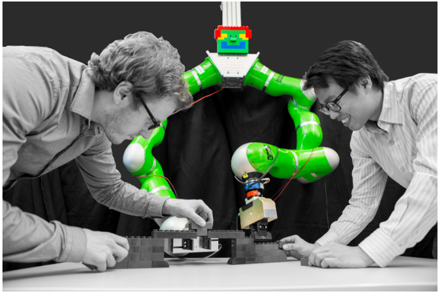
Legible Action Selection in Human-Robot Collaboration

Huaijiang Zhu **Volker Gabler** **Dirk Wollherr**

Institute of Automatic Control Engineering
Technical University of Munich

26th IEEE International Symposium on Robot and Human Interactive Communication, RO-MAN, Lisbon, Portugal, August/29/2017

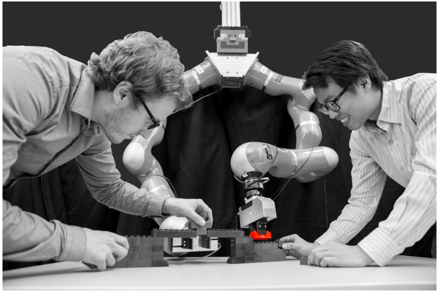
The Vision - The Adaptive Robot-Co-Worker



The adaptive cobot in an assembly process:

- understand ongoing tasks & human behavior
- select single actions accordingly
- support human co-worker

The Vision - The Adaptive Robot-Co-Worker



The adaptive cobot in an assembly process:

- understand ongoing tasks & human behavior
- select single actions accordingly
- support human co-worker

Problem formulation

legible action selection given multiple tasks

- estimate human belief in current task
- act supportive when needed

Requirements for a legible action selection framework

- increase team-efficiency
- supporting actions are taken when needed
- humans accept robot behavior

Related Work

Legible Motion Planning in HRC

Match human expectations of an excited trajectory by adjusting the motion, as

- goal-driven actions [Dragan+ 2013b; Dragan+ 2013a]
- obtained from black-box optimization [Stulp+ 2013; Stulp+ 2015]

× no task knowledge incorporated

Human Centered Probabilistic Decision Frameworks in HRC

Sequential decision-making problem as a Markov Decision Processes, e.g.

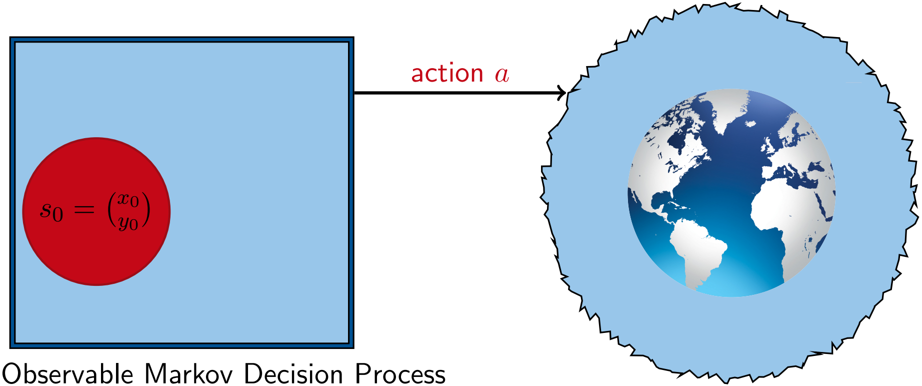
- cost-sensitive action selection based on heuristics [Hoffman+ 2007]
- incorporating human preferences as a single hidden variable [Nikolaidis+ 2015]

× no legibility involved

Contribution

Incorporate legibility in a sequential decision-making problem as a hidden goal Markov Decision Process - **HGMDP**

Mixed Observable Markov Decision Process



Environment

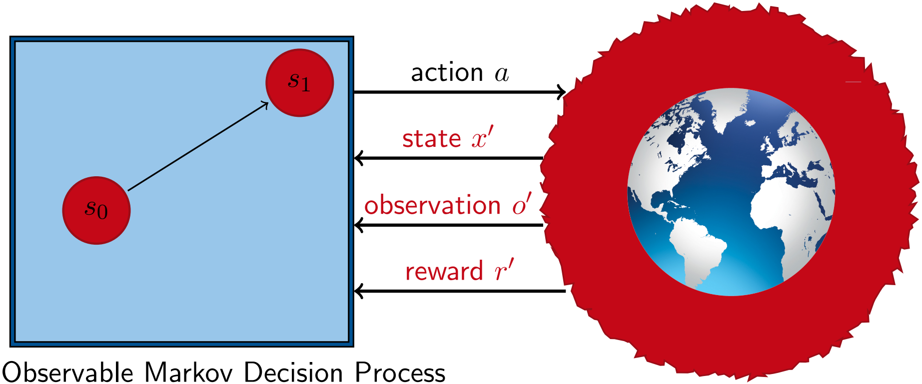
Mixed Observable Markov Decision Process

$s_i = (x_i, y_i) \in \mathcal{S}$ $\quad a \in \mathcal{A}$

fully observable state variable $x \in \mathcal{X}$

hidden state variable $y \in \mathcal{Y}$

Mixed Observable Markov Decision Process



Mixed Observable Markov Decision Process

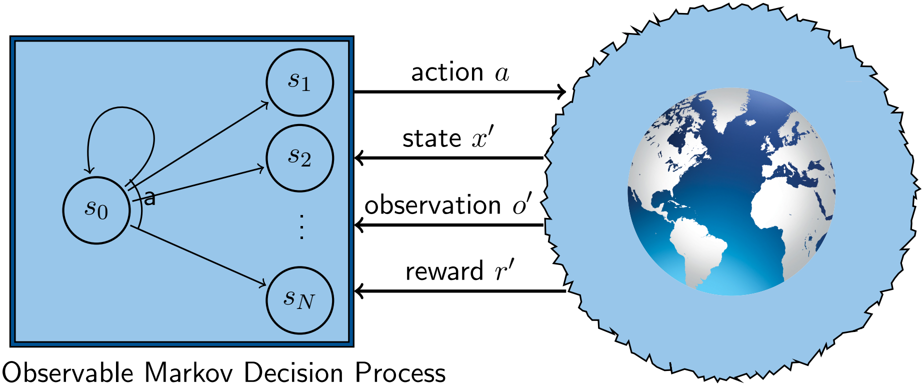
Environment

$s_i = (x_i, y_i) \in \mathcal{S}$ $\quad a \in \mathcal{A}$

fully observable state variable $x \in \mathcal{X}$

hidden state variable $y \in \mathcal{Y}$

Mixed Observable Markov Decision Process



Mixed Observable Markov Decision Process

$s_i = (x_i, y_i) \in \mathcal{S}$ $\quad$ $a \in \mathcal{A}$

fully observable state variable $x \in \mathcal{X}$

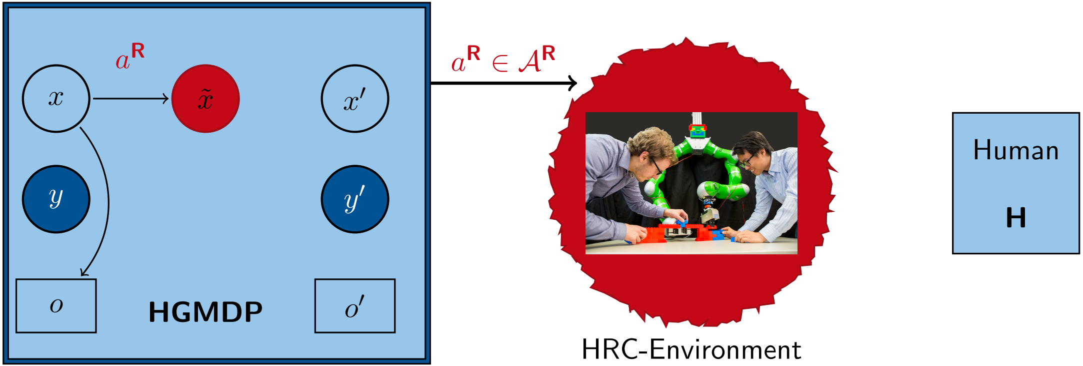
hidden state variable $y \in \mathcal{Y}$

Environment

Hidden Goal Markov Decision Process New

Given a sequential decision problem for a human **H** and a robot **R**.

- finite action sets $\mathcal{A}^{\mathbf{R}}$, $\mathcal{A}^{\mathbf{H}}$
- **task progress** as fully observable $\mathcal{X}$
- **human belief in the goal** as hidden state $\mathcal{Y}$



Hidden Goal Markov Decision Process

New

Given a sequential decision problem for a human **H** and a robot **R**.

- finite action sets $\mathcal{A}^{\mathbf{R}}$, $\mathcal{A}^{\mathbf{H}}$
- **task progress** as fully observable $\mathcal{X}$
- **human belief in the goal** as hidden state $\mathcal{Y}$

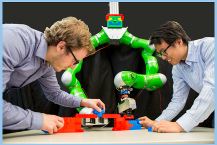
$x \xrightarrow{a^{\mathbf{R}}} \tilde{x} \xrightarrow{a^{\mathbf{H}}} x'$

y $\quad y'$

o $\quad o'$

HGMDP

$a^{\mathbf{R}} \in \mathcal{A}^{\mathbf{R}}$



HRC-Environment

$a^{\mathbf{H}} \in \mathcal{A}^{\mathbf{H}}$

$r'^{\mathbf{H}}$

Human

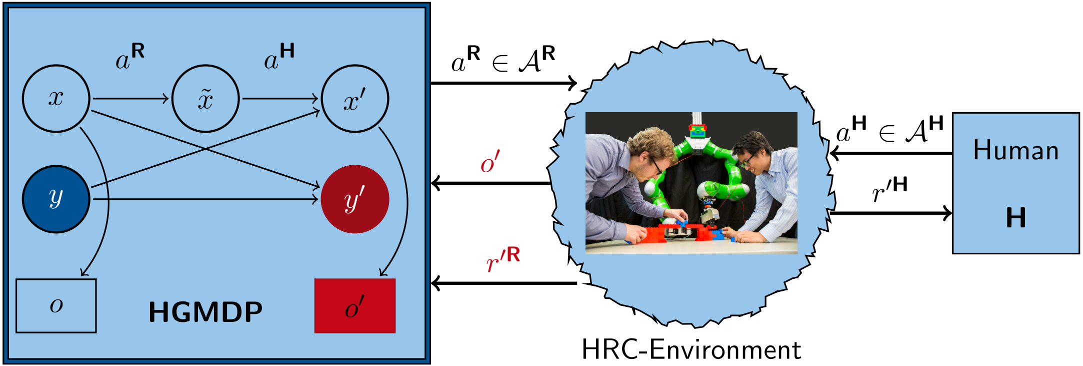
H

Hidden Goal Markov Decision Process

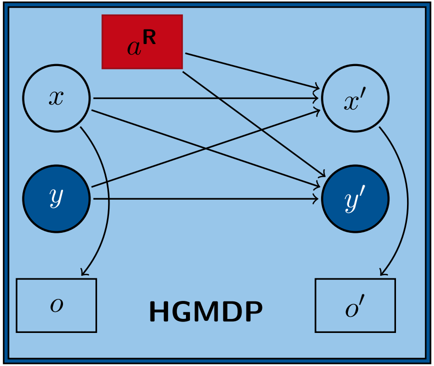
New

Given a sequential decision problem for a human **H** and a robot **R**.

- finite action sets $\mathcal{A}^{\mathbf{R}}$, $\mathcal{A}^{\mathbf{H}}$
- **task progress** as fully observable $\mathcal{X}$
- **human belief in the goal** as hidden state $\mathcal{Y}$



Human Belief Update



We assume

- the robot acts deterministically in $\mathcal{X}$
- **H** follows a stochastic policy $\pi^{\mathbf{H}}(\tilde{x}, y, a^{\mathbf{H}})$
- conditional independence $x' \perp\!\!\!\perp y'$

dynamic Bayesian Network modelling the relation of $a^{\mathbf{R}}$ to Human belief update:

$$\mathbb{P}(y'|x, y, a^{\mathsf{R}}) \propto \mathbb{P}(a^{\mathsf{R}}|x, y)\mathbb{P}(y'|x)$$

Approximatively Solving HGMDP

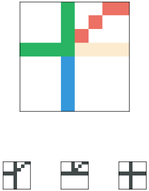
Solving **HGMDP** exactly is PSPACE-complete!

Approximative solution

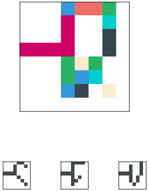
1. feature based state aggregation by mapping single actions to task sets
2. evaluate human belief at every step
   - if **H**'s belief is correct $(y = y^*)$, solve MDP for y^*
   - select strongest belief of **H** $(y_i \neq y^*)$, solve MDP M_i
3. abstracted MDP M_i for false belief of the human
   - states given as $\mathcal{S} = \{\mathcal{X}, y^*, y_i\}$
   - following legible reward model

$$R_{\mathsf{L},i}(x, y_i, a^{\mathbf{R}}) = \mathbb{P}(y^*|x, y_i, a^{\mathbf{R}}) - \lambda \mathbb{P}(y_i|x, y_i, a^{\mathbf{R}})$$

Experimental Setup



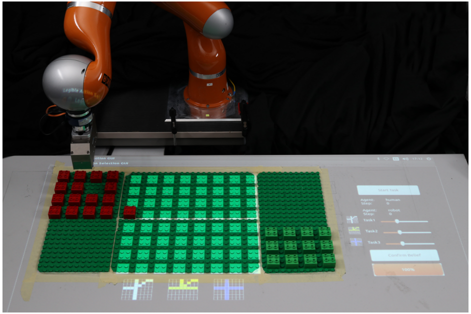
(a) Scenario 1



(b) Scenario 2

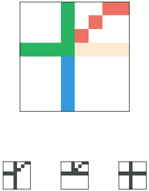
Main experimental setup

- 2 pick-and-place assembly scenarios
- 3 task goals for each scenario
- $n = 10$ participants
- 18 repetitions each

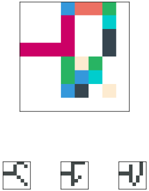


LEGO-assembly scenario with the goal being unknown to the human collaborator **H**.

Experimental Setup



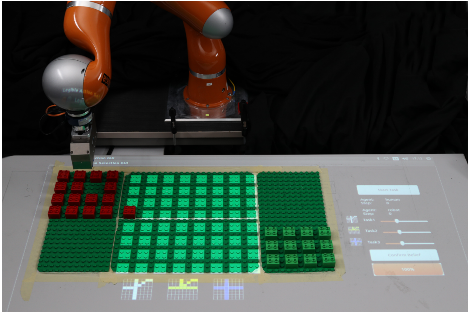
(a) Scenario 1



(b) Scenario 2

Compared policies

- **Efficient**, i.e. shortest distance (**E**)
- **HGMDP** policy (**L**)
- **HGMDP** policy with direct belief feedback (**LF**)



LEGO-assembly scenario with the goal being unknown to the human collaborator **H**.

Experimental Results - Subjective Evaluation

Confirmed Hypotheses: Compared to policy E, participants will rate the robot's actions in the HGMDP

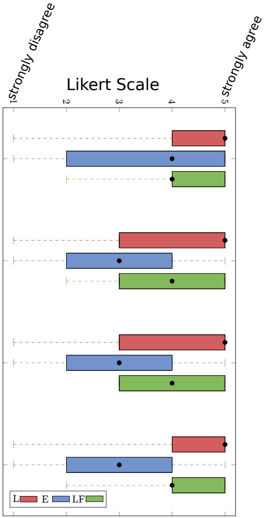
H1 ... more helpful.
H2 ... more responsive.

Q1 *The robot was acting* **efficiently**.

Q2 *The robot* **adapted** *the strategy when I was in doubt about the task.*

Q3 *The robot* **reacted** *when I made errors.*

Q4 *The choice of actions of the robot was* **helpful**.



Experimental Results - Subjective Evaluation

Confirmed Hypotheses: Compared to policy E, participants will rate the robot's actions in the HGMDP

H1 ... more helpful. **(Q1,)Q4** $\rightarrow \checkmark$

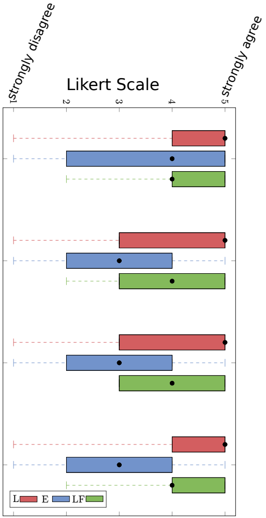
H2 ... more responsive. **Q2, Q3** $\rightarrow \checkmark$

Q1 *The robot was acting* **efficiently**.

Q2 *The robot* **adapted** *the strategy when I was in doubt about the task.*

Q3 *The robot* **reacted** *when I made errors.*

Q4 *The choice of actions of the robot was* **helpful**.



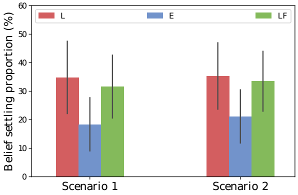
Experimental Results - Empirical Evaluation

Claimed Hypotheses: By applying HGMDP,

H3 ... **H**'s belief converges faster to the correct goal. (supportive agent)

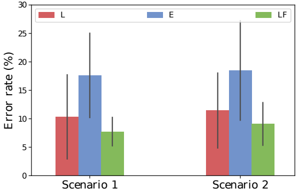
H4 .. the overall error-rate decreases. (prodctivity)

Belief settling proportion



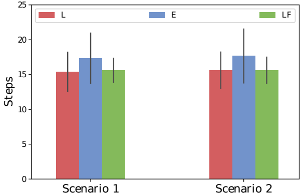
✓ confirming **H3**

Error rate



✓ confirming **H4**

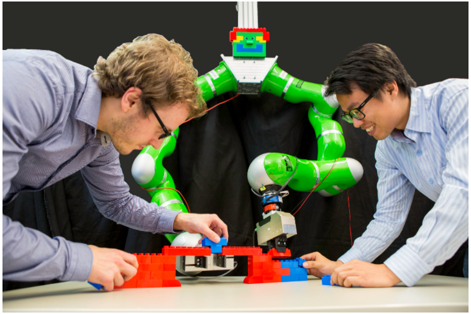
Number of task completion steps



Summary

Conclusion

- outline of **HGMDP**- a sequential and adaptive decision-making framework
- online estimation of human belief
- confirmed four hypothesis in user-study
- distinct improvements in subjective feedback
- increased empirical performance measures



References

A Dragan and S Srinivasa. **Generating Legible Motion**. In: *Robotics: Science and Systems* (2013).

A Dragan, K Lee and S Srinivasa. **Legibility and Predictability of Robot Motion**. In: *HRI*. 2013.

G Hoffman and C Breazeal. **Cost-Based Anticipatory Action Selection for Human-Robot Fluency**. In: *IEEE Trans. Robot.* 23 (2007), pp. 952–961.

S Nikolaidis, R Ramakrishnan, K Gu and J Shah. **Efficient Model Learning from Joint-Action Demonstrations for Human-Robot Collaborative Tasks**. In: *HRI*. 2015, pp. 189–196.

F Stulp, J Grizou, B Busch and M Lopes. **Facilitating Intention Prediction for Humans by Optimizing Robot Motions**. In: *IROS*. 2015.

F Stulp and O Sigaud. **Policy Improvement: Between Black-Box Optimization and Episodic Reinforcement Learning**. In: *JFPDA*. 2013.

Additional Information & Material

Experimental Results - Subjective Evaluation

Confirmed Hypotheses: Compared to policy E, participants will rate the robot's actions in the HGMDP

H1 … more helpful.

H2 … more responsive.

Wilcoxon signed-rank test results

Q1 *The robot was acting efficiently.*
Q2 *The robot adapted the strategy when I was in doubt about the task.*
Q3 *The robot reacted when I made errors.*
Q4 *The choice of actions of the robot was helpful.*

Question	Overall Comparison	L vs E	L vs LF	E vs LF
Q1	**0.0009**	**0.0013**	0.8591	**0.0004**
Q2	$< \mathbf{0.0001}$	$< \mathbf{0.0001}$	0.2789	**0.0002**
Q3	$< \mathbf{0.0001}$	$< \mathbf{0.0001}$	0.5525	$< \mathbf{0.0001}$
Q4	$< \mathbf{0.0001}$	$< \mathbf{0.0001}$	0.8552	$< \mathbf{0.0001}$

Feature Based State Aggregation

General Assembly Scenario

Given M tasks $\mathcal{T} = \{T_1, T_2 .. T_M\}$, there exists

- a set of all task components $\mathcal{C} = \bigcup_{i=1}^{|\mathcal{T}|} T_i$
- a set tasks $\mathcal{P}_i = \{T_j | c_i \in T_j\}$ each components belongs to

State Aggregation

Defining an equivalence relation over $\mathcal{C}$ and $\mathcal{P}$

$$\mathcal{R} = \left\{(c_m, c_n) | \mathcal{P}_m = \mathcal{P}_n \;\; c_m, c_n \in \mathcal{C}\right\}$$

$$\Pi = \{[c]_{\mathcal{R}} | c \in \mathcal{C}\}$$

obtains the final state aggregation by an additional error-counter $\varphi_e(x)$:

$$\Phi(x) : x \mapsto |\Pi \cap T_x| \mapsto \left[\varphi_e(x), \varphi_1(x), \varphi_2(x), \ldots, \varphi_{|\Pi|}(x)\right]^{\mathsf{T}}$$

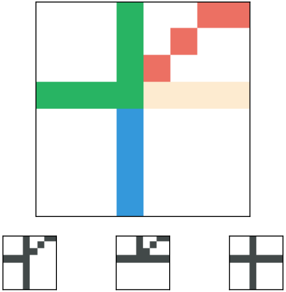
Example State Aggregation

State aggregation - Scenario 1

- $\mathcal{T} = \{T_1, T_2, T_3\}$ tasks
- $\mathcal{C} = \{c_1, c_2 .. c_{19}\}$ legal positions
- $\mathcal{P} = \{P_1, P_2, P_3, P_4\}$ task mappings shwon in green, blue, red and beige
- $|E| = \{7, 4, 4, 4\}$ maximum counter per set P_i
- $|\varphi_e(x)| \leq 4$ error counter

Define a mapping of single task components to tasks:

$$\Phi(x) : x \mapsto [\varphi_e(x), \varphi_1(x), \varphi_2(x), \varphi_3(x), \varphi_4(x)]^\mathsf{T}$$



(c) Scenario 1

Transition Probability Functions

We assume that **H** always acts greedily according to her goal expectation

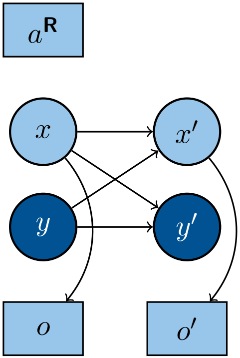
$$\pi^{\mathbf{H}}(x, y, a^{\mathbf{H}}) \propto \exp\left(\beta_2 R^{\mathbf{H}}(x, y, a^{\mathbf{H}})\right)$$

The robot acts deterministically such that

$$T_X(x, y, a^{\mathbf{R}}, x') = \mathbb{P}(x'|x, y, a^{\mathbf{R}}) = \pi^{\mathbf{H}}(\tilde{x}, y, a^{\mathbf{H}})$$

As shown in the DBN that $x' \perp\!\!\!\perp y'$ holds, such that

$$T_Y(x, y, a^{\mathbf{R}}, x', y') = \mathbb{P}(y'|x, y, a^{\mathbf{R}})$$



Transition Probability Functions

We assume that **H** always acts greedily according to her goal expectation

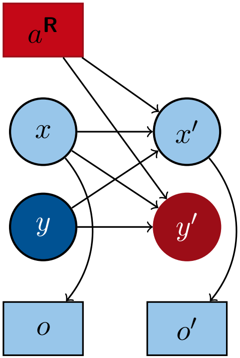
$$\pi^{\mathbf{H}}(x, y, a^{\mathbf{H}}) \propto \exp\left(\beta_2 R^{\mathbf{H}}(x, y, a^{\mathbf{H}})\right)$$

The robot acts deterministically such that

$$T_X(x, y, a^{\mathbf{R}}, x') = \mathbb{P}(x'|x, y, a^{\mathbf{R}}) = \pi^{\mathbf{H}}(\tilde{x}, y, a^{\mathbf{H}})$$

As shown in the DBN that $x' \perp\!\!\!\perp y'$ holds, such that

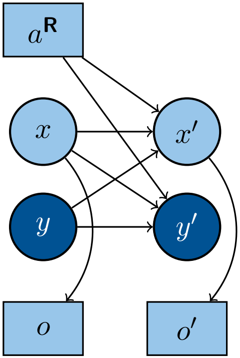
$$T_Y(x, y, a^{\mathbf{R}}, x', y') = \mathbb{P}(y'|x, y, a^{\mathbf{R}})$$



Goal Inference

The goal inference in **HGMDP**is obtained from the distribution of y at each transition according to

$$b'(y') \propto \mathbb{P}(o|x', y', a^{\mathbf{R}}) \sum_{y} T_{XY}(x, y, a^{\mathbf{R}}, x', y')b(y)$$



The observation is modeled deterministically

$$\mathbb{P}(o|x', y', a^{\mathbf{R}}) = \begin{cases} 1, & \text{if } o = a^{\mathbf{H}} \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

The state transition function is given by

$$T_{XY}(x, y, a^{\mathbf{R}}, x', y') = \pi^{\mathbf{H}}(\tilde{x}, y, a^{\mathbf{H}})\mathbb{P}(y'|x, y, a^{\mathbf{R}})$$

Incorporating Legibility in Reward Model

Human inference probability for **R**'s actions:

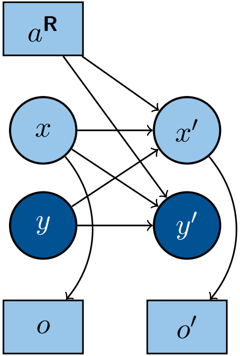
$$\mathbb{P}(a^{\mathbf{R}}|x,y) \propto \exp\left(\beta_1 R^{\mathbf{R}}(x,y,a^{\mathbf{R}})\right)$$

Given the actual goal y^*, the legible reward model results in:

$$R_{\mathsf{L}}(x,y,a^{\mathbf{R}}) = \mathbb{P}(y^*|x,y,a^{\mathbf{R}}) - \lambda \sum_{y' \in \mathcal{Y} \setminus \{y^*\}} \mathbb{P}(y'|x,y,a^{\mathbf{R}})$$

In return, this results in the following update rule for **H**'s belief:

$$\mathbb{P}(y'|x,y,a^{\mathbf{R}}) \propto \begin{cases} p_{\mathsf{c}}\mathbb{P}(y'|x,a^{\mathbf{R}}), & \text{if } y = y' \\ \frac{1-p_{\mathsf{c}}}{|\mathcal{Y}|-1}\mathbb{P}(y'|x,a^{\mathbf{R}}), & \text{otherwise} \end{cases}$$



Approximatively Solving HGMDP

Approximate **HGMDP**based on the current human belief y by a fully observable MDP M_i with $\mathcal{S} = \{\mathcal{X}, y^*, y_i\}$ and

$$T_i = \mathbb{P}(x', y' | x, y, a^{\mathbf{R}}) = \begin{cases} \pi^{\mathbf{H}}(\tilde{x}, y_i, a^{\mathbf{H}}), & \text{if } y' = y_i \\ 0, & \text{if } y' \neq y_i \end{cases}$$

as well as

$$R_{\mathsf{L},i}(x, y_i, a^{\mathbf{R}}) = \mathbb{P}(y^* | x, y_i, a^{\mathbf{R}}) - \lambda \mathbb{P}(y_i | x, y_i, a^{\mathbf{R}})$$

Obtain HGMDPPolicy

Resulting in the overall policy to solve **HGMDP**

$$\pi_{\mathsf{L}}(x, b(y), a^{\mathbf{R}}) = \begin{cases} \pi^* \left(M_i(\mathcal{S} := \mathcal{X}) \right) & \text{if } \arg\max b(y) = y^* \\ \hat{\pi}_{\mathsf{L}}(x, \underset{y \in \mathcal{Y} \setminus \{y^*\}}{\arg\max} b(y), a^{\mathbf{R}}) & \text{else} \end{cases}$$

General Approach - Hidden Goal Markov Decision Process

Problem Definition

Given a sequential decision problem for a human **H** and a robot **R**.

- Finite action sets $\mathcal{A}^{\mathbf{R}}$, $\mathcal{A}^{\mathbf{H}}$.
- Reward functions given as $R^{\mathbf{R}}, R^{\mathbf{H}}$,.
- Obtain $a^R = \text{argmax} \sum_i^N R$.

Hidden Goal Markov Decision Process

Given as $M = (\mathcal{X}, \mathcal{Y}, I_Y, \mathcal{A}^{\mathbf{R}}, \mathcal{A}^{\mathbf{H}}, \mathcal{O}, T_X, T_Y, Z, R^{\mathbf{R}}, R^{\mathbf{H}}, R_{\mathsf{L}}, \gamma, y^*)$.

- $\mathcal{X}$: fully observable task state.
- $\mathcal{Y}$: hidden variable representing human goal expectation (y^* as the actual goal).
- $\mathcal{O}$: set of observations, given as the actual human action.
- $T_X = \mathbb{P}(x'|x, y, a^{\mathbf{R}})$ and $T_Y = \mathbb{P}(y'|x, y, a^{\mathbf{R}}, x')$: transition probability functions.
- Z: probability distribution to observe o.
- γ: discount factor $\in [0, 1]$.

Hypotheses and Measurements

Claimed Hypotheses

Compared to the *efficient* policy, Participants will rate the robot's actions in the **HGMDP**...

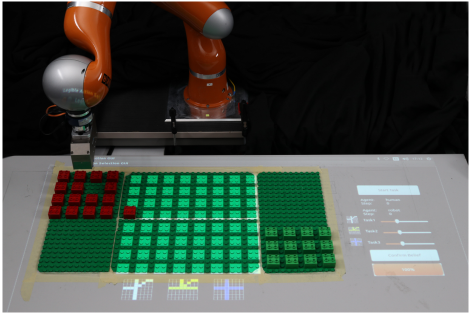
- **H1** more helpful.
- **H2** more responsive.

By applying the **HGMDP**, ...

- **H3** **H**'s belief converges faster to the correct goal.
- **H4** the overall error-rate decreases.

Experimental Measurements

- subjective questionnaire (**H1, H2**)
- belief settling proportion, w.r.t. to steps (**H3**)
- error-rate over all runs (**H4**)



LEGO-assembly scenario with the goal being unknown to the human collaborator **H**.

Outsourced