UiO : University of Oslo

# Model-based Bayesian Reinforcement Learning for Dialogue Management


Pierre Lison

*Language Technology Group, University of Oslo*



August 26, 2013

**Interspeech**

# Motivation

- Hand-crafting dialogue policies is hard!
  - Noise & uncertainty (e.g. speech recognition errors)
  - Large number of possible trajectories
- *Alternative*: automatically optimise dialogue policies from (real or simulated) experience
- Two types of approaches:
  - **Model-free** reinforcement learning
  - **Model-based** reinforcement learning

# Motivation

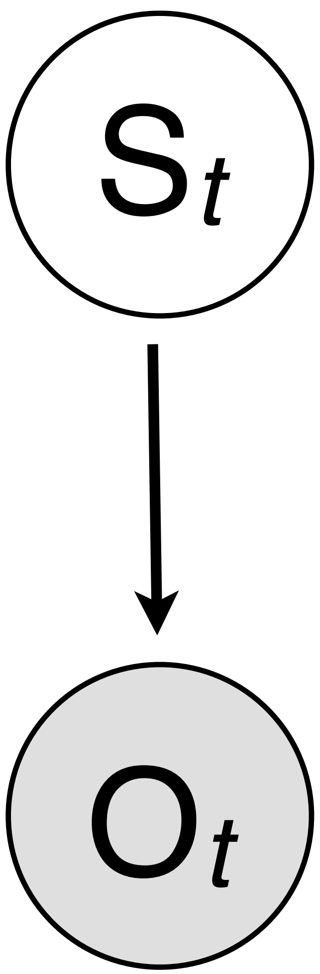
- Hand-crafting dialogue policies is hard!
  - Noise & uncertainty (e.g. speech recognition errors)
  - Large number of possible trajectories
- *Alternative*: automatically optimise dialogue policies from (real or simulated) experience
- Two types of approaches:
  - **Model-free** reinforcement learning
  - **Model-based** reinforcement learning

Focus of this talk

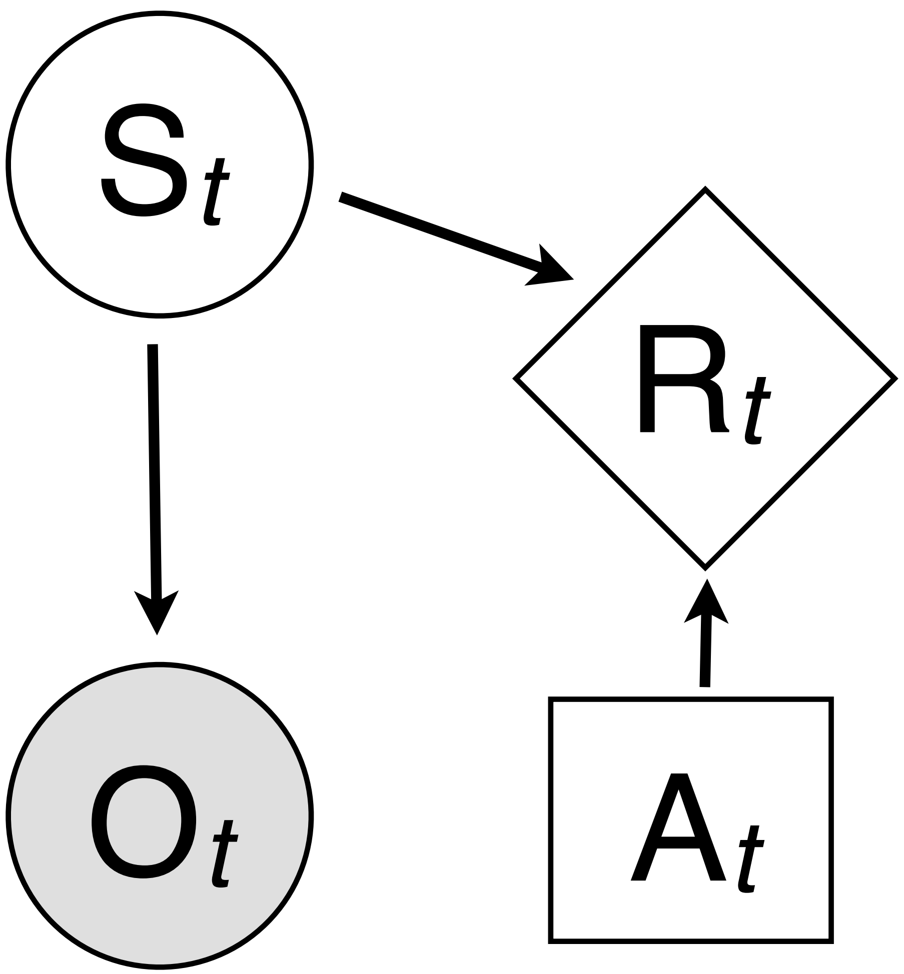
# Motivation

- Model-based reinforcement learning:
  - Collect interactions and use them to estimate explicit models of the domain
  - Use the resulting models to plan the best action
- *Key advantage*: can exploit prior knowledge to structure the domain models
- We present an experiment showing the benefits of this approach

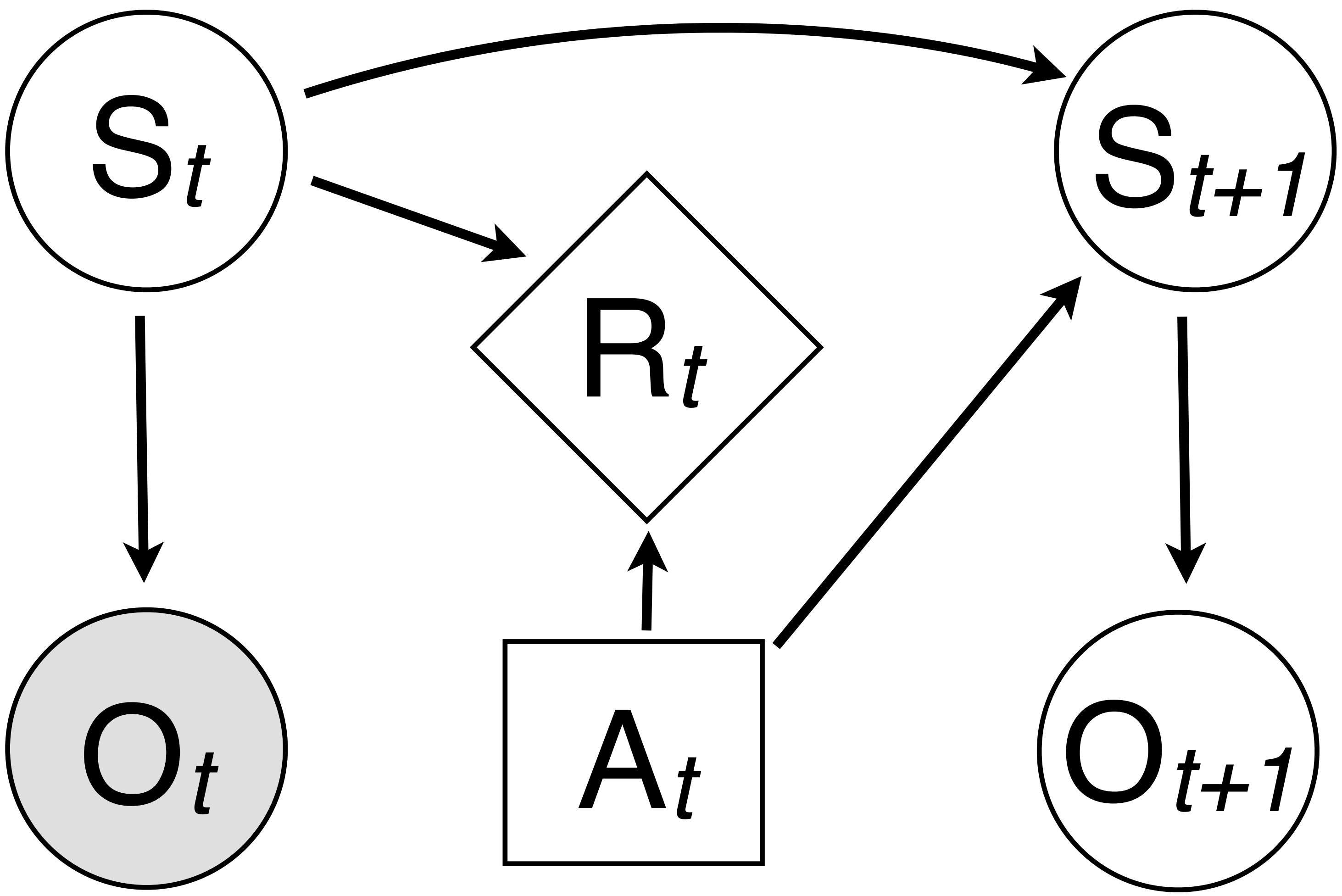
# POMDPs



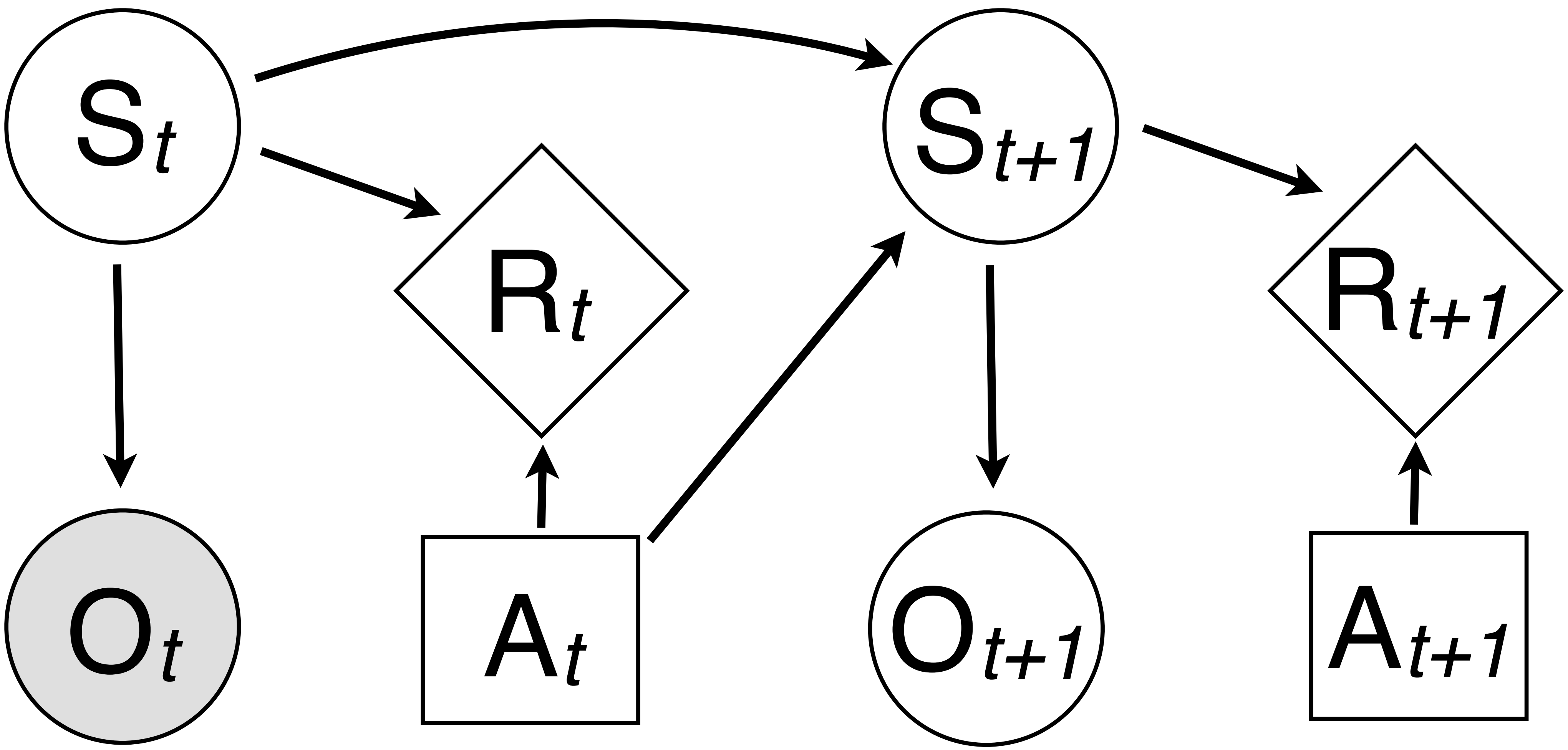
# POMDPs



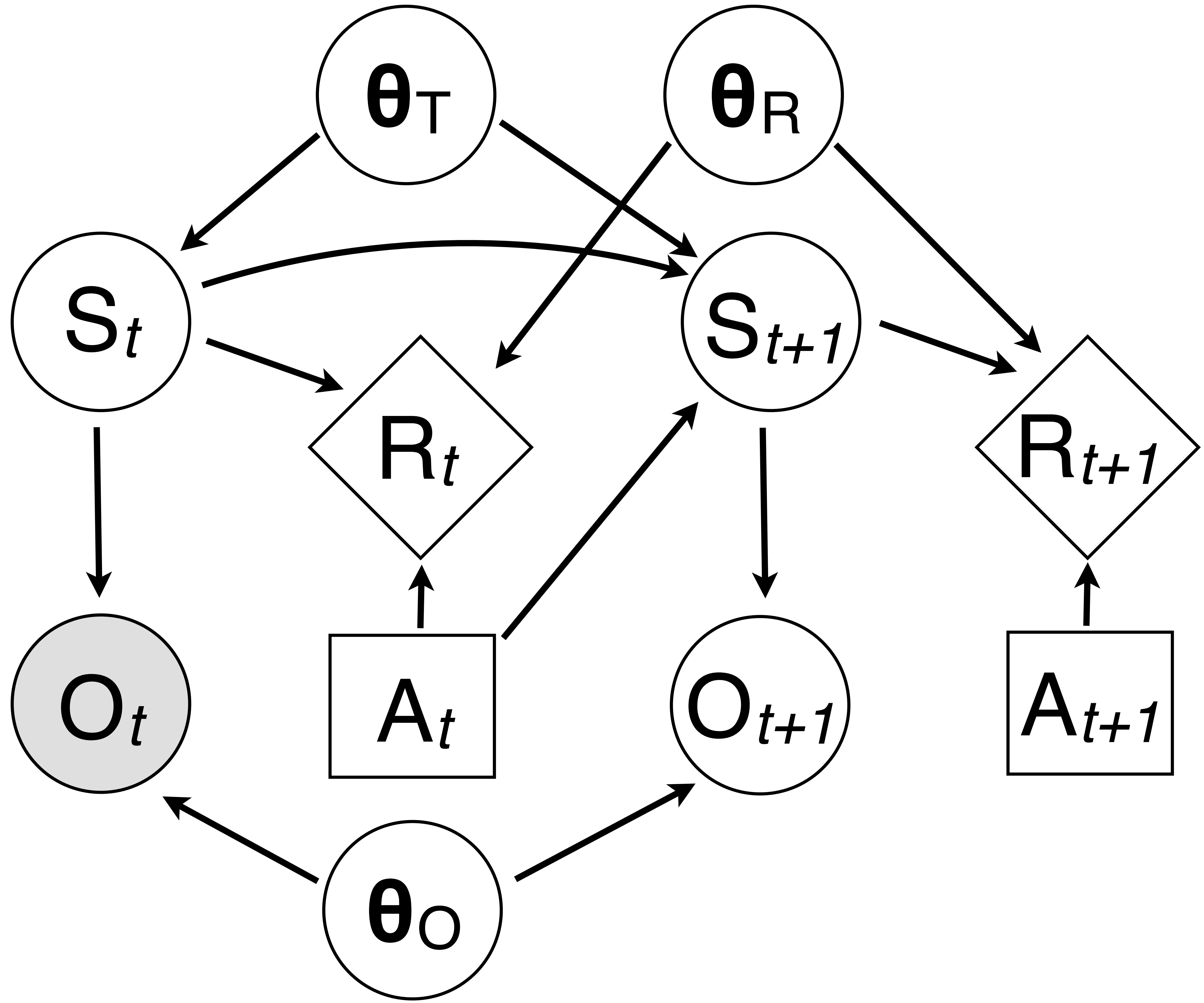
# POMDPs



# POMDPs



# POMDPs with model uncertainty



# Approach

- After each observation, the parameters are updated via *Bayesian inference*
  - Parameter distributions gradually narrowed down to the values that best fit the observed data
- Forward planning is used to select the next action to execute at runtime
  - Three source of uncertainty: state uncertainty, stochastic action effects, and model uncertainty

# Abstraction

- Dialogue domains often have large, complex state and action spaces
- Need *generalisation/abstraction techniques* to avoid the «curse of dimensionality»
- The framework of **probabilistic rules** offers such abstraction language
  - Capture domain structure through (parametrised) rules mapping conditions to probabilistic effects
  - Drastic reduction in the number of parameters

# Probabilistic rules

- Structured if...then...else cases associating conditions to distributions over effects:

> **if** (condition$_1$ holds) **then**
> P(effect$_1$)= $\theta_1$, P(effect$_2$)= $\theta_2$, ...
> **else if** (condition$_2$ holds) **then**
> P(effect$_3$) = $\theta_3$, ...
> ...

- Probabilistic rules serve as *high-level templates* for a Bayesian network

[P. Lison, «*Probabilistic Dialogue Models with Prior Domain Knowledge*», SIGDIAL 2012]

# Probabilistic rules: example

$r_1$: $\forall$ X:

**if** ($a_m$ = AskConfirm(X) $\wedge$ $i_u \neq$ X) **then**

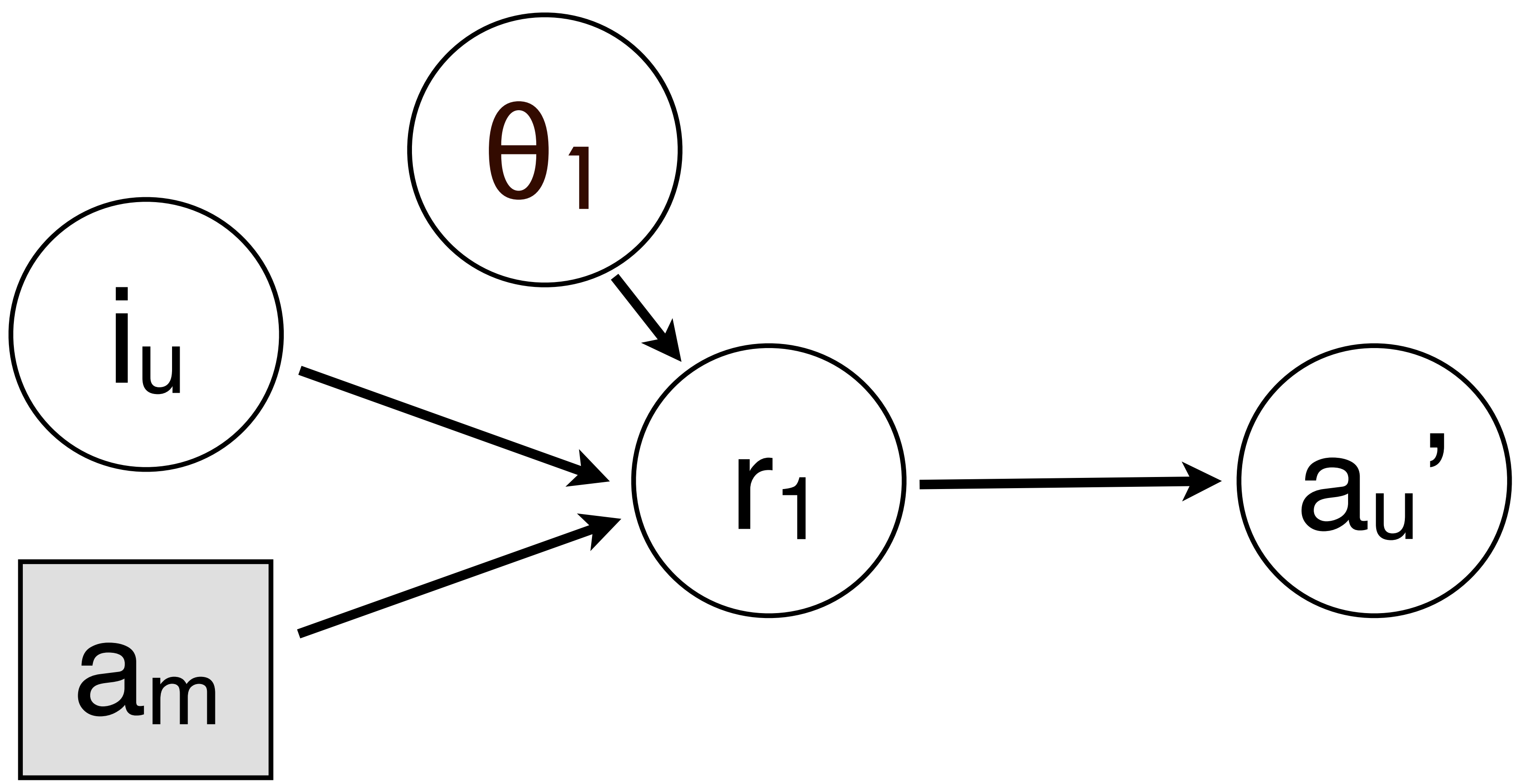
[P($a_u'$ = Disconfirm) = $\theta_1$]

# Probabilistic rules: example

$r_1$: $\forall$ X:

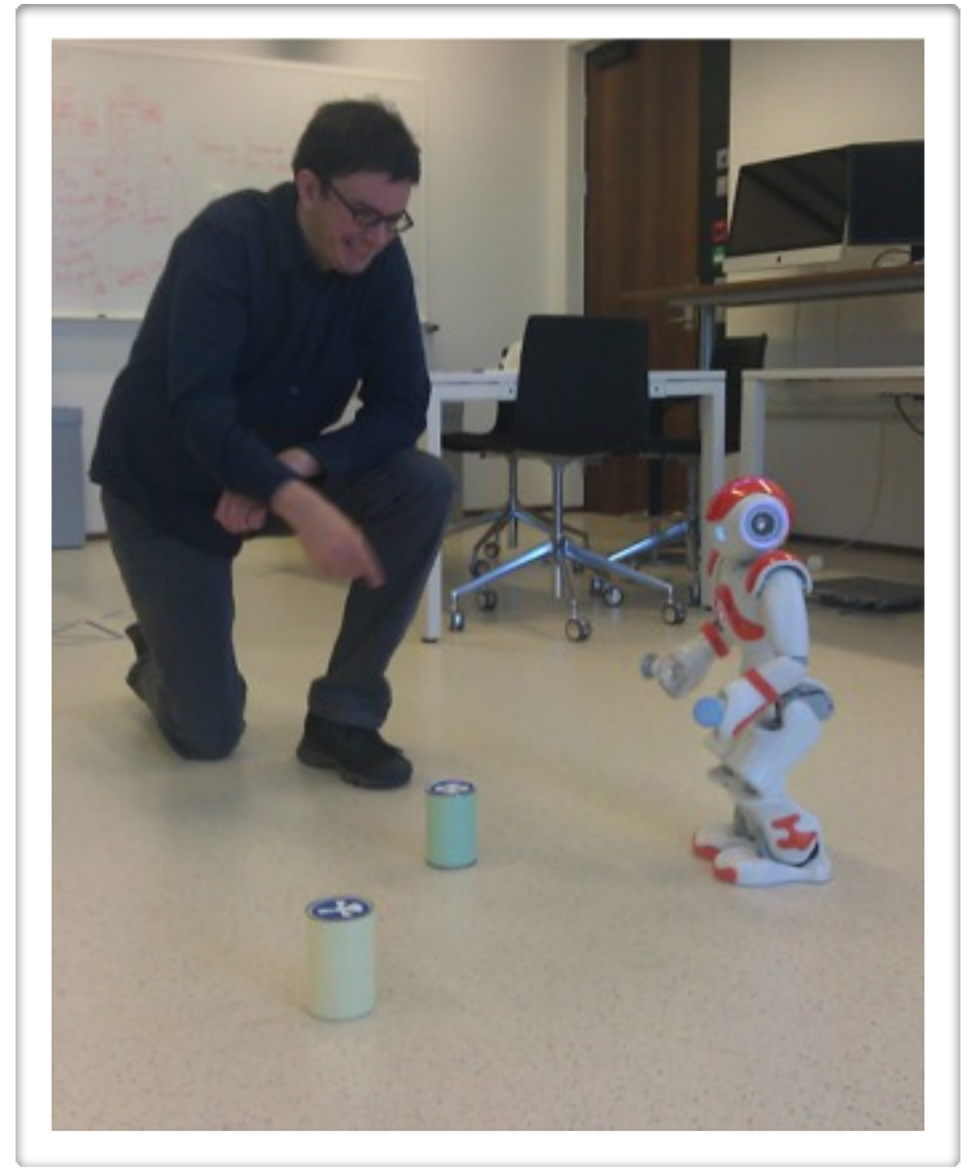
**if** ($a_m$ = AskConfirm(X) $\wedge$ $i_u \neq$ X) **then**

[P($a_u'$ = Disconfirm) = $\theta_1$]



# Evaluation

- Evaluation of the learning approach in a simulated environment:
  - Human-robot interaction domain (with Nao robot)
  - Simulator constructed from Wizard-of-Oz data
  - **Goal**: estimate the *transition model* of the domain (reward model is given)



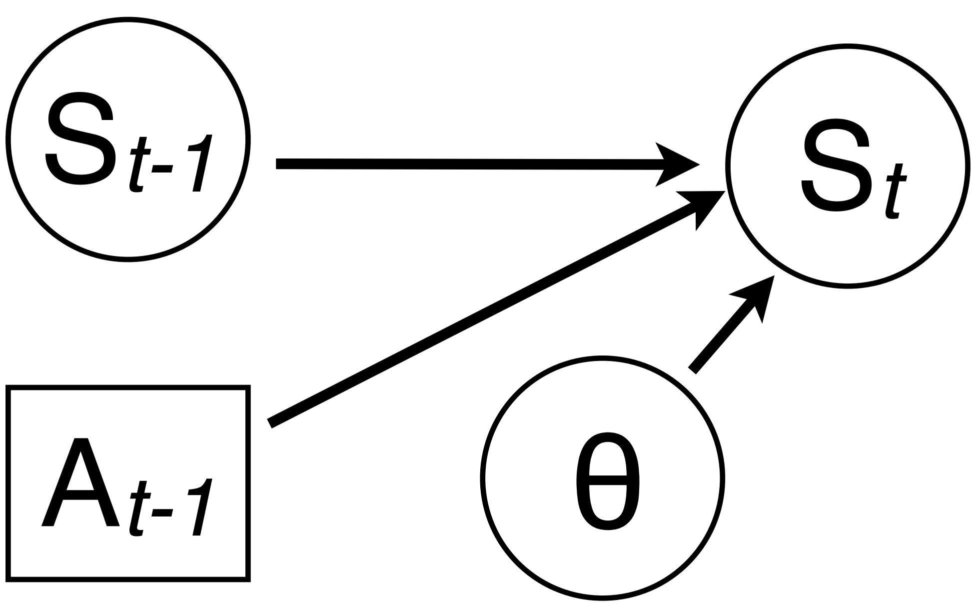
# Simulator

- Simulation models:
  - *User modelling*: how the user is expected to react to the system actions
  - *Context modelling*: how the system actions change the state of the environment
  - *Error modelling*: how understanding errors can occur
- Collected and annotated Wizard-of-Oz data to empirically estimate these models

# Experimental setup

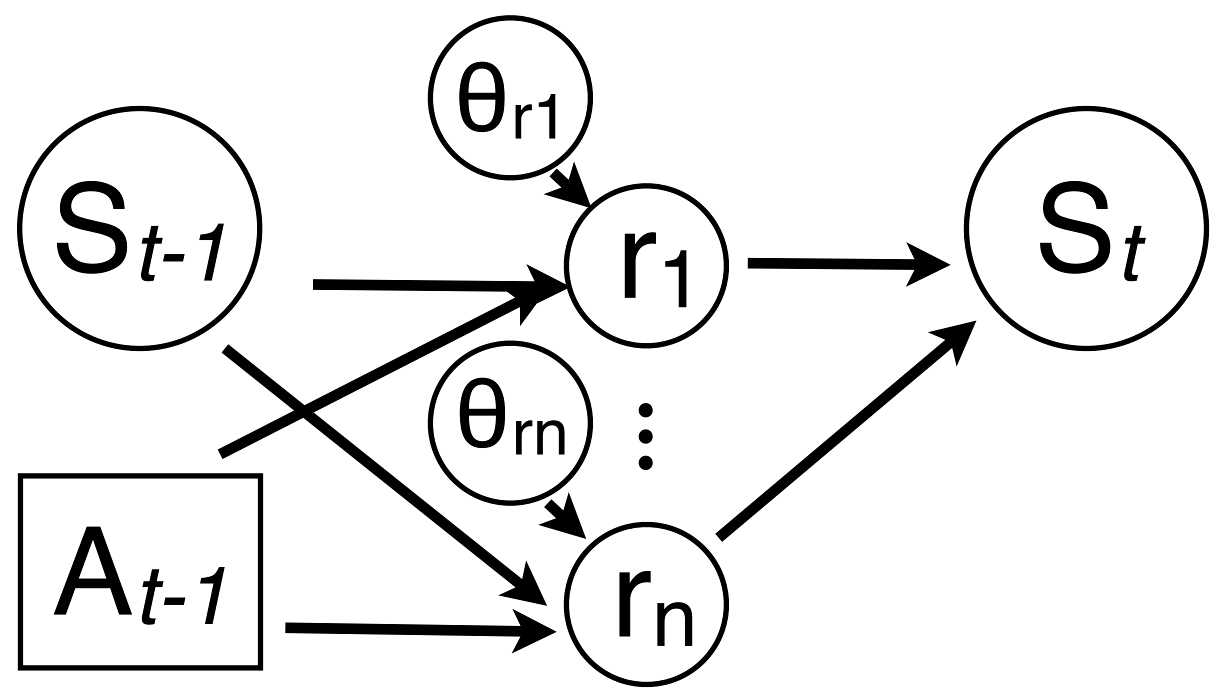
- Two alternative formalisations of the transition model:

**Baseline:**



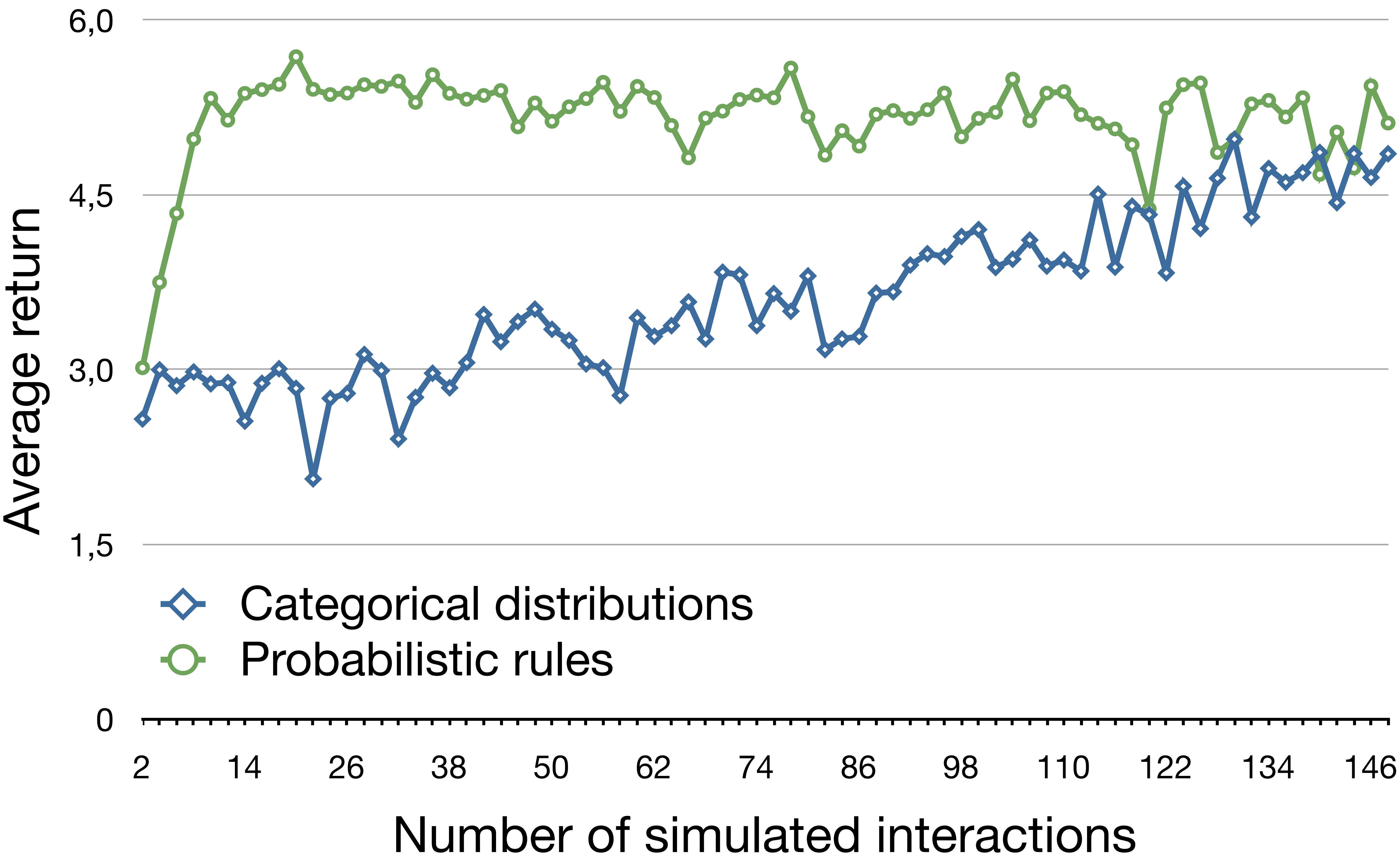

Classical (factored) categorical distributions

**Our approach:**




Model structured with probabilistic rules

# Results: average return



# Conclusion

- *Hybrid* approach to dialogue policy optimisation:
  - Domain models structured with probabilistic rules
  - Rule parameters estimated via model-based Bayesian RL
- Experiment shows that the rule-structured model outperforms a classical factored model
- **Future work**:
  - Evaluate the approach with real interactions
  - Combine offline and online planning