Interactive Learning and Decision Making: Foundations, Insights & Challenges

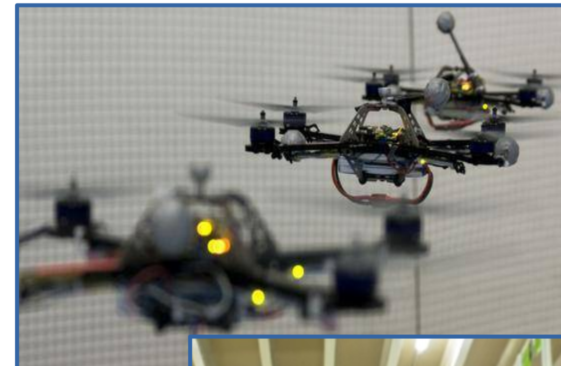
Frans A. Oliehoek

UNIVERSITY OF LIVERPOOL

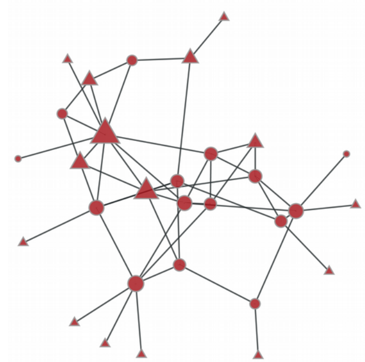
TU Delft

Early Career Spotlight Track – IJCAI 2018, Stockholm

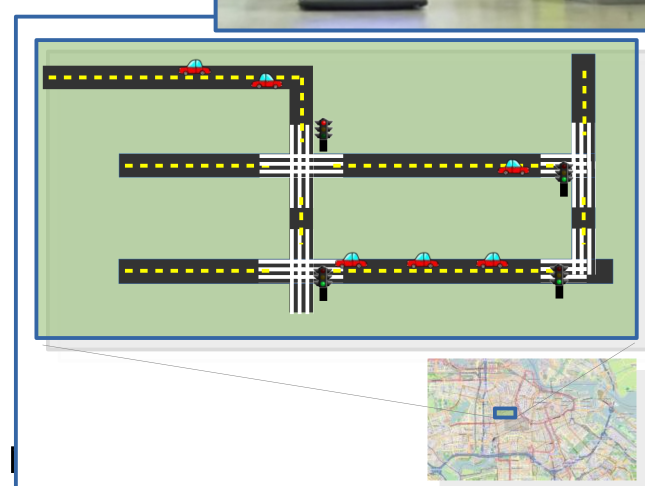
Goal: designing intelligent agents



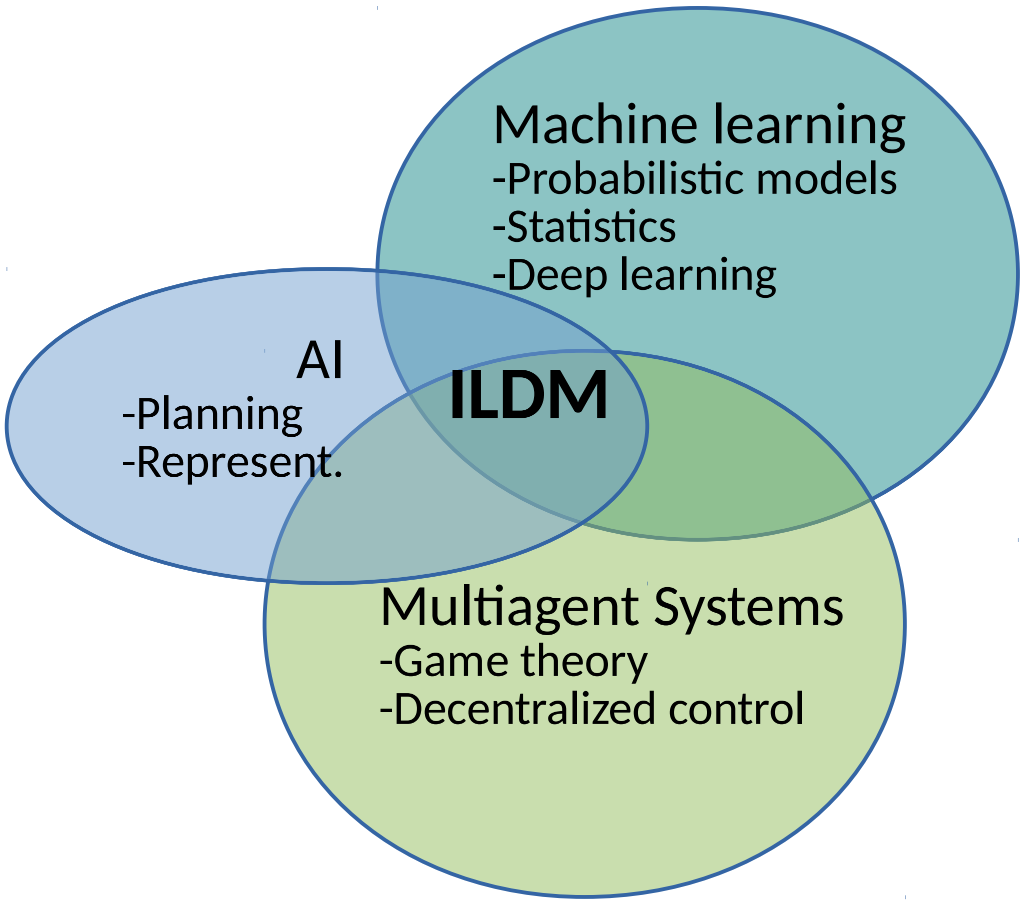
- Design intelligent agents (systems, robots) for complex environments



- Agents will **interact** with...
  - ...each other,
  - ...humans, and
  - ...their unknown environments



What is Interactive Learning & Decision Making (ILDM)?



ILDM =

sequential decision making

+ interaction

Interactive

The Oxford dictionary defines

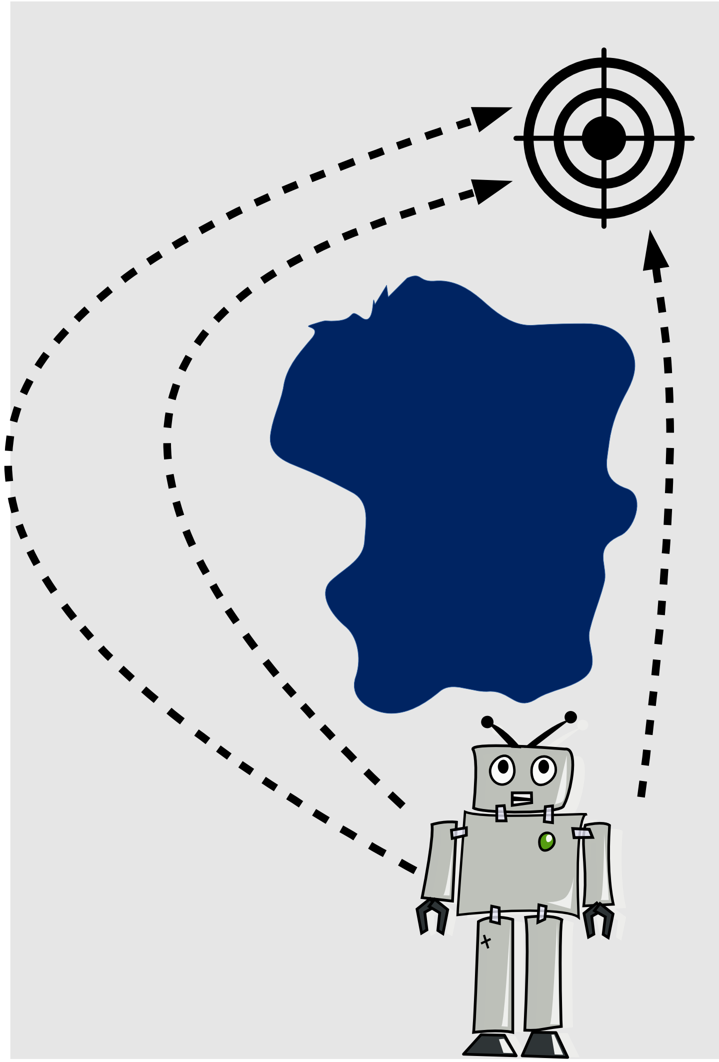
interactive (adj)

- (of two people or things) influencing each other,
- Allowing a two-way flow of information [...]

→ the key characteristic: **two-way flow of influence**.

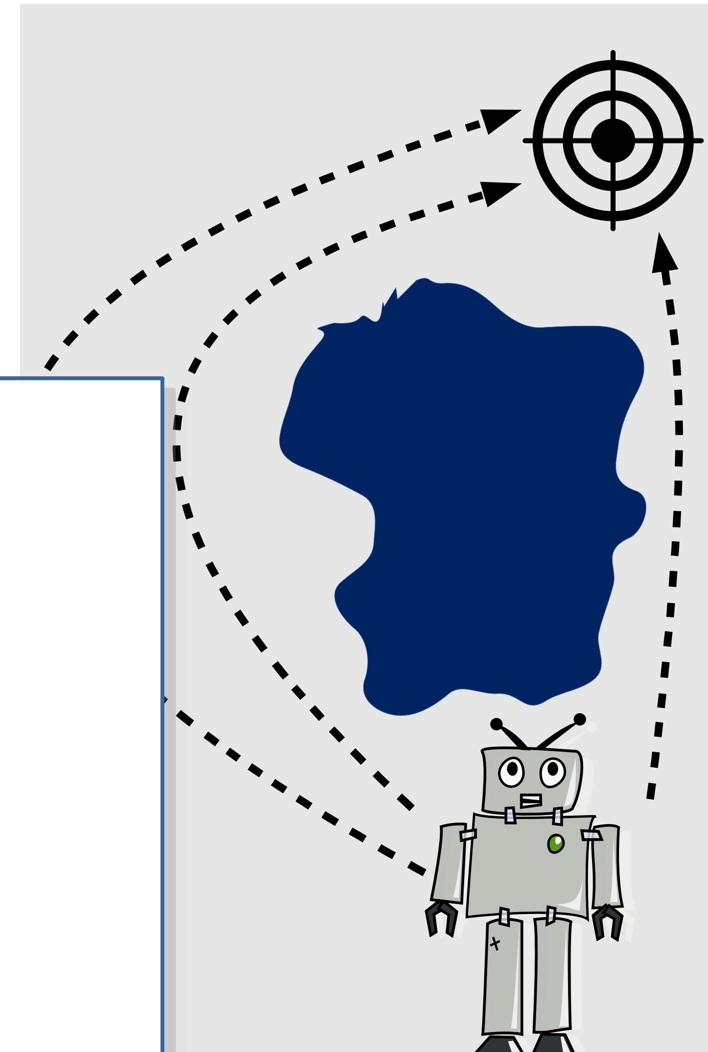
Sequential Decision Making (SDM)

- Actions over multiple time steps
- SDM problems are complex…
  - **immediate** vs **long-term rewards**
  - deal with **uncertainties** (stochasticity, partial information)
- Manual programming is difficult
  - Instead: “programming via rewards”
  - planning / reinforcement learning



Sequential Decision Making (SDM)

- Actions over multiple time steps
- SDM problems are complex...



And **interaction adds to the complexity**...

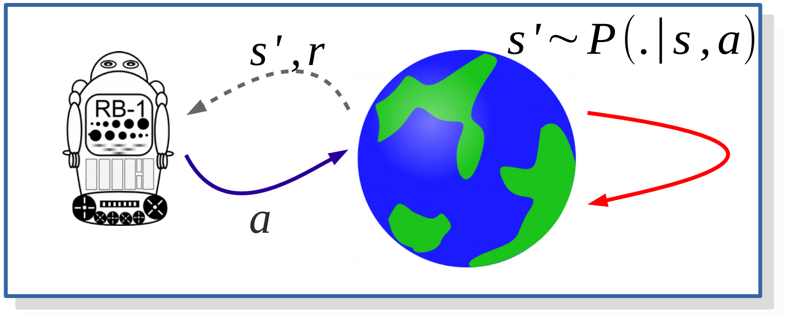
- intelligent agents will live in a **multiagent world**: multiple agents / humans

→ Near-impossible to manually program
→ How? Need principled methods!

Foundations

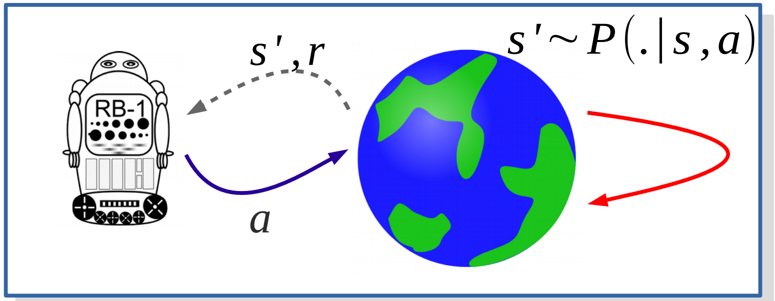
ILDM... How? Formal Models!

- Community realized that action uncertainties need representation
  - MDP embraced by RL community
- For a long time: “POMDPs are intractable...”
  - scalability has seen great progress (e.g., POMCP)
  - more and more applications emerging!



ILDM... How? Formal Models!

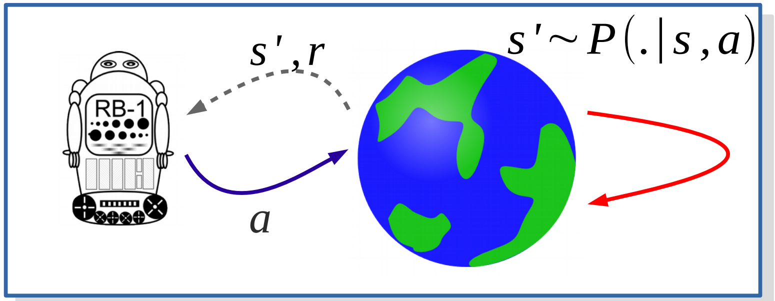
- Community realized that action uncertainties need representation
  - MDP embraced by RL community
- For a long time: “POMDPs are intractable...”
  - scalability has seen great progress (e.g., POMCP)
  - more and more applications emerging!



No:
“selecting optimal actions under partial observability” is intractable!

ILDM... How? Formal Models!

- Community realized that action uncertainties need representation
  - MDP embraced by RL community



- For a long time: “POMDPs are intractable...”
  - scalability has seen great progress (e.g., POMCP)
  - more and more applications emerging!

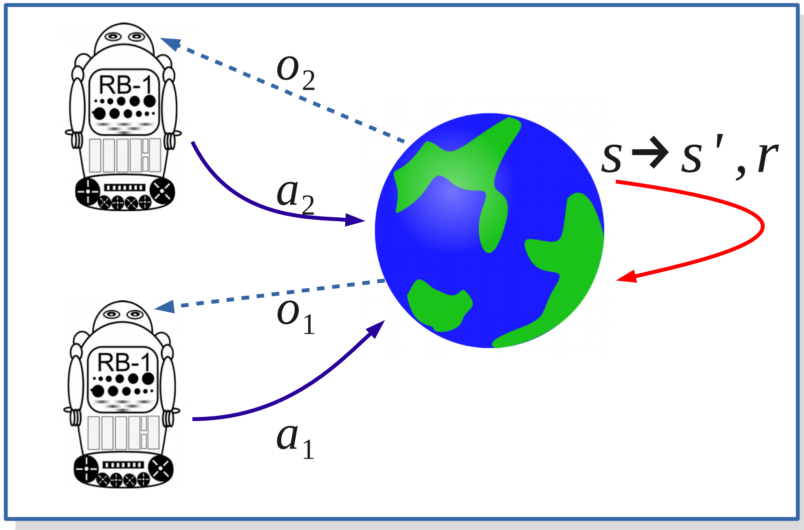
No:
“selecting optimal actions under partial observability” is intractable!

- Bottom line: we have seen great progress due to

$\rightarrow$ **minimal models** that **represent all relevant aspects** $\leftarrow$

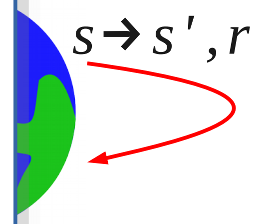
- So... if your problem has multiple decision makers
  $\rightarrow$ you should probably represent that.

Decentralized POMDPs



- A minimal framework for
  - multiple cooperative agents
  - stochastic environments
  - state uncertainty
- A Dec-POMDP $\langle S, A, P_T, O, P_O, R \rangle$
  - n agents
  - S – set of states
  - A – set of **joint** actions $\quad a = \langle a_1, a_2, \ldots, a_n \rangle$
  - P_T – transition function $\quad P(s' | s, a)$
  - O – set of **joint** observations $\quad o = \langle o_1, o_2, \ldots, o_n \rangle$
  - P_O – observation function $\quad P(o | a, s')$
  - R – reward function $\quad R(s, a)$
- Act based on **individual** observations

Decentralized POMDPs



Yes, these are horribly complex to solve optimally...

- ►NEXP-complete [Bernstein et al. 2000]
- ►but no easy way out - this is a minimal model.

...but we are making steady progress

- ►E.g., multi-robot systems - Christopher Amato et al.

- PR2 + 2 turtlebots for faster drinks delivery!

What does it buy us?

- Optimal plans need to trade-off:
  - immediate vs long-term reward (as in MDPs)
  - knowledge gathering vs exploitation (as in POMDPs and/or RL)
  - exploiting individual knowledge vs being predictable

- Using Dec-POMDPs (and similar models) we can study quantitatively and qualitatively the effect of interaction.

Some Insights

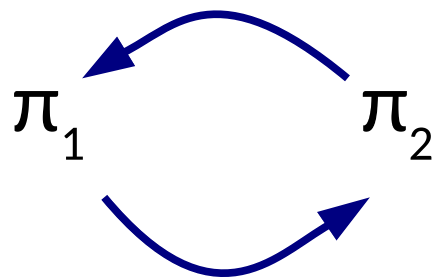
- This talk, zoom in on just **two insights**:
  - **influence-based abstraction (IBA)**
  - **transfer planning (TP)**

See paper for a more...!

Insights:
Influence-based Abstraction & Search

Multiagent Influences

- Interaction: the combination of policies matters

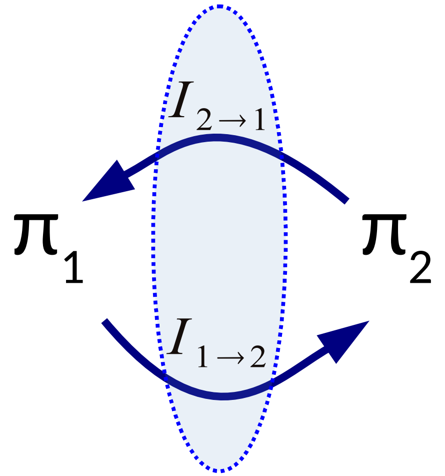


Policies influence each other

best response π_2 depends on π_1

Multiagent Influences

- Interaction: the combination of policies matters

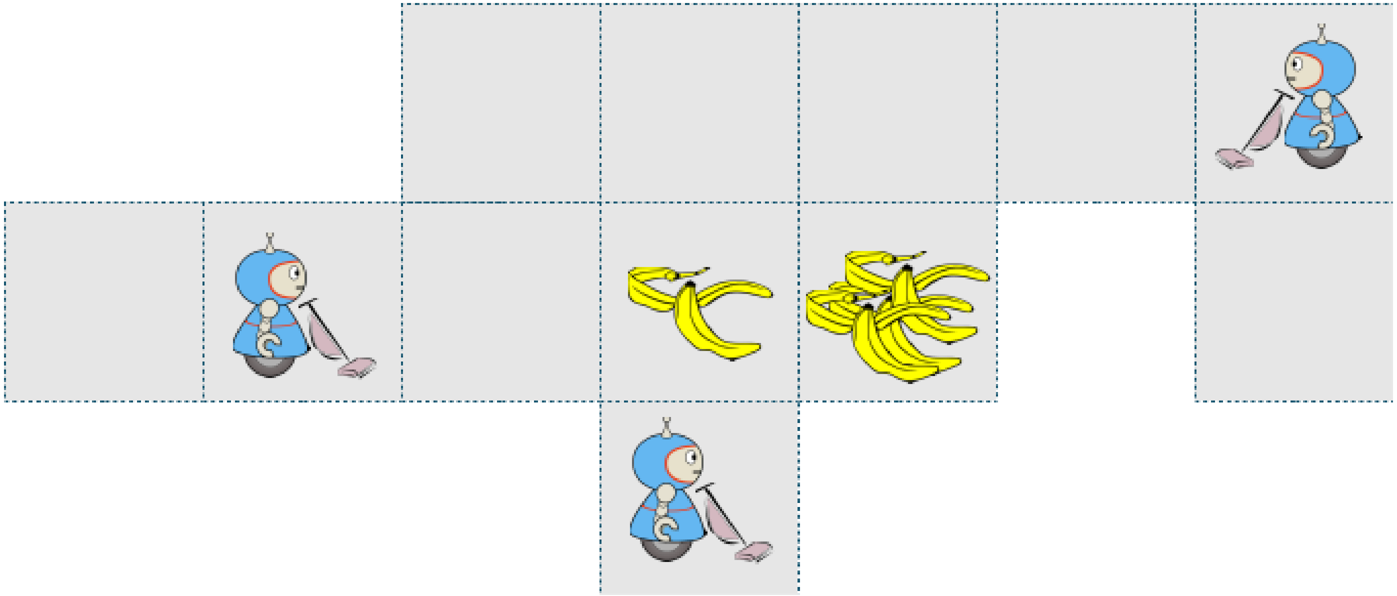


Influence-Based (Policy) Abstraction:
summarize the influence of the policy of the other

[Becker et al. 2003, Becker et al., 2004, Varakantham et al., 2009, Witwicki and Durfee, 2010b, Velagapudi et al., 2011, Oliehoek et al. 2012]

Influence representations – Intuition

- Spatial Task Allocation Problems
  [Claes et al. AAMAS 2015, 2017]

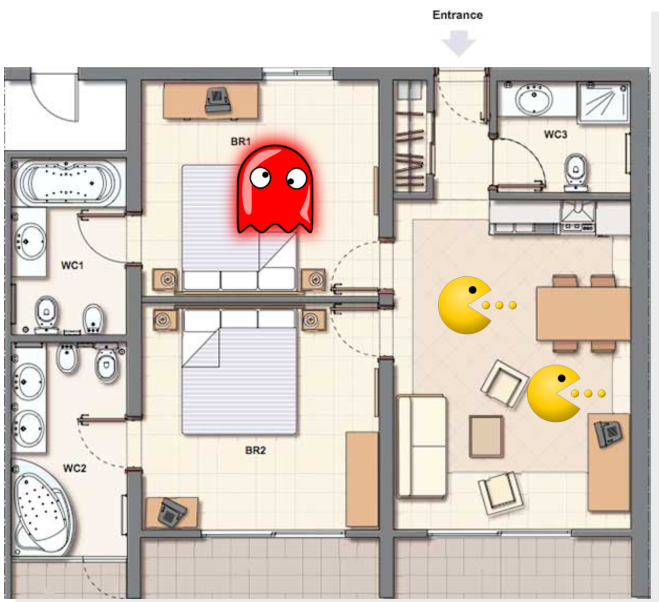


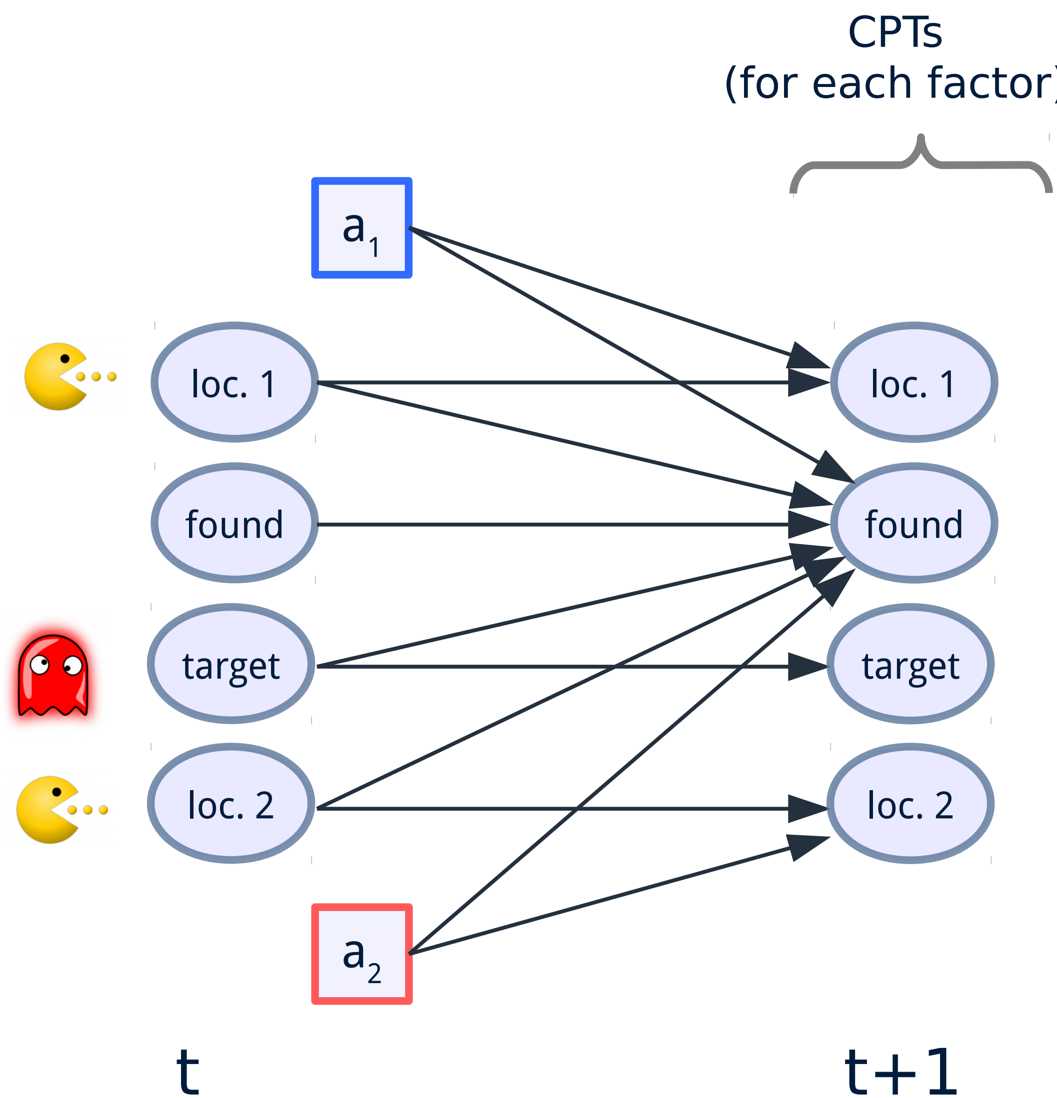
- Task locations known

→ what is the prob. that each task will be addressed by teammates?

Example: HouseSearch [Witwicki et al. AAMAS, 2012]

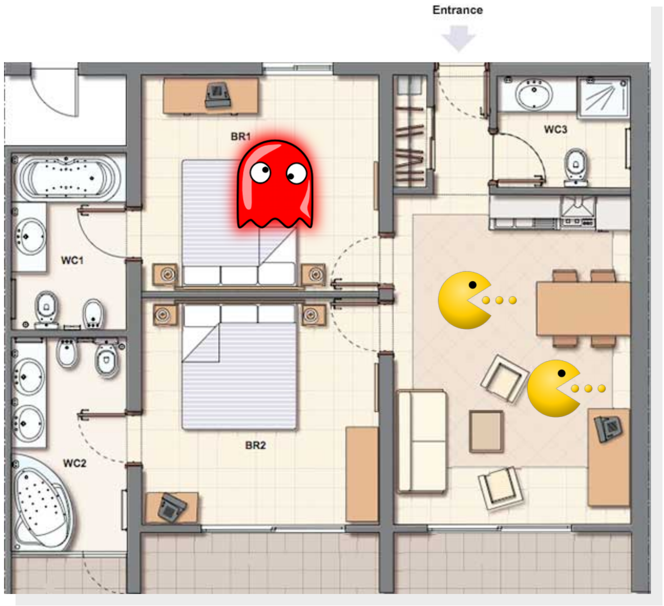
- Team of agents
- A target

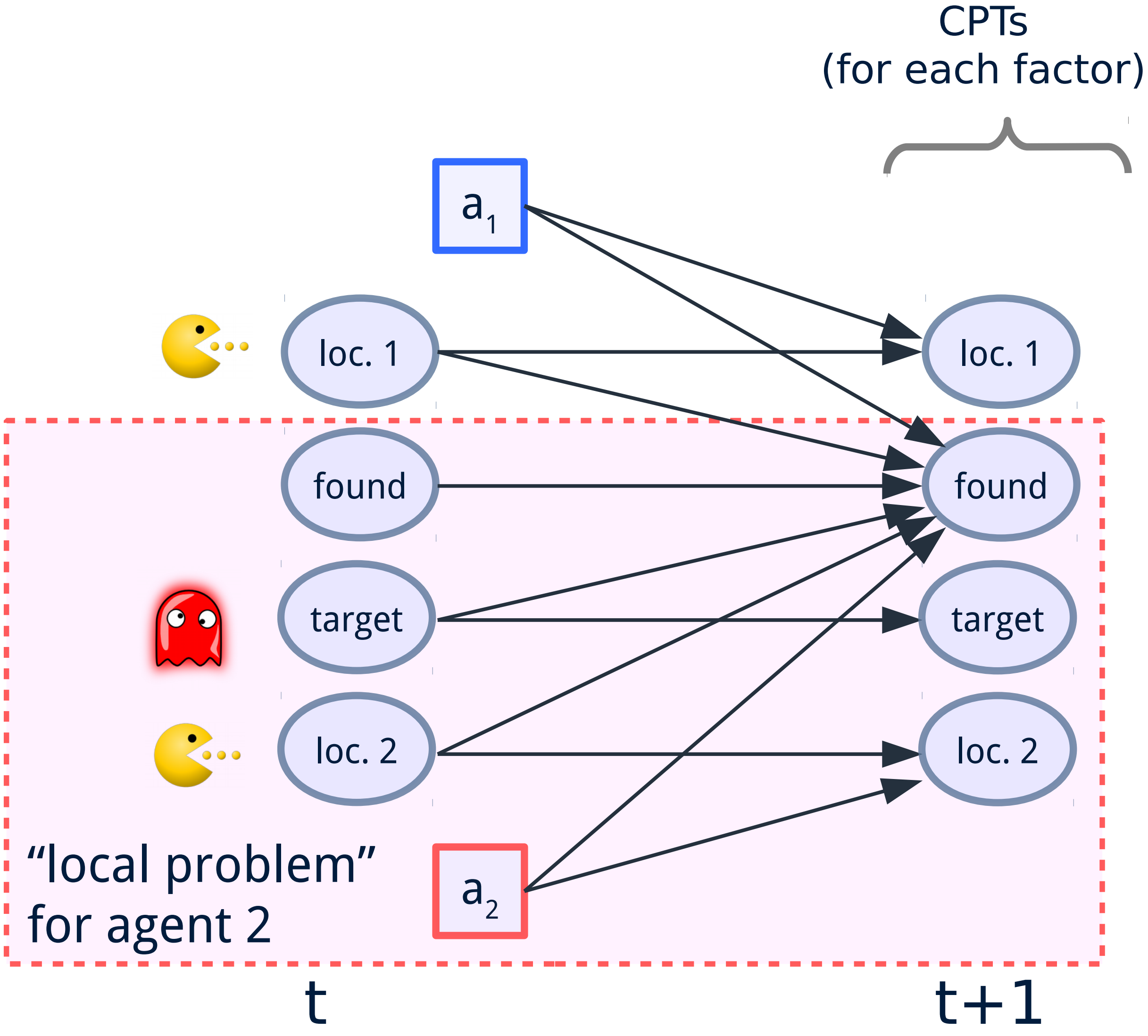




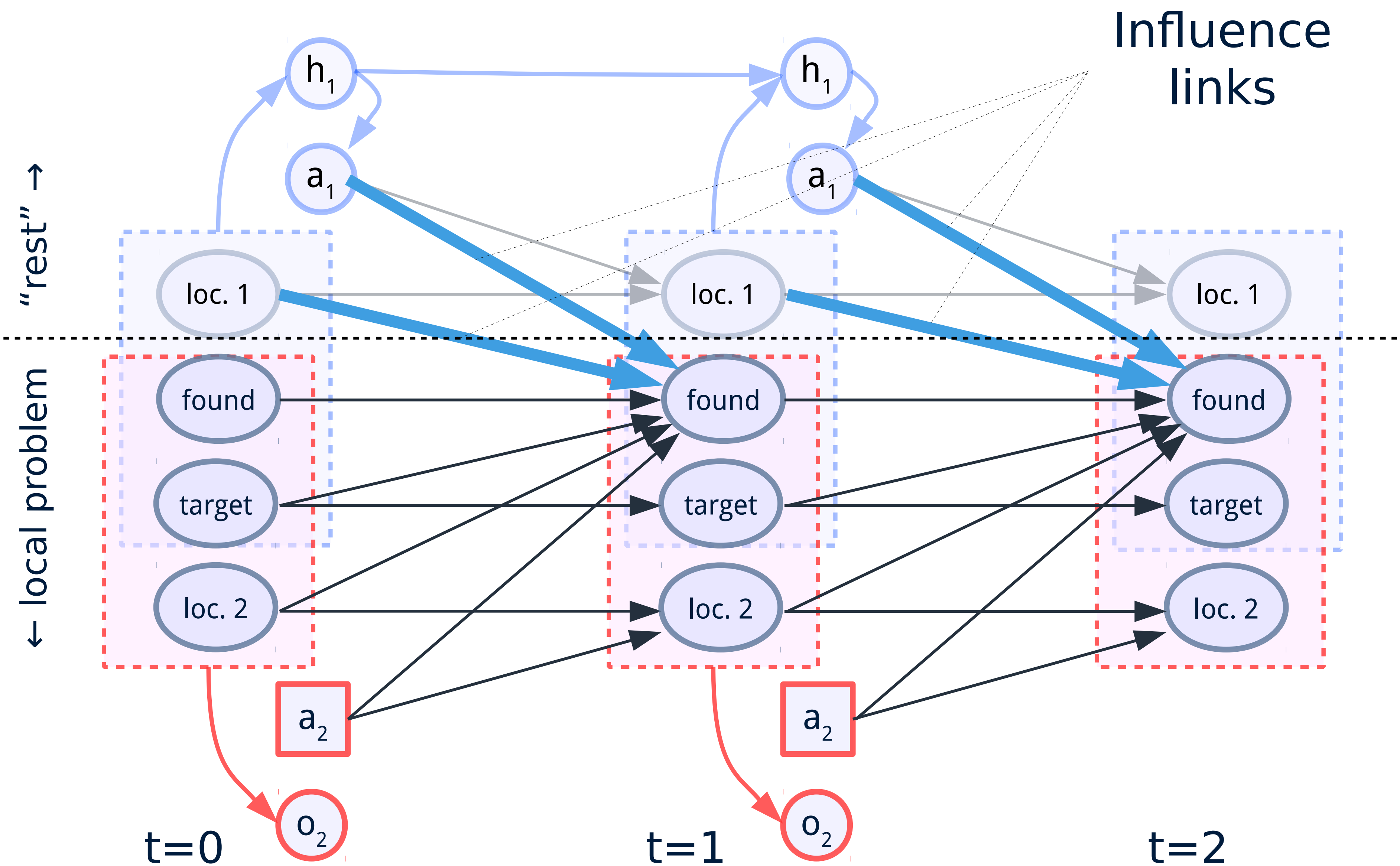
Example: HouseSearch [Witwicki et al. AAMAS, 2012]

- Team of agents
- A target

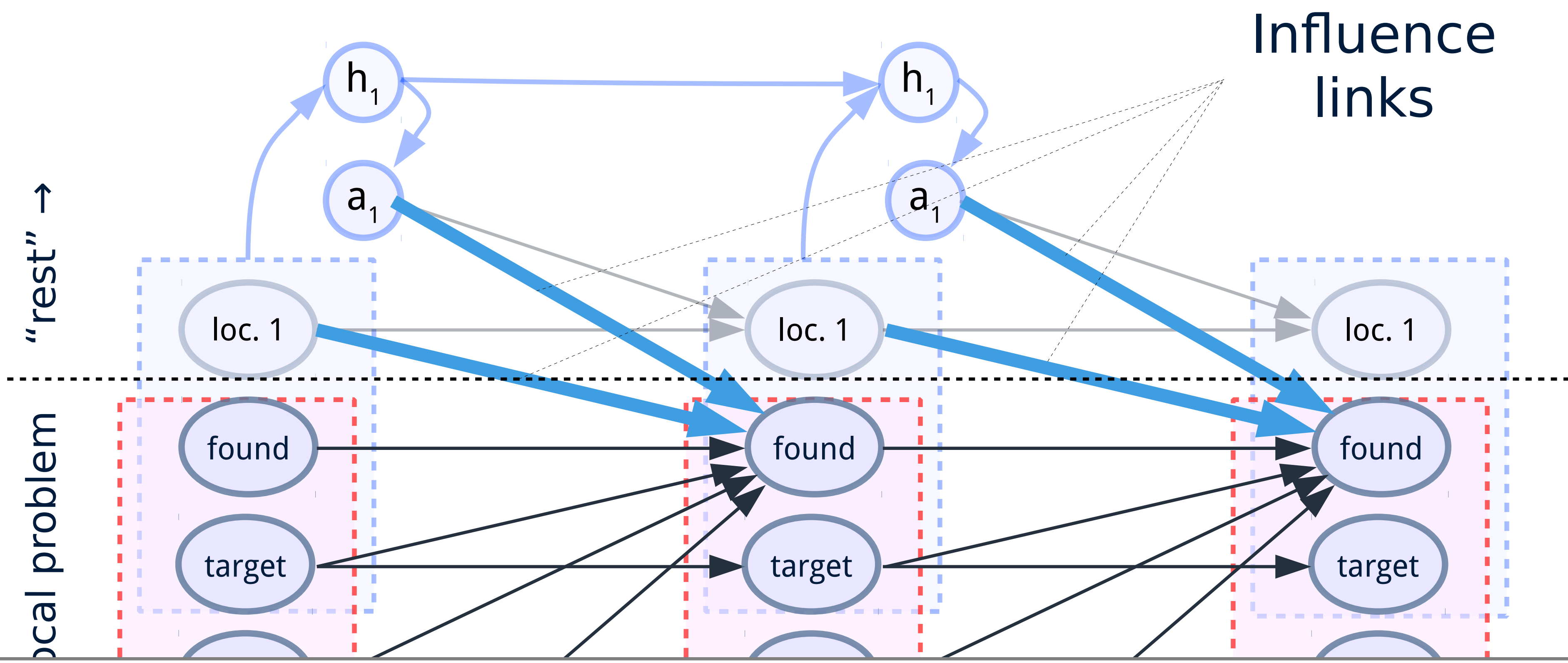




Agent 2's Perspective



Agent 2's Perspective

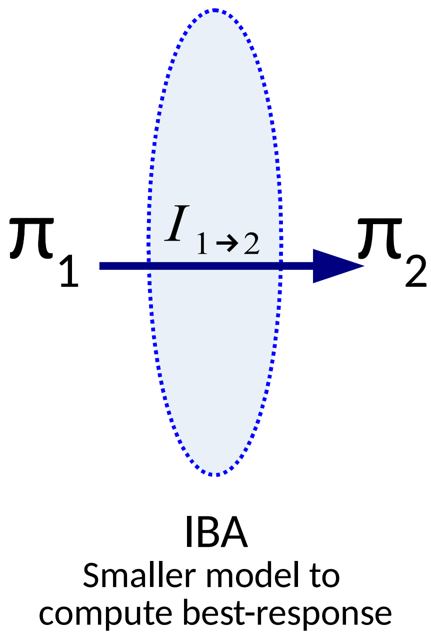


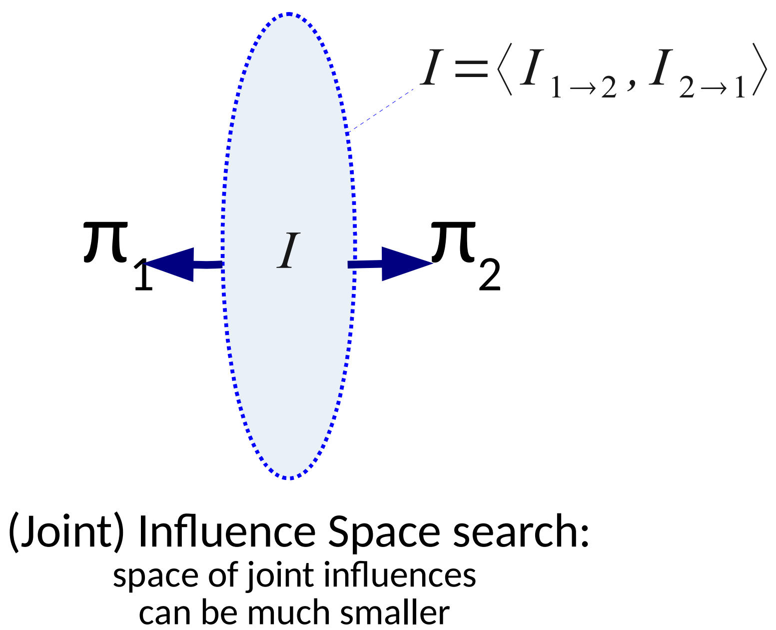
- ▸ If we new the values of **influence sources** in advance...
- ▸ But can compute a distribution over them
  - need to condition on some stuff, D, in the local problem
- ▸ So, an **influence point** is a collection { $P(x_{sources} \mid D)$ }

An inference task

Why/How To Use Influences?

- **Influence-based abstraction (IBA)**: local best-response
- **Influence search**: joint optimization

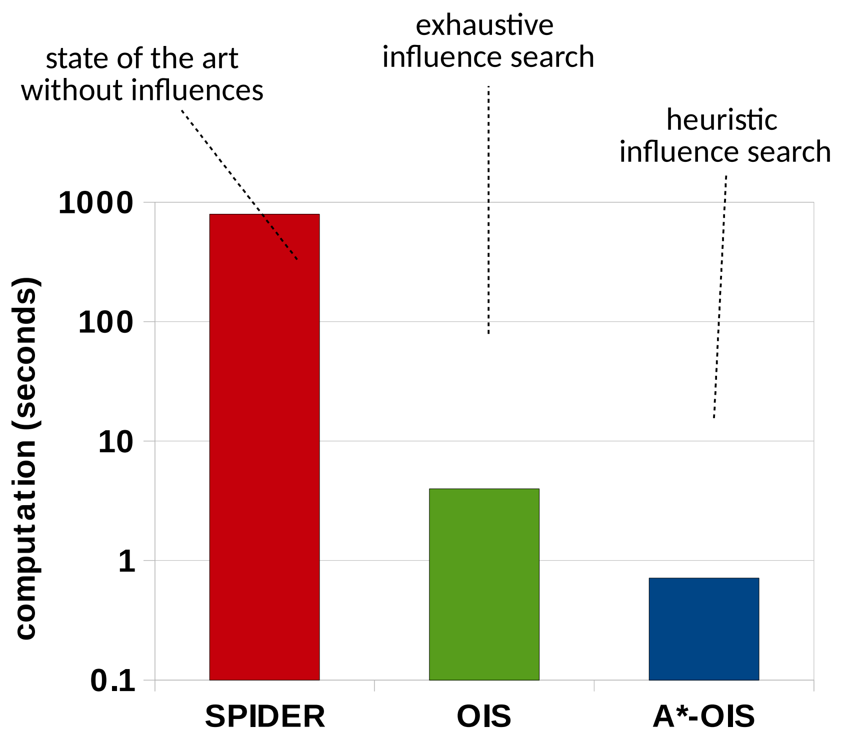


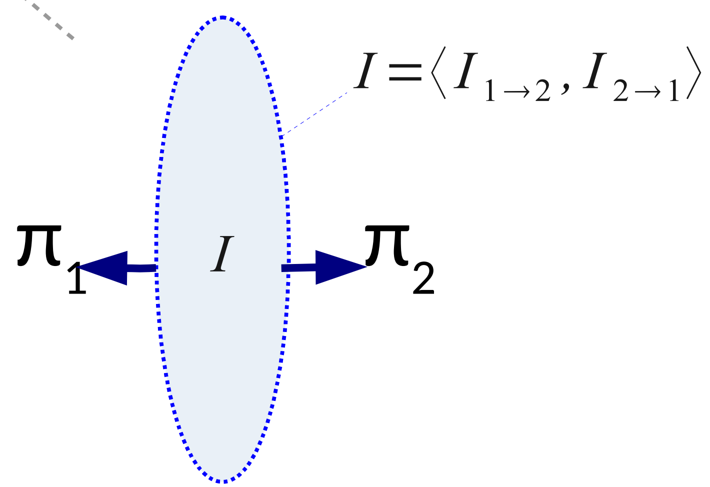


Why/How To Use Influences?

- **Influence-based abstraction (IBA)**: local best-response
- **Influence search**: joint optimization

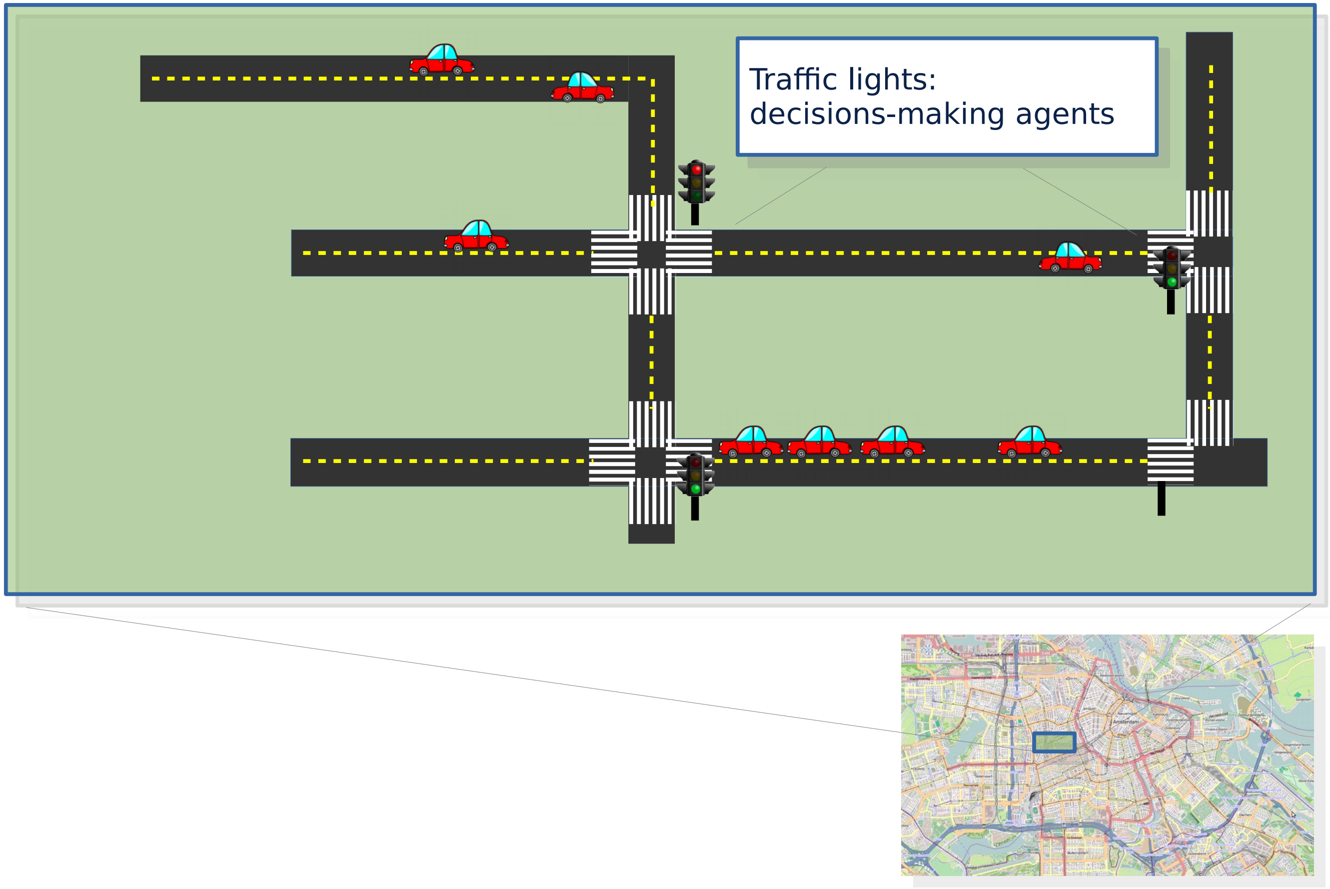
Heuristic Influence Search:
large speed-ups [Witwicki et al. AAMAS 2012]





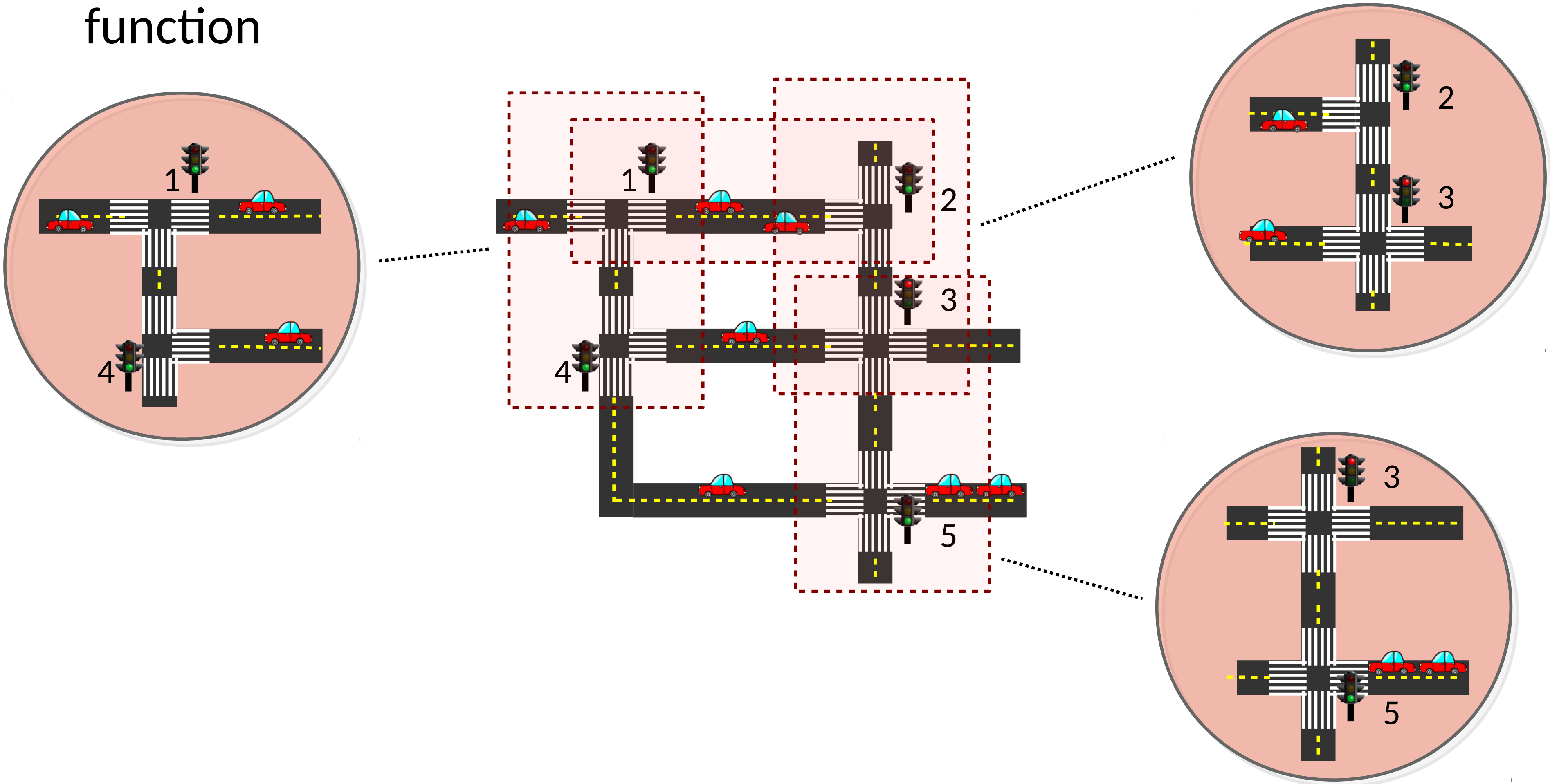
(Joint) Influence Space search:
space of joint influences
can be much smaller

Insights:
Scaling via Transfer Planning



Transfer Planning (TP)

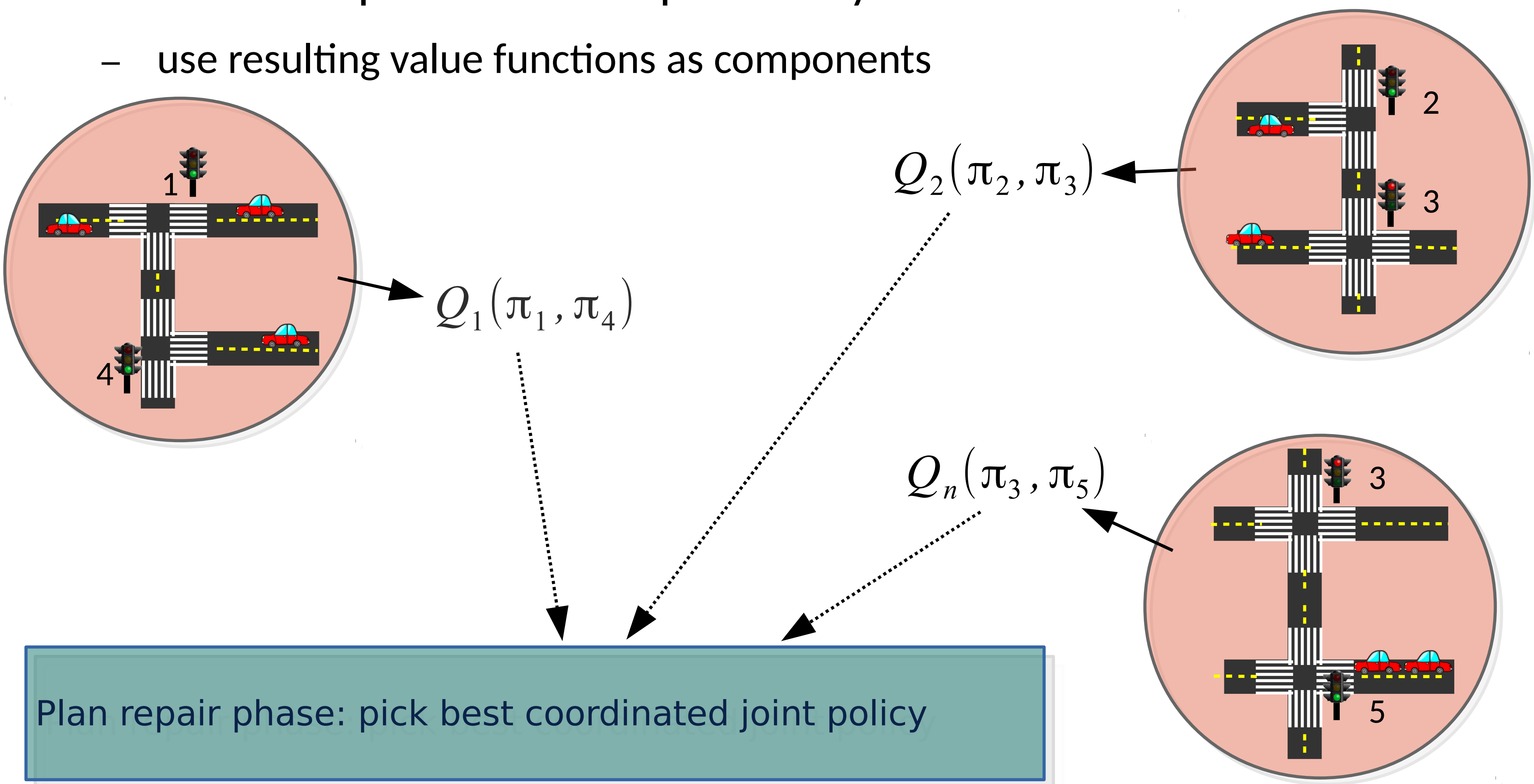
- Define **source problem** for each component of the approximate value function



Frans A. Oliehoek, Shimon Whiteson, and Matthijs T. J. Spaan. Approximate Solutions for Factored Dec-POMDPs with Many Agents. In Proceedings of the Twelfth International Conference on Autonomous Agents and Multiagent Systems, pp. 563–570, 2013.

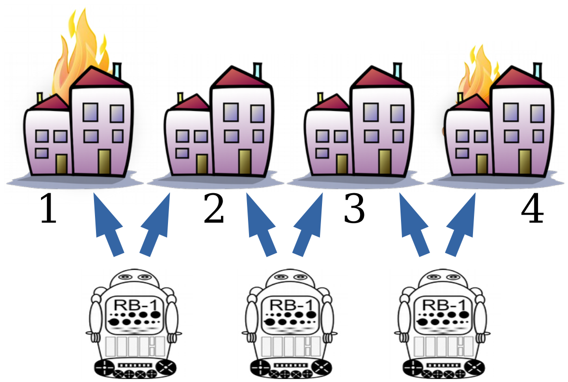
TP for Dec-POMDPs

- Solve source problems independently
  - use resulting value functions as components

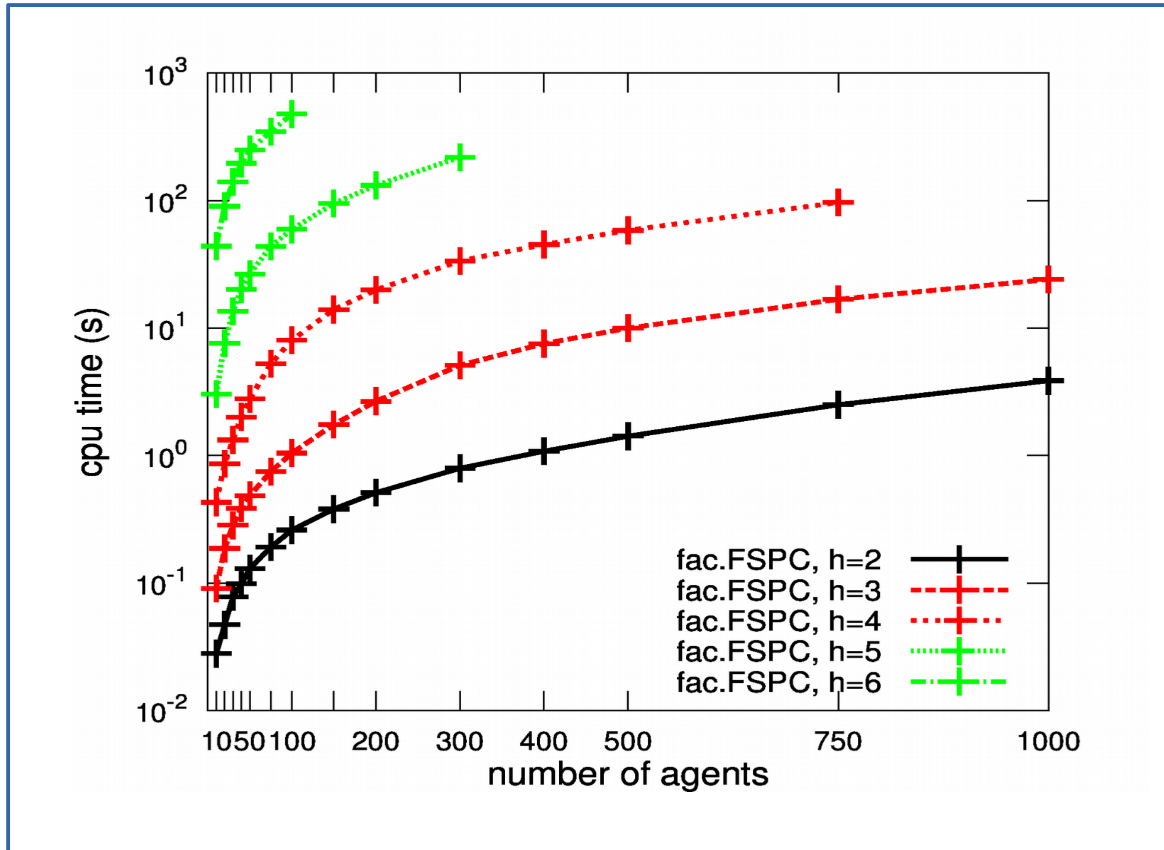


What can we say about TP?

- Unprecedented scalability: 100s of agents



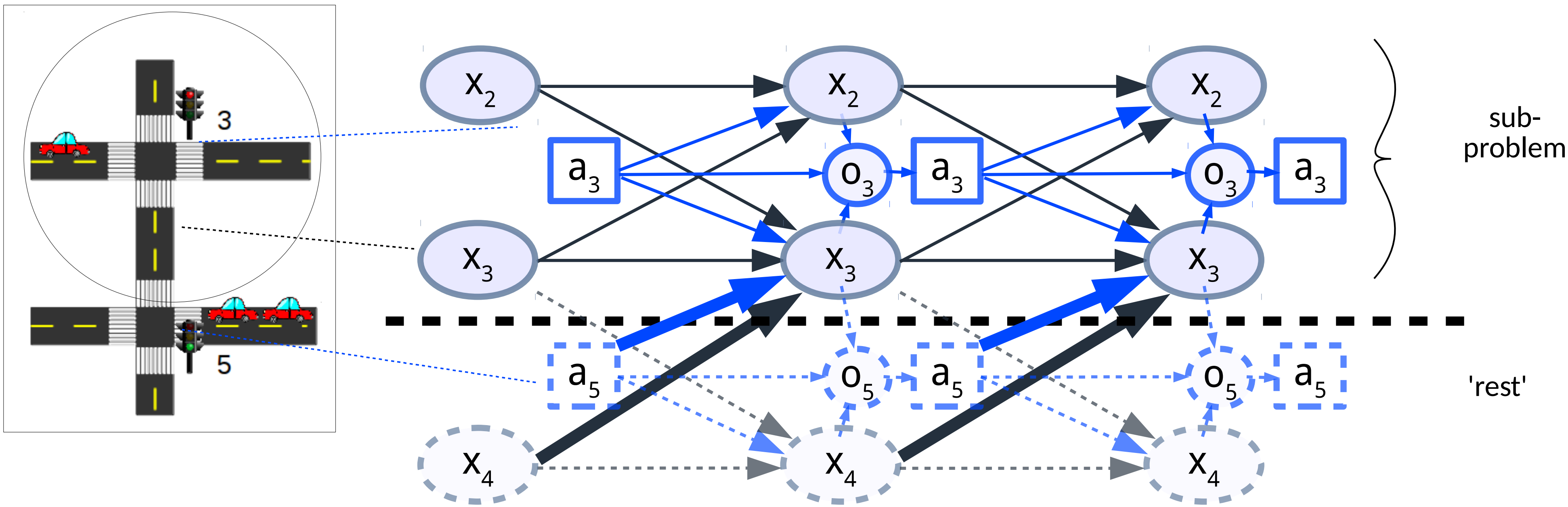
“Firefighting Graph (FFG)” benchmark



But completely heuristic:
No guarantees on solution quality....

Bounds via Influence-optimism

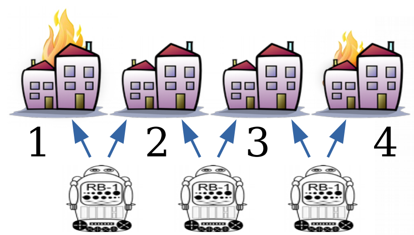
- Compute local upper bound on value [Oliehoek et al. IJCAI 2015]

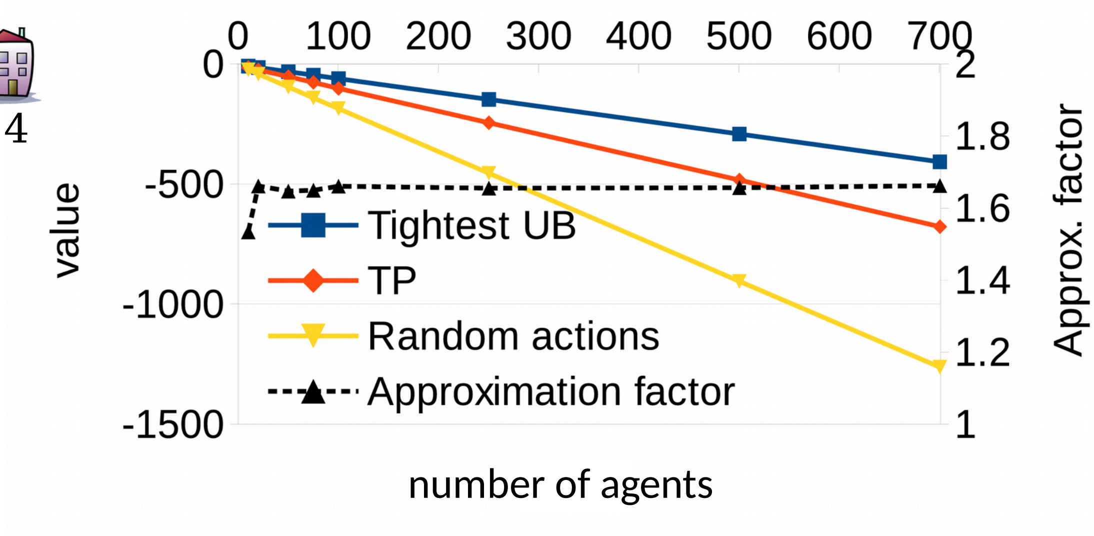


Approach:
- ► Be optimistic about the influence sources!

Upper Bounds for Large Problems

- Provide UB for interpretation of Transfer Planning [Oliehoek et al. '15 IJCAI]

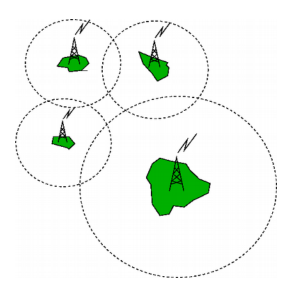




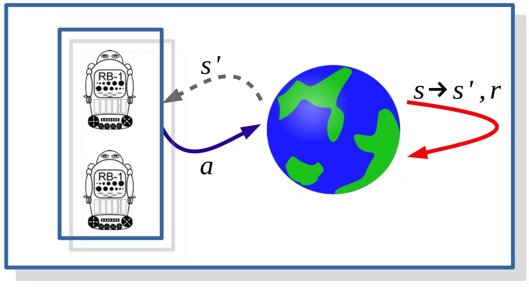
essentially optimal

n	50	75	100	250
V^{TP}	−71.99	−111.07	−148.70	−382.47
$\hat{V}^{IO}$	−72.00	−107.06	−144.00	−360.00
EAF	1.00	1.04	1.03	1.06

Table 1: Empirical approximation factors for Aloha ($h = 3$) with varying number of agents.



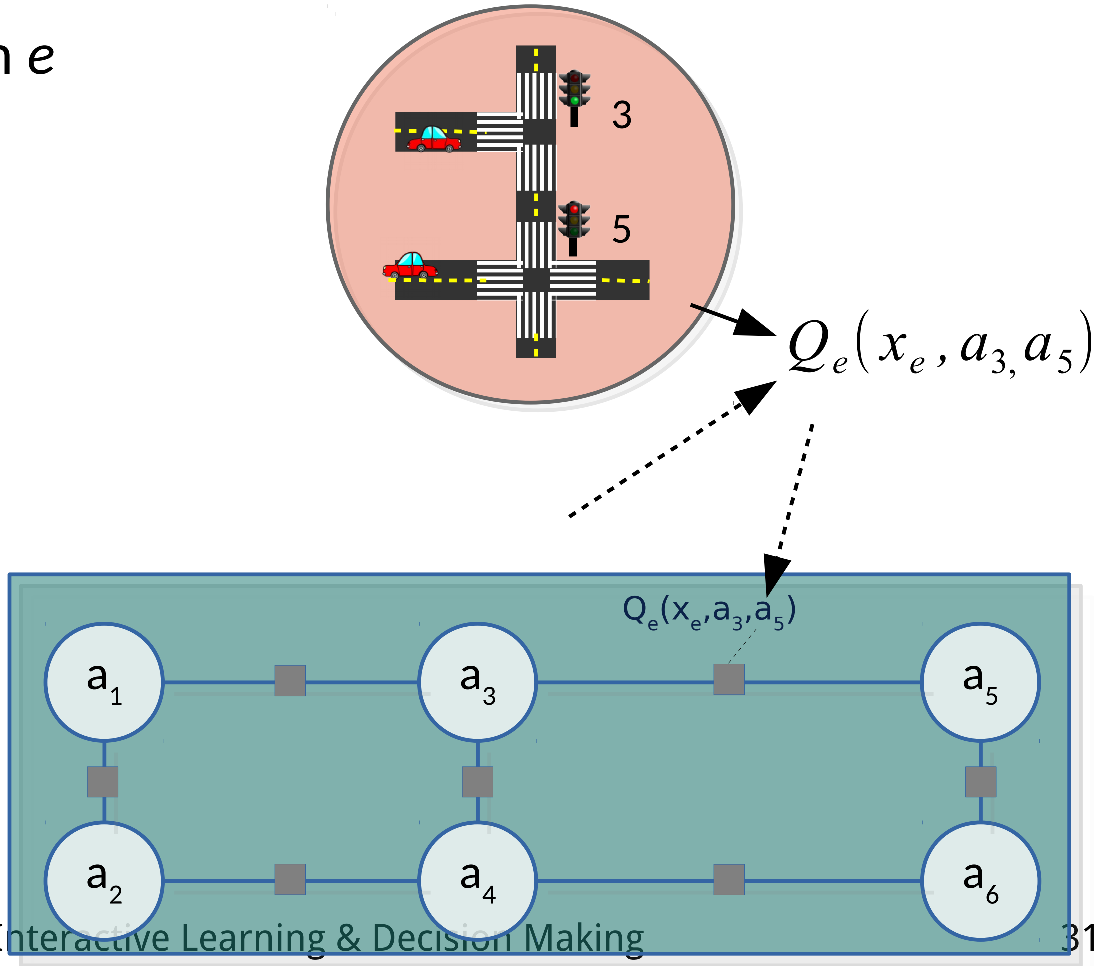
Transfer Planning for MMDPs



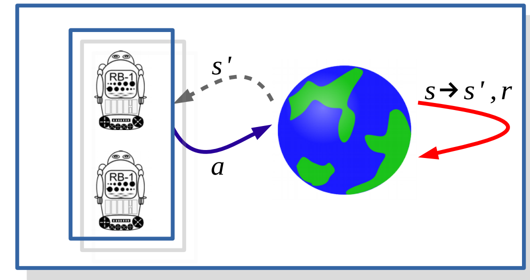
- Can also use TP when agents can synchronize: multiagent MDPs (MMDPs)

- **off-line planning phase**:
  - solve each source problem *e*
  - compute $Q_e(x_e,a_e)$ for each

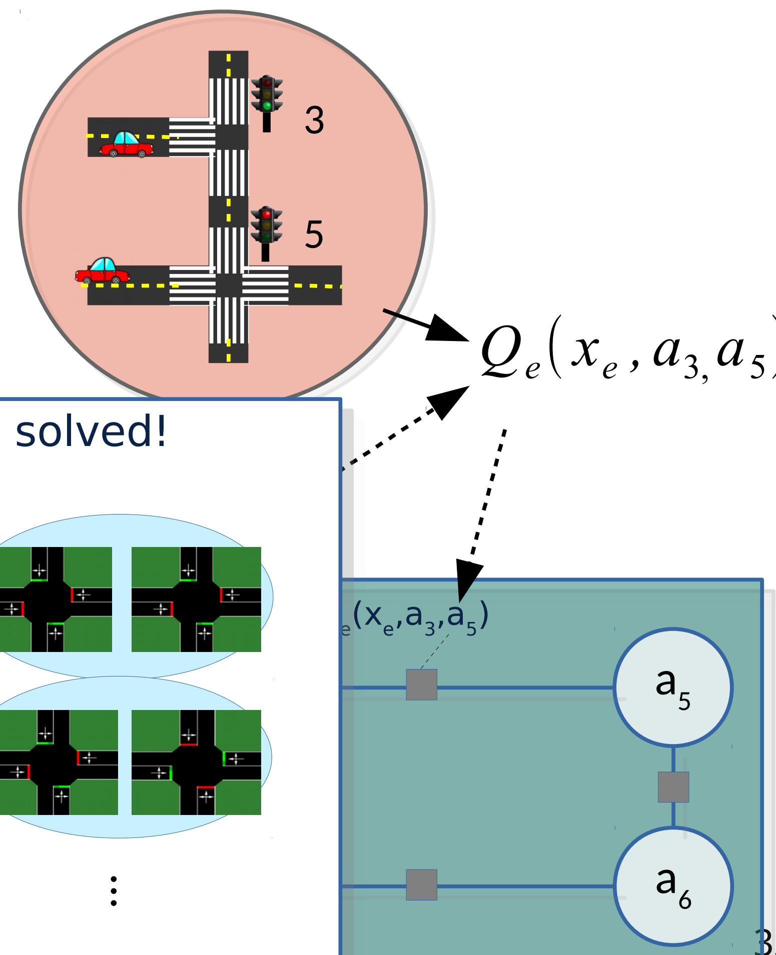
- **on-line execution phase**:
  - coordinated action via message passing



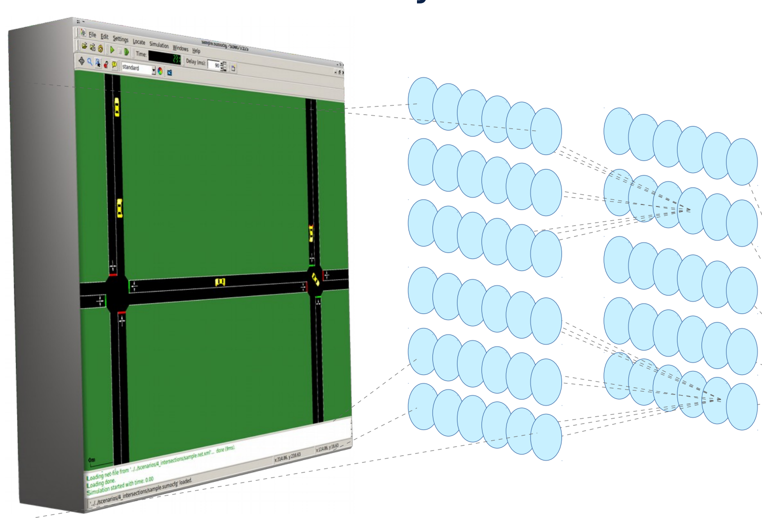
Transfer Planning for MMDPs

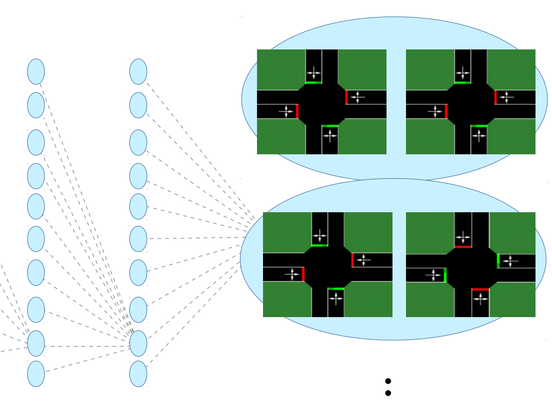


- Can also use TP when agents can synchronize: multiagent MDPs (MMDPs)
- **off-line planning phase**:
  - solve each source problem *e*
  - compute $Q_e(x_e,a_e)$ for each
- **on-line execution phase**:
  - coordinated action



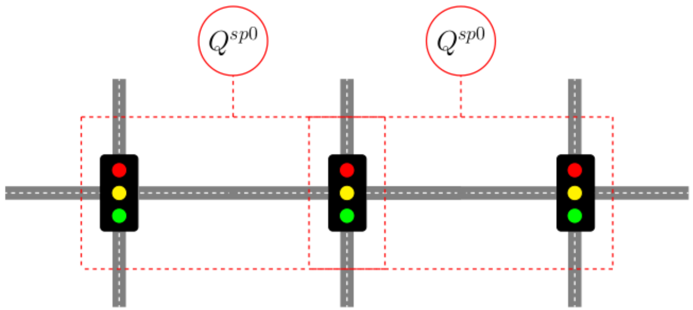
And free to choose way in which source problems are solved!

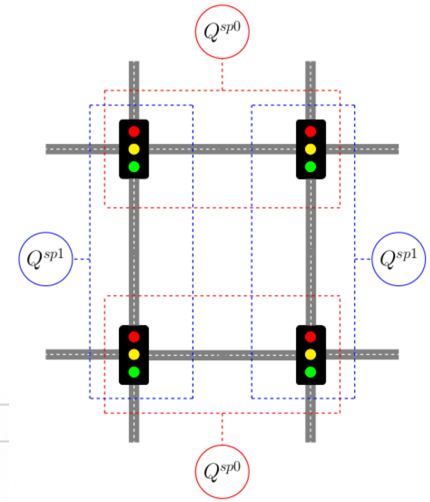


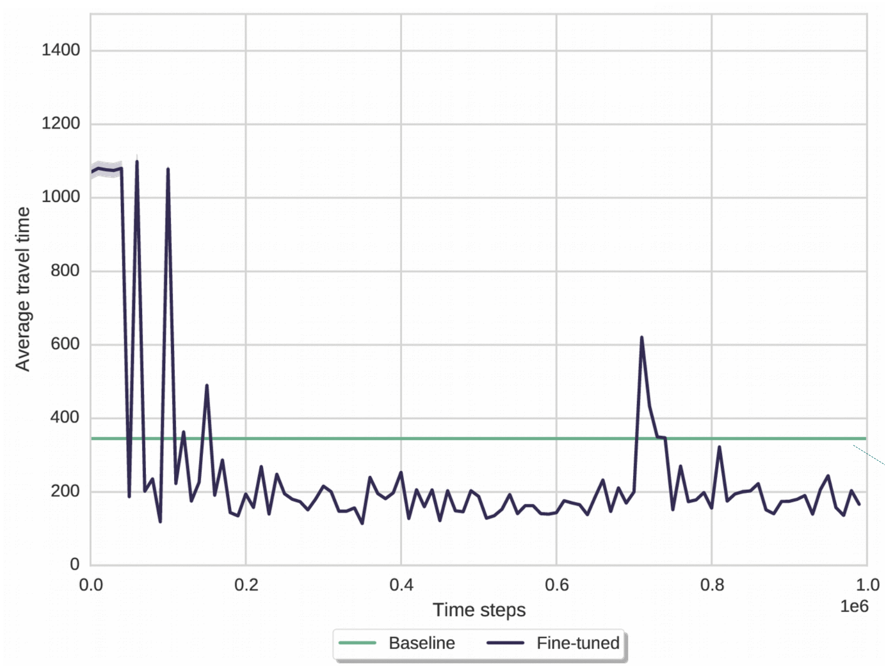


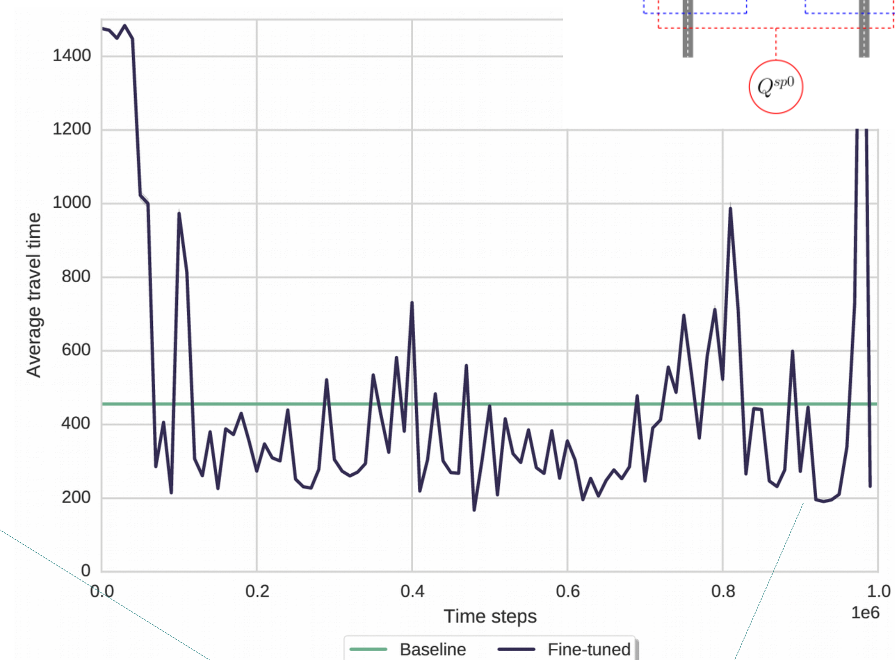
Coordinated Deep RL

- "DQN-TP" [Van der Pol & Oliehoek, 2016]





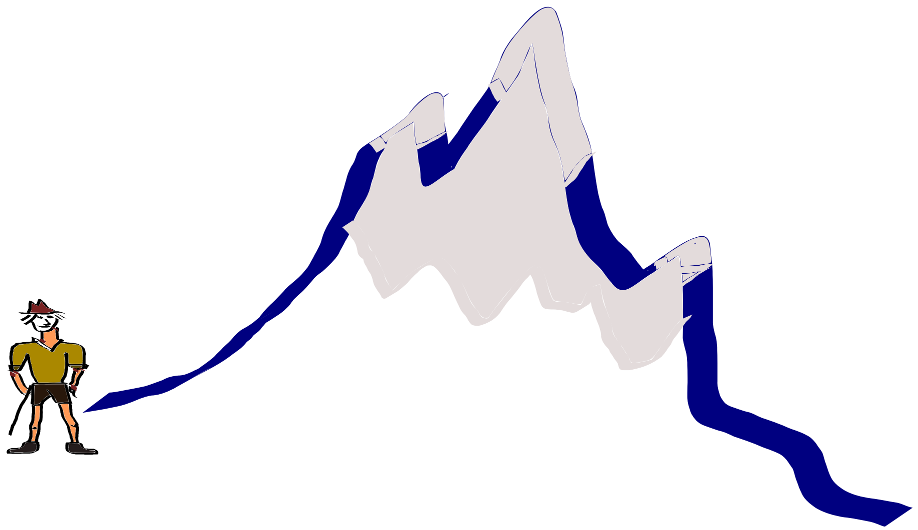




Factor 2 improvements over baseline [Kuyer et al. 2008 ECML]

- Also: http://www.fransoliehoek.net/trafficvideo

Challenges



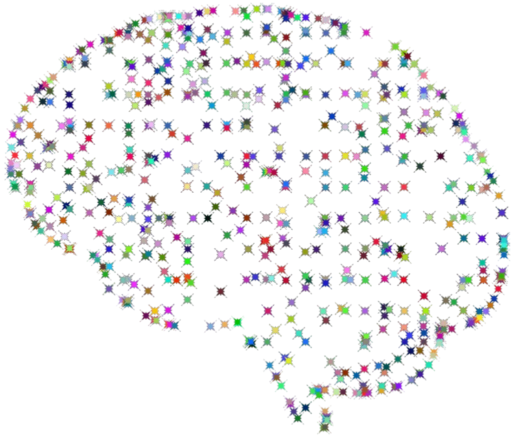
MARL (incl. “deep” MARL)

- Deep RL has shown ability to scale to impressive domains
- But… much MARL does not take into account interaction explicitly
  - Individual Q-learners: may work… or not.
- Some deep MARL does. [E.g., Foerster et al.'16, Mordatch&Abeel'18, Foerster et al.'18, etc.]
- Challenges:
  - **scalability** in number of agents remains a challenge
  - truly **decentralized** learning remains challenging
    - e.g., policy gradient works for Dec-POMDPs [Peshkin et al 2000] but still requires observation of return of entire system.

Learning Models (incl. of other agents and humans)

- Researchers turning to model-based RL: learning ("world") models.
- But how can we learn models of interaction/interactive settings?
  - E.g. how to model human behavior in a warehouse?
  - Progress? Need benchmark simulators?

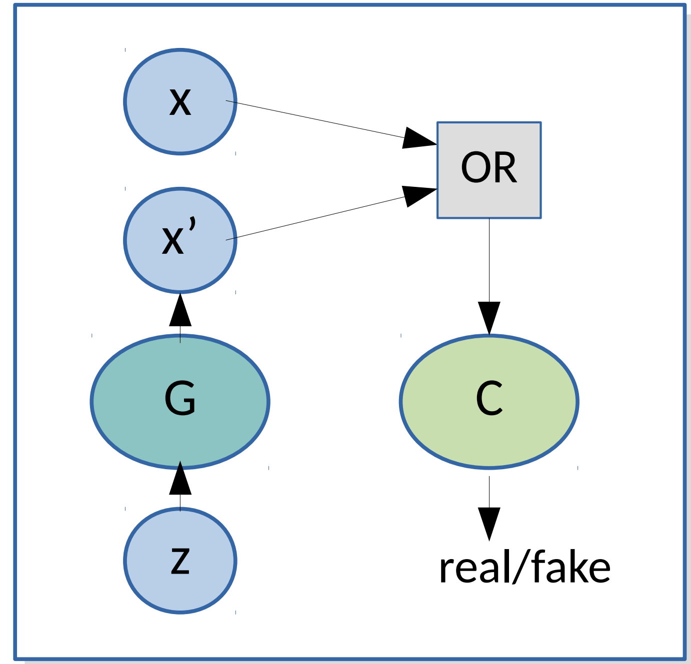
Understanding Interactive Learning

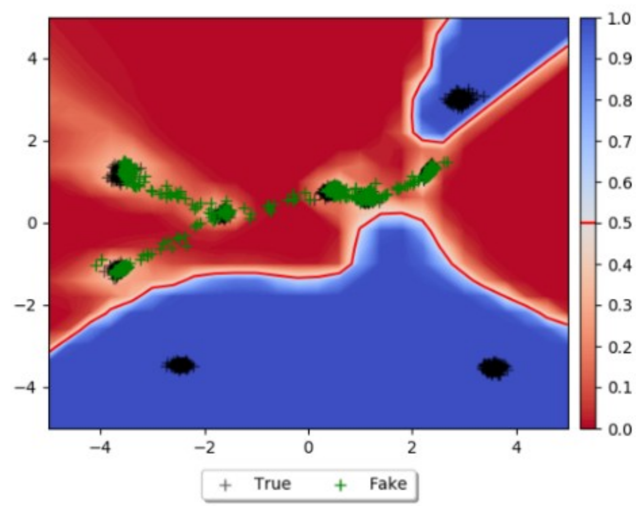


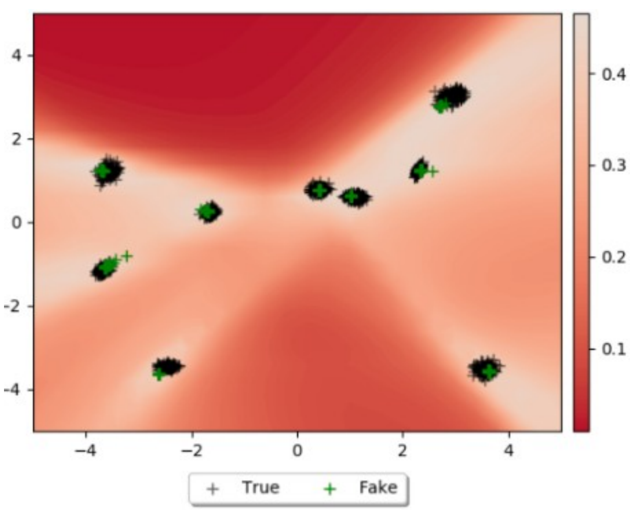
- Interaction at basis of successful learning paradigms:
  - Self-play
  - Learning from demonstration
  - Active learning
  - Learning in competition (e.g., GANs)
- Challenges:
  - better understanding when/how these work?
  - insights from MAL and game theory to improve these?

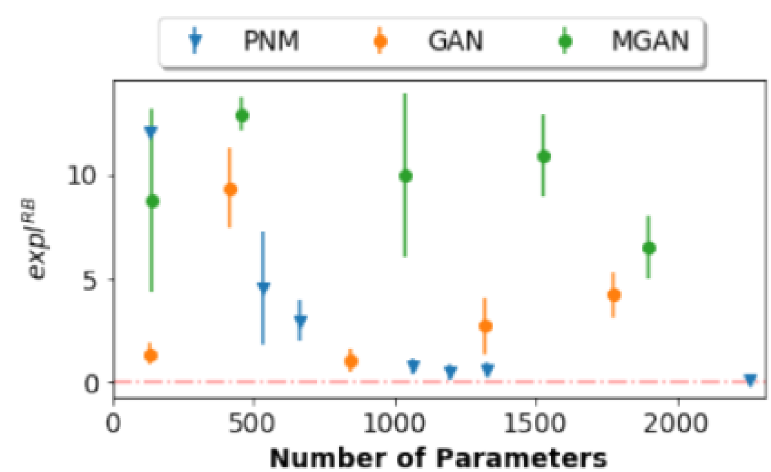
E.g. GANs [Oliehoek et. al 2018 arxiv]

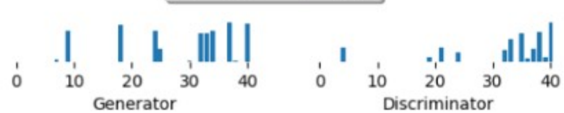
- Using game-theoretic methods:
  - avoid ‘local Nash equilibria’
  - better, more robust solutions







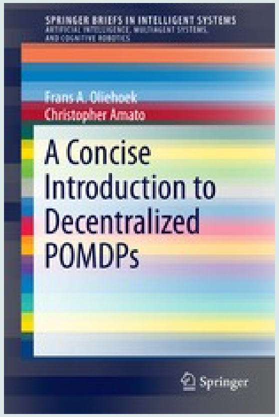




Summary

- Many of the problems have interactive aspects: 2-way stream of influence
- Main message: important to explicitly **think about interaction**, and **represent it** in the frameworks we consider
- This will lead to:
  - better multiagent RL
  - better HRI / HCI
  - and even better “single-agent” machine learning (reinforcement learning, active learning, GANs, etc.)

Guidance in the (Dec-)(PO)MDP zoo?



available from my website!

Acknowledgments / Join my lab!

- Collaborators/mentors:

  (amongst others) Nikos Vlassis, Frans Groen, Sammie Katt, Christopher Amato, Shimon Whiteson, Matthijs Spaan, Stefan Witwicki, Leslie Kaelbling, Rahul Savani, Jie Zhang, Athirai Irissappane, Diederik Roijers, Yash Satsangi, José Gallego-Posada, Elise Van der Pol, Edwin D. De Jong, Roderich Groß, Daniel Claes, Hendrik Baier, Daniel Hennes, Karl Tuyls, ...

- References: see paper!

- Funding agencies:

  Netherlands Organisation for Scientific Research

  EPSRC Engineering and Physical Sciences Research Council

  European Research Council Established by the European Commission

I am still looking for one postdoc
to join the ERC INFLUENCE project!

https://www.fransoliehoek.net/wp/vacancies/