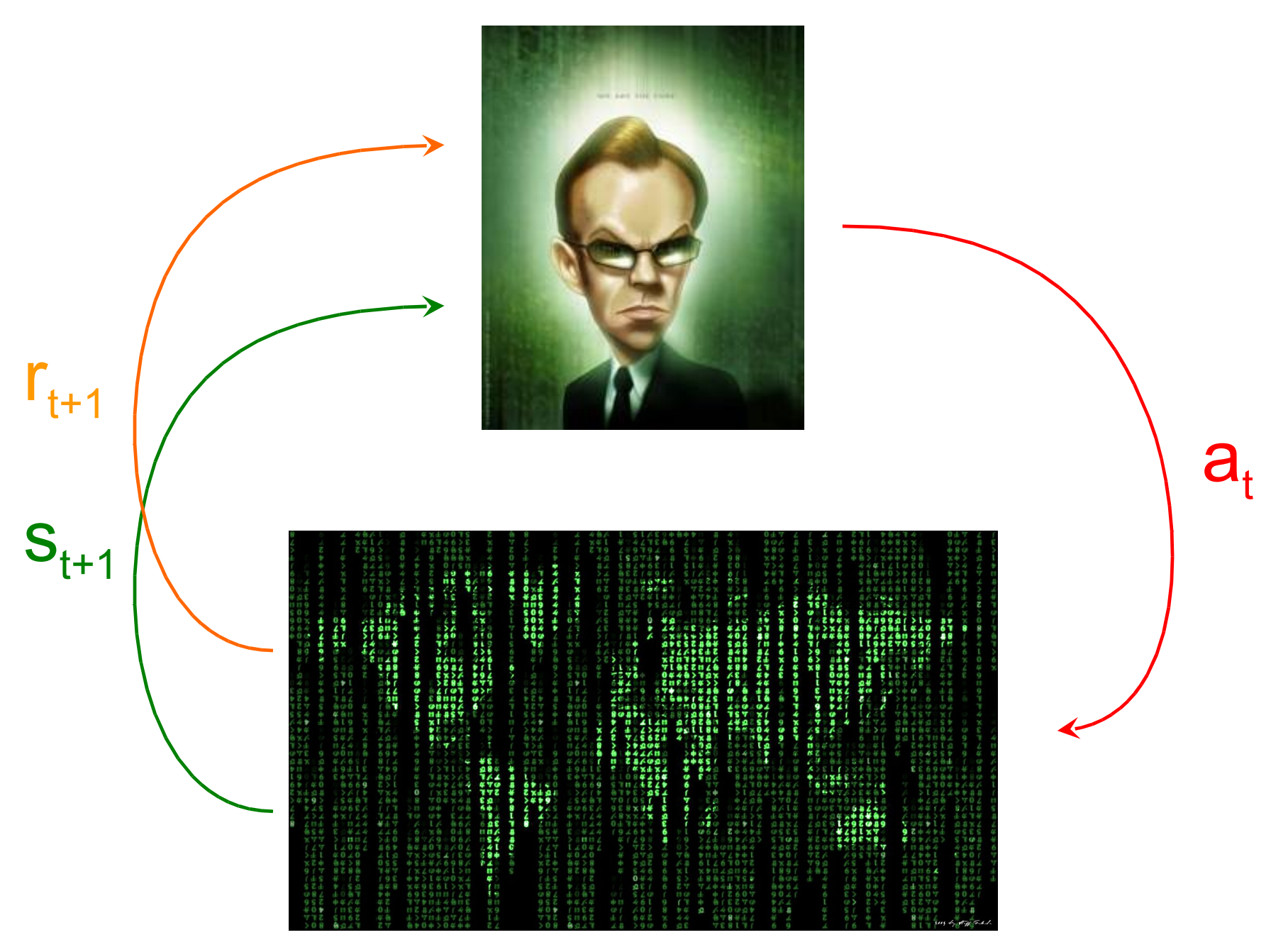
Model-Assisted Approaches for Relational Reinforcement Learning:

Some challenges for the SRL community

Tom Croonenborghs, Jan Ramon, Hendrik Blockeel and Maurice Bruynooghe

1.Relational Reinforcement Learning (RRL)



Given:

- set of possible states S
- set of possible actions A
- unknown transition function

$$\delta : S \quad A \quad S$$

- unknown reward function

$$r : S \quad A$$

Find policy $\pi^* : S \qquad A$, maximizing

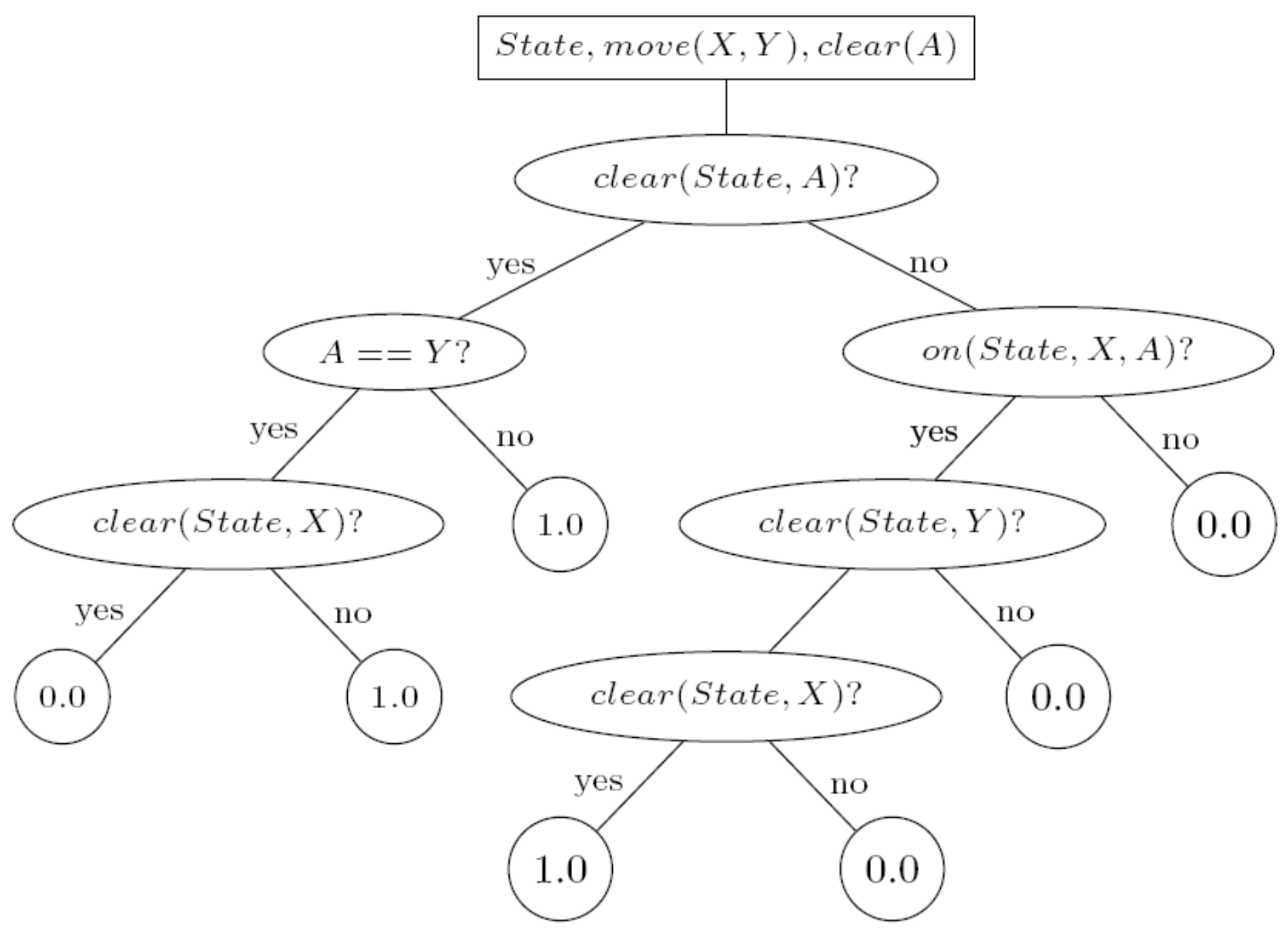
$$V^{\pi}(s_t) = \quad_{i=0} \gamma^i r_{t+i}$$

2. Model-Assisted Approaches for RRL

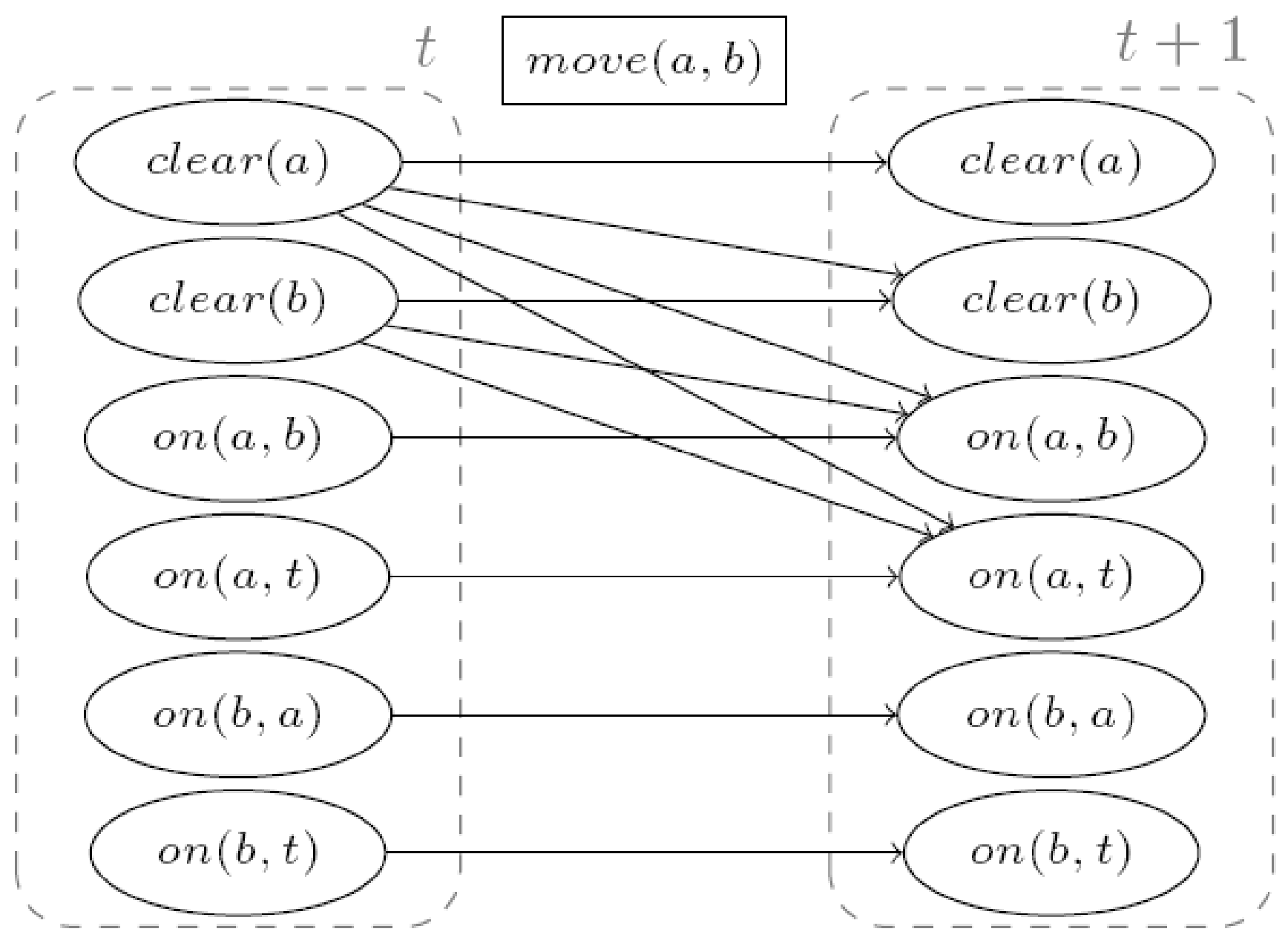
- The (full or fully correct) MDP is not always available
  - Learning extra information can be useful
- Learn a model of the world while performing RL
  - Knowing world dynamics ⟺ exploiting this knowledge
    - ✓ Learning good models might require a policy that reach important places
    - ✓ Knowledge of the world may be essential to learn a good policy
- Relational domains
  - Learning a (good) model is more challenging
    - Even impossible ⇒ how to handle this uncertainty?
- First indirect RRL approach
  - Learn transition and reward probability distributions
  - Improve policy by performing a (small) local search starting in current state

3.1. Learning the transition function

- Represent as a Relational Dynamic Bayesian Network
  - Assume that random variables describing the next state only depend on the current state and the action
    - Only parameter learning
- Conditional probability distribution modeled as a relational probability tree for every state predicate
  - Learned with the TG-algorithm



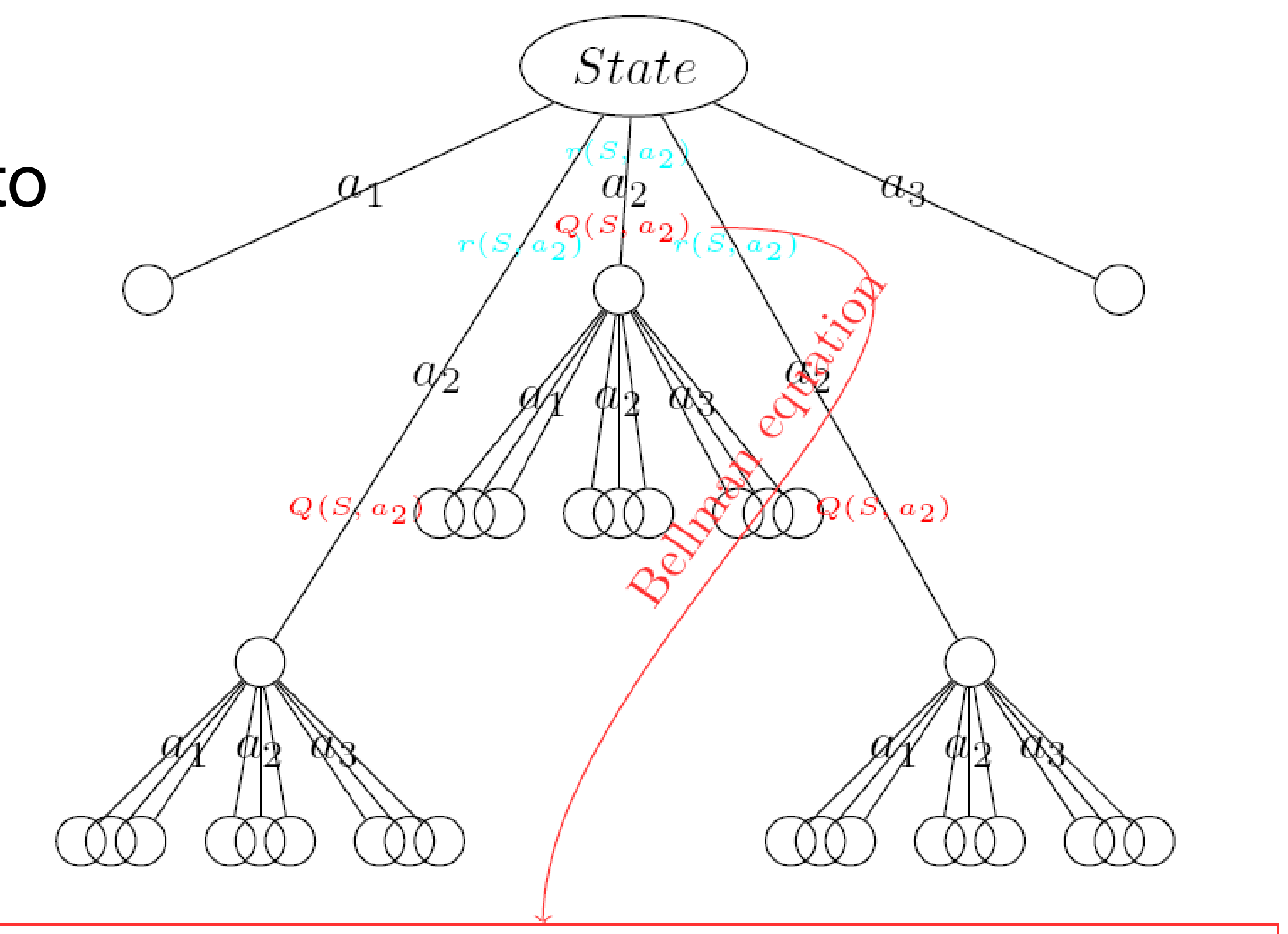
Probability tree showing the probability that block *A* will be clear, given the fact that action *move(X,Y)* is executed in state *State*.



Example grounded RDBN showing the dependencies between two successive states for the *move(a,b)* action.

3.2. Q-Learning with Lookahead

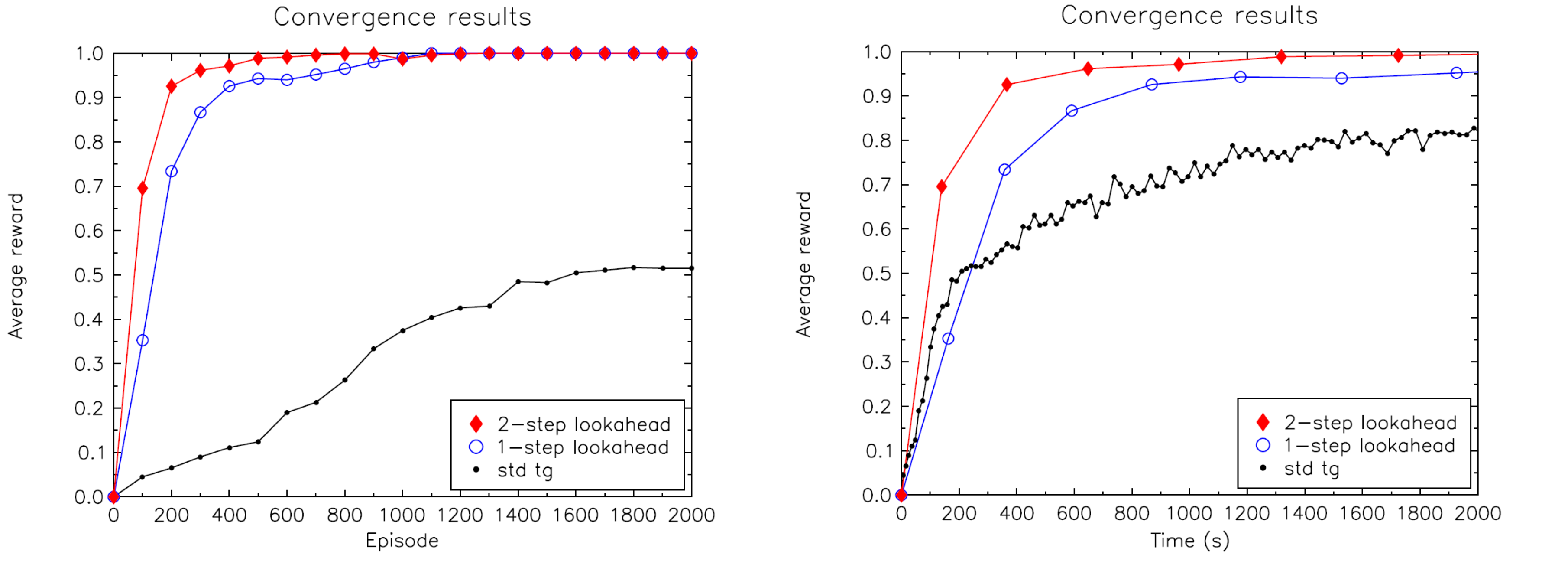
When an action needs to be chosen, instead of using the Q-values for the current state, the agent can look some steps ahead to obtain more informative Q-values.



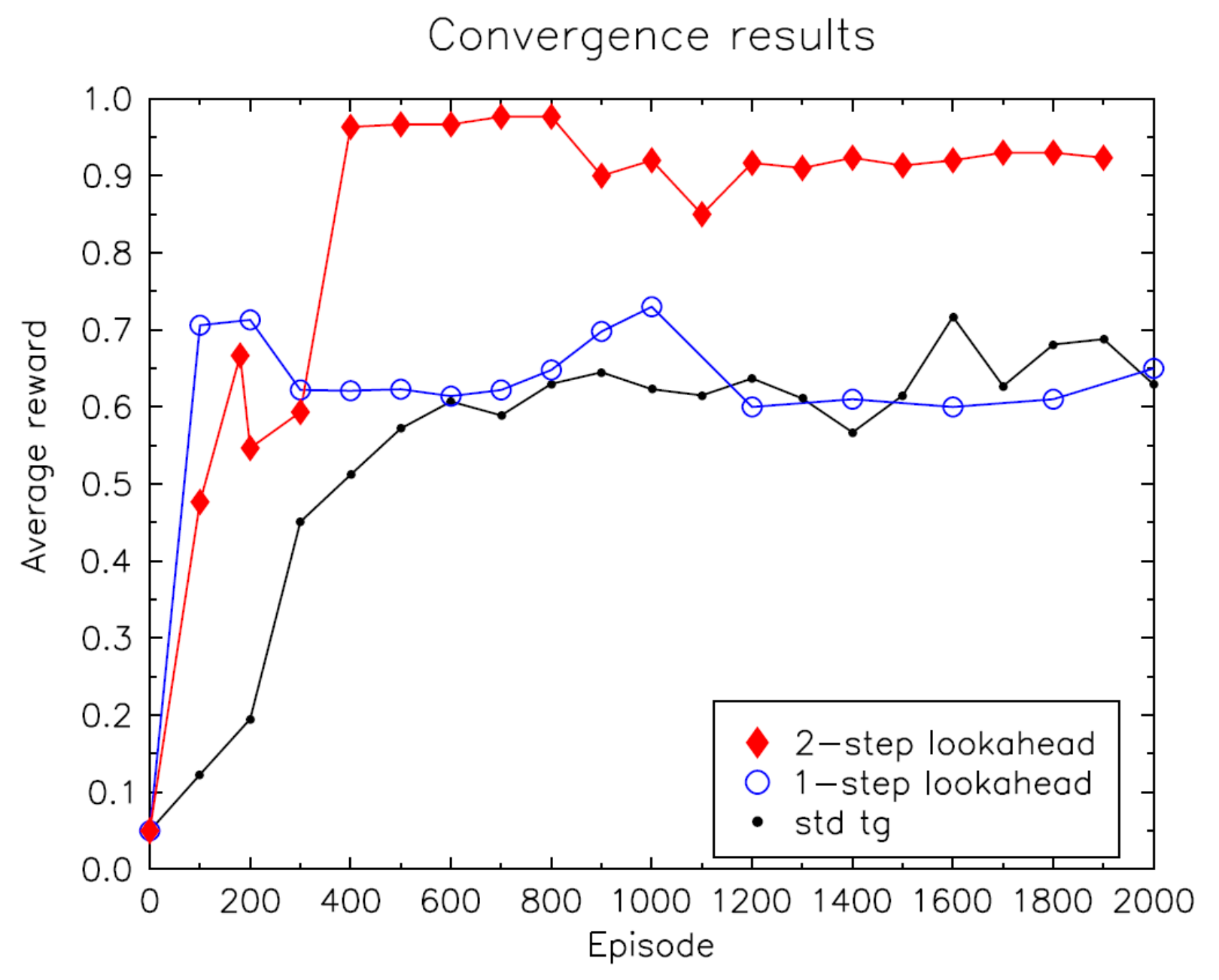
$$Q(S, a_2) = \frac{1}{M} \sum_0^M \left(r(S, a_2) + \gamma \frac{1}{N} \sum_0^N max_a Q(S', a) \right)$$

3.3. Preliminary Experiments

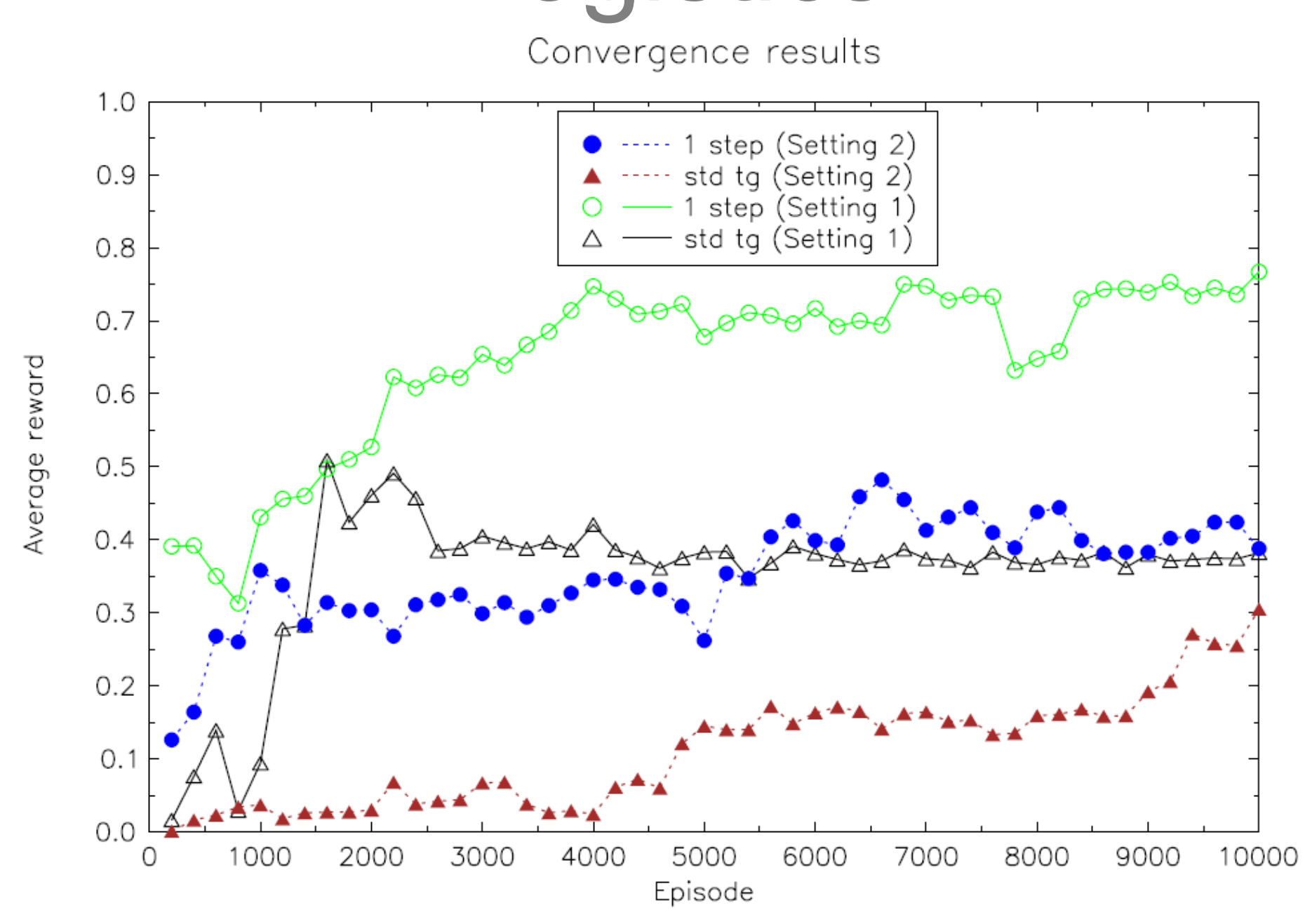
on(A.B) in blocks world with 5 blocks



stack



logistics



4. Challenges and Open Problems

- Evaluate different components of the learned model
- Efficient sampling strategies
- Efficient planning techniques

5. Conclusions

- First model-assisted RRL approach
- Incrementally learn a RDBN to model the transition function
- Improved convergence speed by looking some steps ahead

Acknowledgements: This research was funded by the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT Vlaanderen) and by the Fund for Scientific Research (FWO) of Flanders.

Contact: Tom.Croonenborghs@cs.kuleuven.be or http://www.cs.kuleuven.be/~tomc