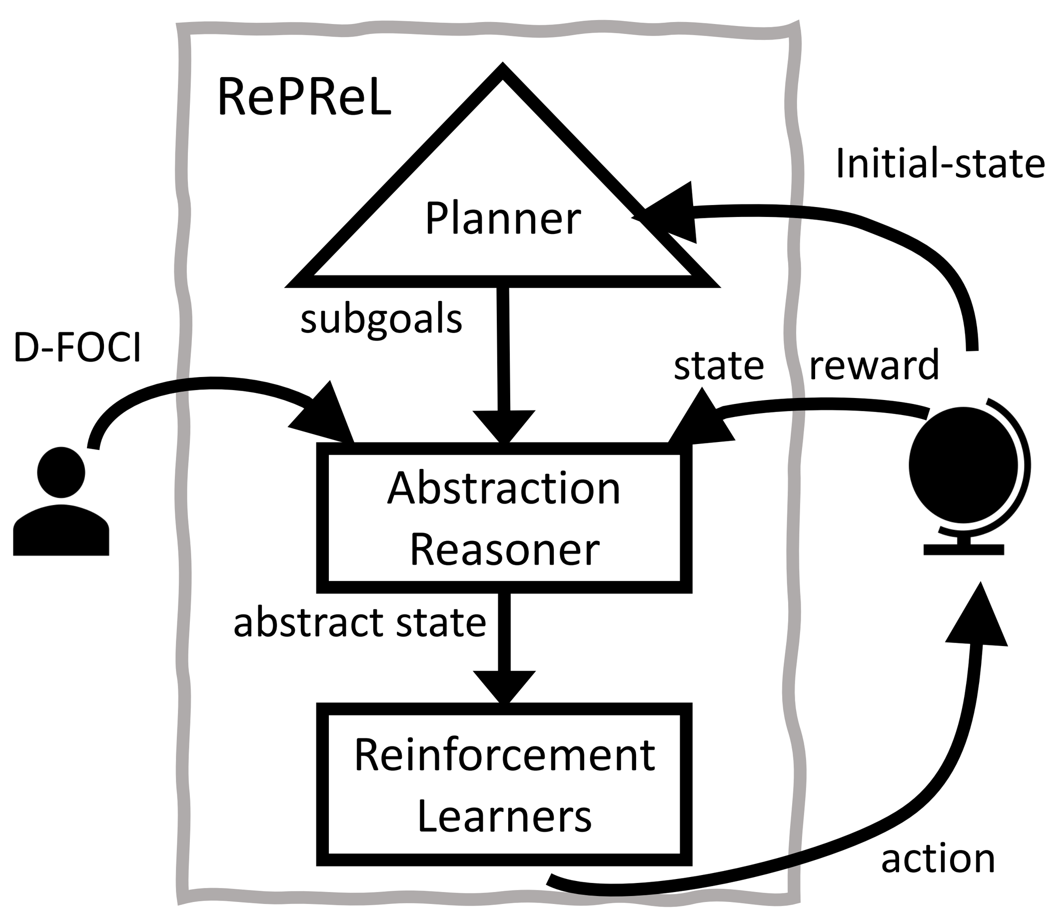


Planning | Execution

- For a task like going to airport, the state space of planning a route differs from the state space of driving a car.
- Humans do not have access to dynamic nature of the state space while planning, nor do they have computational resources to plan for all possible traffic events.
- Humans can still plan and execute the task efficiently.

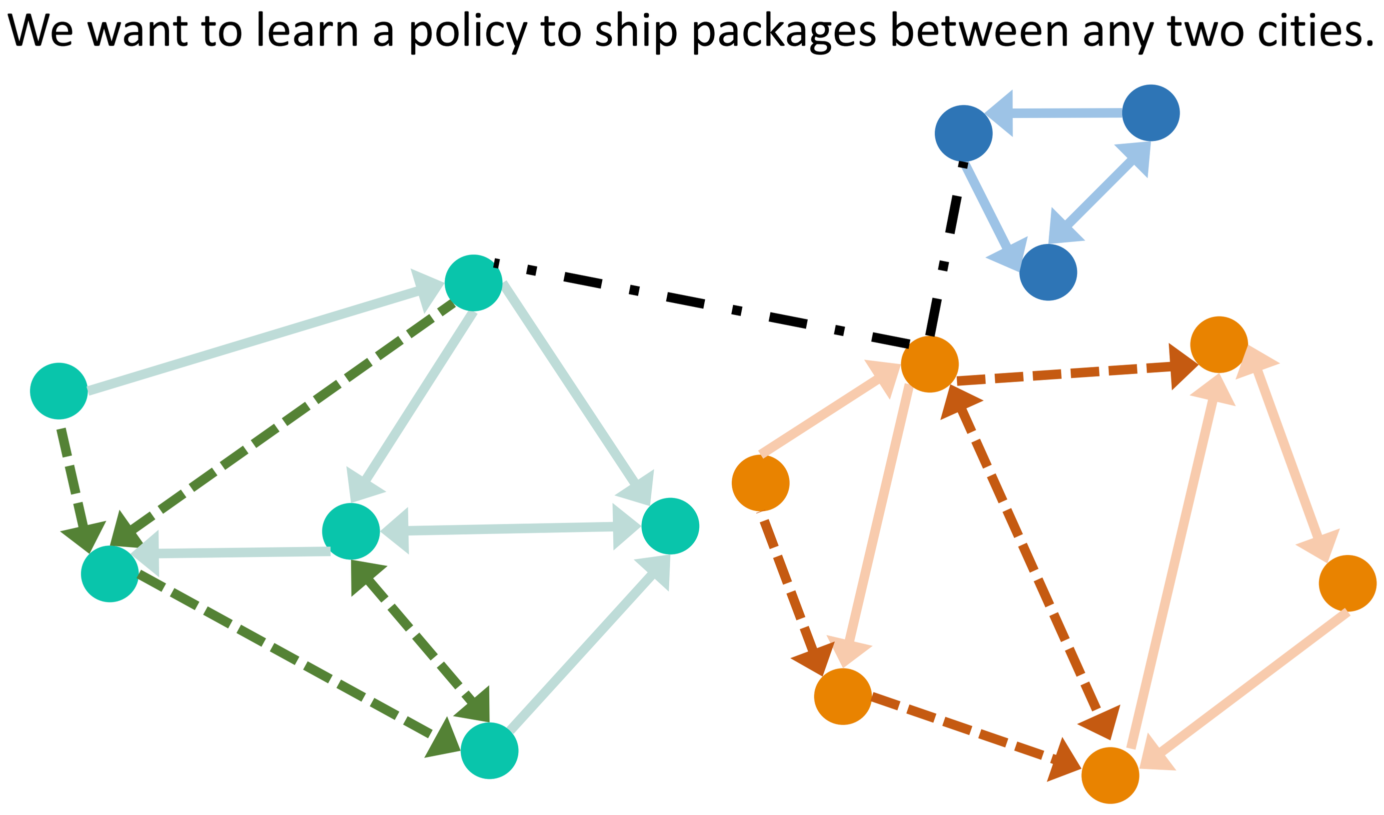
The key principle that enables humans to deal with these informational and computational challenge is **abstraction**.

RePReL

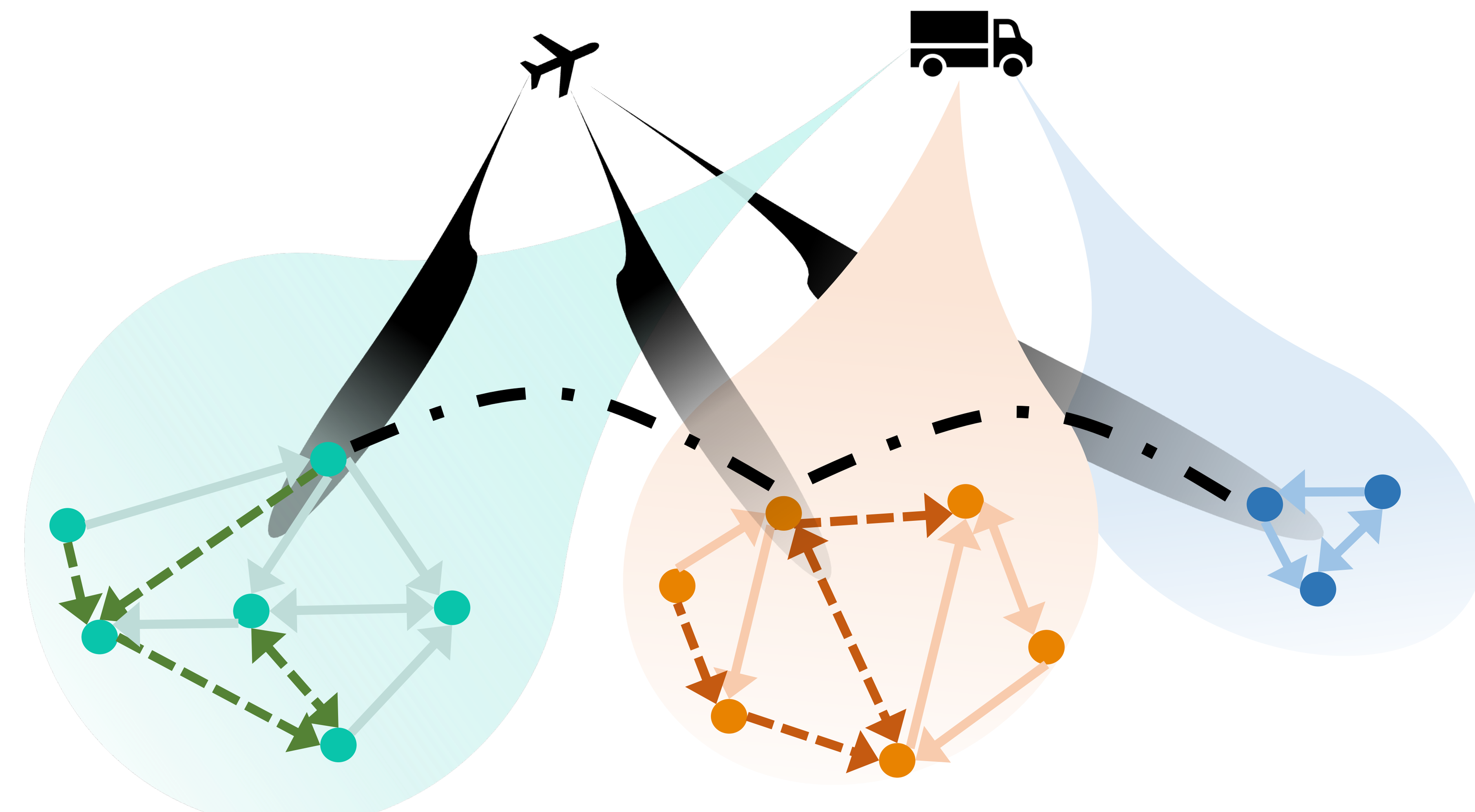


Toy Example

- Consider a toy relational logistics domain with 3 countries green, orange and blue.
- Each country has multiple cities. Two cities can be connected by a high-way or a road.
- There is a flight which goes to a major city in all three countries.
- We want to learn a policy to ship packages between any two cities.



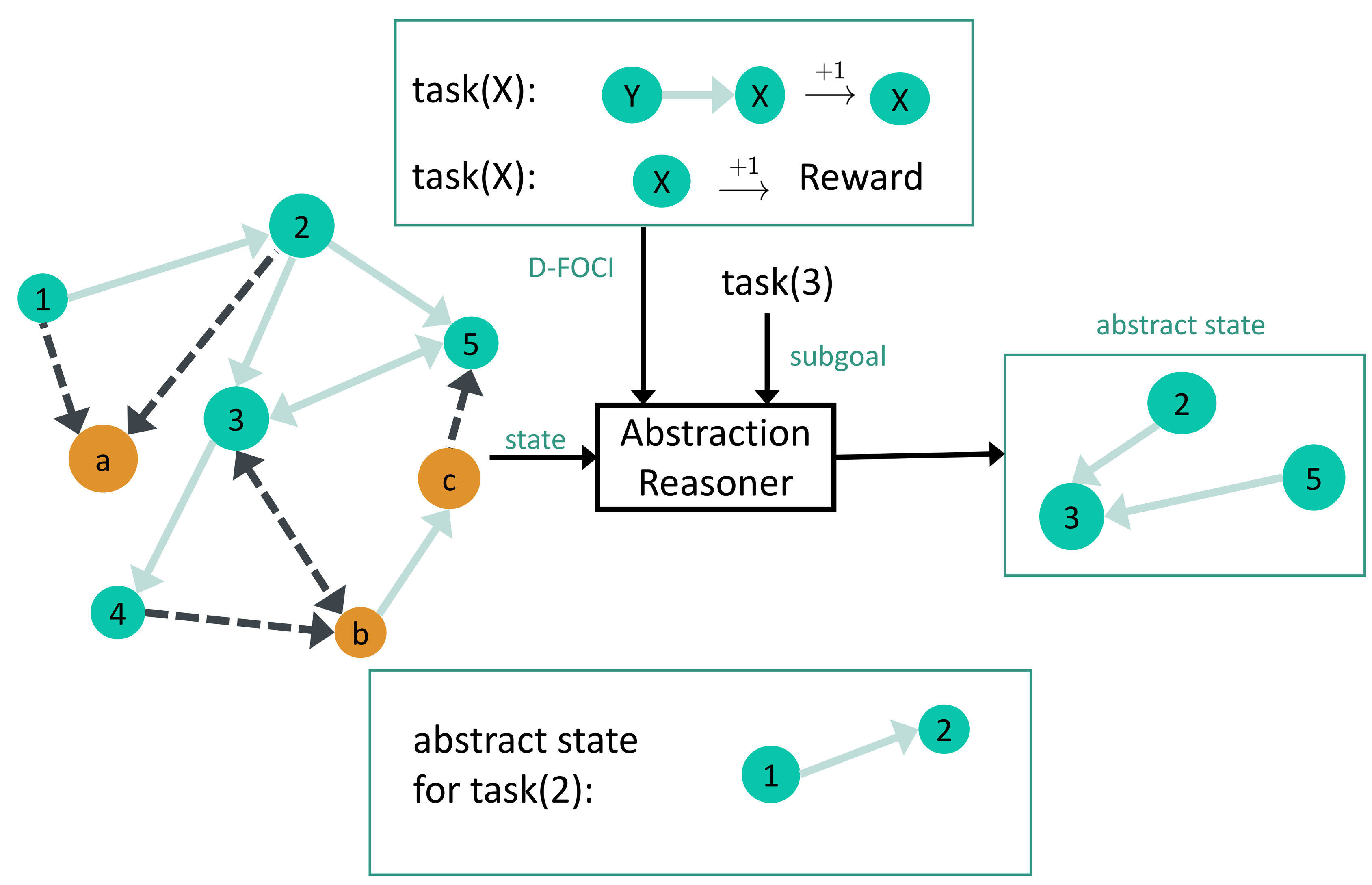
D-FOCI Statements for task-specific abstraction



- Instead of learning an End-to-End RL policy, RePReL decomposes the task and learns 2 RL agents.
- One agent learns how to transport the package in a country internally with truck. Other learns how to transport it internationally with flight.
- Agents can use task-specific state representation and make decisions effectively using Dynamic First-Order Conditional Influence statements.

D-FOCI template $\text{sub-task}: X1 \xrightarrow{+1} X2$

D-FOCI statement (with $\xrightarrow{+1}$) encodes that the literals in set X2 (at time step t) are influenced by literals in set X1 (at time step t-1) when the given "sub-task" is being executed.



Guarantees

- Task-specific state abstraction obtained using D-FOCI statements is model agnostic abstraction.
- This state abstraction is safe and preserves optimality if the MDP satisfies the influences encoded in the D-FOCI statements.

A **model agnostic abstraction** ϕ preserves the transition function and reward function of the ground MDP in the abstract MDP. That is, $\phi(s_1) = \phi(s_2)$ if and only if any action a and abstract state $\bar{s}$ satisfies following conditons,

$$\sum_{\{s'_1|\phi(s'_1)=\bar{s}\}} R_o(s_1, a, s'_1) = \sum_{\{s'_2|\phi(s'_2)=\bar{s}\}} R_o(s_2, a, s'_2)$$

$$\sum_{\{s'_1|\phi(s'_1)=\bar{s}\}} P_o(s_1, a, s'_1) = \sum_{\{s'_2|\phi(s'_2)=\bar{s}\}} P_o(s_2, a, s'_2)$$

Human knowledge can help provide task-specific abstractions that enable efficient learning and effective generalization across tasks and objects.