Accepted Publication to the 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
October, 1-5, 2018, Madrid, Spain
DOI: 10.1109/IROS.2018.8594076

Online inference of human belief for cooperative robots

Moritz C. Buehler[1,2] and Thomas H. Weisswange[2]

Abstract— **For human-robot cooperation, inferring a human's cognitive state is very important for an efficient and natural interaction. Similar to human-human cooperation, understanding what the partner plans and knowing, if he is situation aware, is necessary to prevent collisions, offer support at the right time, correct mistakes before they happen or choose the best actions for oneself as early as possible.**

We propose a model-based belief filter to extract relevant aspects of a human's mental state online during cooperation. It performs inference based on human actions and its own task knowledge, modeling cognitive processes like perception and action selection. In contrast to most prior work, we explicitly estimate the human belief instead of inferring only a single mode or intention. Since this is a double inference process, we focus on representing the human estimates of environmental state and task as well as corresponding uncertainties.

We designed a human-robot cooperation experiment that allowed for a variety of cognitive states of both agents and collected data to test and evaluate the proposed belief filter. The results are promising, as our system can be used to provide reasonable predictions of the human action and insights into his situation awareness. At the same time it is inferring interpretable information about the underlying cognitive states – A belief about the human's belief about the environment.

I. INTRODUCTION

Robots and other (partially) autonomous technical systems are more and more present in our daily and working life, examples include vacuum cleaning robots, manufacturing robots, intelligent application software or advanced driver assistance systems. Particularly classical manufacturing robots often have very narrow use-cases, because they are limited to work in dedicated areas to protect people from harm (separation of space). This often demotes them to mere tools, that is humans are delegating defined functions to those technical systems (functional separation), e.g. a vacuum cleaning robot is clearly supposed to clean a room. Those approaches become more and more inappropriate due to the high complexity and interdependence of modern technical systems. For a next generation, it is important to investigate interaction between humans and autonomous technical systems to create interfaces that are comprehensible and efficient, enabling a natural cooperation. Application scenarios for human-robot cooperation include cooperative manufacturing, shared autonomy teleoperation, path planning in crowded environments or semi-autonomous traffic. For the cooperation of humans and machines, complex aspects known from human-human interaction become important, including situation awareness, predictability, trust, intention, desires or affect. Using these aspects requires some basic understanding of the human cognition and its representation of the environment.

Human situation awareness (SA) in particular covers many important parts of the human cognitive state. SA is a topic of interest especially in aviation research [1] since many years. It proved to be an efficient concept to evaluate and enhance the performance of a human operator in his interaction with a machine. Situation awareness goes beyond the perception of relevant pieces of information, like flight variables (e.g. altitude) or the environmental states (e.g. location of other aircrafts). The human has to relate different information fragments (low altitude is necessary while landing, but dangerous during flight) and anticipate the consequences. Along these lines, we propose that a robot also needs to be situation aware to be able to solve more complex tasks. For a collaborative setting, SA will also include awareness of the interaction partner and his internal processes, for example if a human partner is situation aware himself. With this knowledge a robot can warn or direct the human's attention, provide additional information or adapt its own strategy to prevent failures or improve overall task performance. For this reason, we want to model the relevant cognitive processes of a human partner involved in solving a collaborative task. But since we can not directly access them, we will infer the hidden information indirectly through reasoning over the human's actions and active sensing strategies. Indeed, humans are really good in inferring the mental states of their human partners [2].

One interesting requirement for SA is the understanding of the task goals and valuations. In machine learning research one concept of inferring these is Inverse Reinforcement Learning (IRL) [3]. Reinforcement learning tackles the problem of an agent selecting actions to optimize accumulated future reward in a dynamic environment. IRL in contrast infers the reward function from an observed optimal action sequence. It was previously proposed to specify the goal for a robot in collaborative settings, implicitly taking into account some of the cognitive state variables of a human partner [4]. In [5] IRL was used in an automotive context to predict a human's most likely future action and to incorporate it into the planning process of a robotic agent. IRL focuses on inferring the reward function based on human actions, however, to appropriately handle situation awareness, additional aspects have to be considered.

Among human actions, gaze is an important information source to estimate which aspects of the environment the human might be aware of. Gaze as active observation is proposed to obtain the object of human attention awareness for a collaborative setting in [6]. Similarly, in an automotive

[1]with Control Methods & Robotics Lab, TU Darmstadt, Germany
[2]with Honda Research Institute Europe GmbH

context, [7], [8], estimation of a human's focus of attention is extracted from gaze information to estimate his situation awareness. [9] investigate gaze as communication, that is used to coordinate actions by looking at the other. In the remainder of this paper, we will call active observations like this gathering actions, because they can be used for our inferences in a way similar to task-progressing actions. This can, for example, also be seen in [10], where the authors use IRL on gaze behavior to infer the internal valuation of awareness of certain information for a given task.

The combination of both task actions and gathering actions, should provide a more complete view over the human cognitive states. As task structure and perception are not free of noise and might only be partially observable, the cognitive human state will include a belief, i.e. a probabilistic representation of his environment. Previous belief inference concepts are proposed in [11] and [12] interpreting an agent's action and information gain. [11] use belief recursion to an arbitrary depth (i.e. he believes that I believe...) for including the cognitive state of an interaction partner into the decision making. They apply it to a game-like multi-agent scenario with prespecified policies for the (artificial) agents and were able to solve cooperative problems. In contrast, [12] observe a human moving in a grid world while approaching a desired food truck. They want to explain the observed behavior retrospectively by inferring the human's desires and beliefs based on his actions and a simple perception model. The human belief is sampled uniformly and inferred after task completion. The results are compared to simpler (heuristic) human models and to human observer assessments.

In this paper, we develop a general concept for inferring belief and situation awareness of a human cooperation partner from information gathering and task actions. We use an application scenario, inspired by human-robot cooperative manufacturing, with a shared task and goal and the potential for supportive warning actions and predictive adaptation of the robot strategy. This direct interaction target requires a formulation for online estimation of a human's belief, in contrast to [12]. Our approach introduces a parametrized approximation of the human belief that concentrates on significant aspects of the cognitive state. Through this, inferring the complete human belief over the environment reduces to the inference of the parameters which form the basis for estimating his situation awareness.

The remaining paper is structured as follows: First, we will describe our general concept for filtering the human belief based on his task and gathering actions. Therefor, we introduce a parametrization of the human belief to support online inference. We describe our cooperative human-robot experiment and specify the inference processes for this scenario. The collected data is finally used to test and evaluate our inference process.

II. Model-based online belief filter

We present our approach for inferring the human belief. Our goal is the enhancement of human-robot cooperation, where the robot should adapt its behavior based on the human's situation awareness. This is done through inference of the human belief regarding relevant parts of the common task during execution. Therefore, we use human actions as well as his information gathering and respect its consequences for the human state of mind to generate a complete view.

We formalize our problem in the Partially Observable Markov Decision Process (POMDP) framework [13]. The environmental state s is changed to the next state s' through actions a by the agents, according to the dynamics, called the transition function, $T(s'|s,a)$. A reward signal provides possibly delayed feedback on the quality of the behavior of the agents. It is assumed that they select their actions trying to maximize the overall reward. The system state s is not directly accessible by the agents, but has to be inferred from the perceived observations $o(a,s)$. Based on this information, an agent can construct a belief over the state, $b = p(s)$, which means integrating action and observation histories into a probability distribution of the current state.

We assume that the human creates a belief b_H based on which he selects his actions. The belief of the human is only partially observable to the robot, it can not look into the human brain. Instead it only knows the human's action and information gathering activities, which provide information of the human observations. Inferring the human belief results in a probability distribution over a probability distribution, which quickly gets to complex for interesting real-world scenarios. Therefore, we approximate the human belief through a sparse parametrized distribution and infer its parameters. With an appropriate parameter selection, we will show that we can still recover a representation containing the task relevant aspects.

A. Representation

The belief $b_H = p(s,T)$ describes the human's internal state as distribution over the state and transition function (the transition belief could be subsumed in the state belief by defining the base MDP in a different way). Constructing a probability distribution $p(b_H)$ is unfeasible for real world scenarios.

We parametrize an approximate belief distribution $b_H \approx q(s,T|\theta)$ with parameters θ that focuses on regions, where the belief is high. The most probable belief $\mu = \operatorname{argmax}(b_H)$ is represented together with its probability $\alpha = b_H(\mu)$ as parameters $\theta = (\mu, \alpha)$. In a continuous space, one could alternatively approximate the human belief through a normal distribution with parameters mean and precision. Representing only one state in the belief can be limiting, especially for multimodal cases. This can be easily overcome by separately modeling the k most probable states.

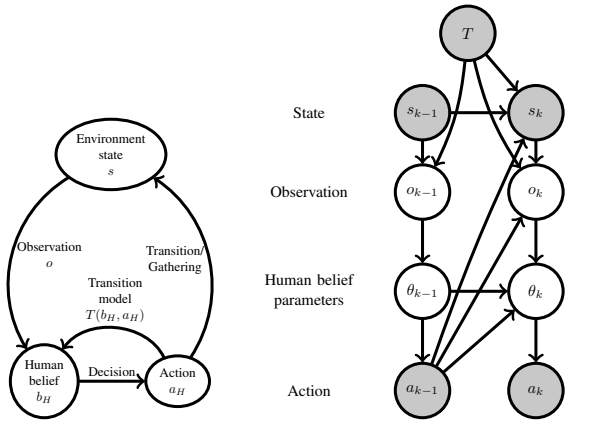
In the following, we consider a general parametrization θ for the human belief and develop our filter equations to infer the distribution $p(\theta)$ over the parameters.

B. Filter structure

We model the human as rational agent acting in a POMDP. The cognitive human processes that we respect are shown in

Fig. 1a. Based on his belief, the human will decide on the next action, that can be information gathering (e.g. gaze shift) leaving the state unchanged or an action that leads to a state transition of the environment. Information gathering will result in an observation, providing information the human uses to update his belief. For task actions, the human will respect the expected transition of the environmental state, updating his state belief according to his transition belief.

In Fig. 1b the temporal and causal relations of the filter variables are shown in the form of a Dynamic Bayesian Network. It includes the relations from Fig. 1a to update its representation in the same way, we suppose the human does, according to action and information gathering. Since the human action decision is supposed to be based on his belief, we can use the actual decision as noisy corrective information to reduce deviation between belief estimate and the true belief.



(a) Structure of modeled mental processes

(b) Temporal filter structure

Fig. 1: Model structure

C. Expected human observations

Human observations are not directly accessible, but we can attribute them based on human information gathering. For example, if someone looks at an object (directs his gaze towards it), one assumes that he observed that object and integrated its presence in his internal environment representation. Therefore, based on (information gathering) action by the human, a_H, we conclude on his observation, $p(o|a_H)$ (perception model). An observation will lead to an update of human belief, $p(s, T|o) \sim p(o|s, T)p(s, T)$. We therefore have to update the parameters θ

$$p(\theta|a_H) \sim \sum_{s,T,o} p(o|s, T)q(s, T|\theta)p(\theta)p(o|a_H), \quad (1)$$

where the observation probabilities $p(o|s, T)$ depend on the concrete, application specific perception model.

D. Actions

For human actions, we have to do two further effects, feedback based on human action selection and the state transition. The first is based on the assumption, that the human decides approximately rationally and we calculate the action probabilities according to a softmax model as

$$p(a_H|\theta) \sim \exp(\tau Q(a, s|\theta)), \quad (2)$$

where $Q(a, s|\theta)$ is the action value function depending on the belief parameters θ. τ is a temperature hyperparameter, characterizing the human degree of rationality.

The corrective update can be computed as

$$p(\theta|a_H) \sim p(a_H|\theta)p(\theta). \quad (3)$$

The action value function Q could be obtained from a POMDP solver, e.g. [14]. Using an online method like this, the distribution over θ might be sampled to reduce the required processing load.

For task progressing actions, the human expects a transition of the system state s to a new state s'. He will update his belief and the new distribution of the parameters after state transition $p(\theta')$ becomes:

$$p(\theta'|a_H) = \sum_{s,T,\theta,s'} p(\theta'|s', \theta) \underbrace{p(s'|s, T, a_H)}_{T(s', s, a_H)} q(s, T|\theta)p(\theta|a_H), \quad (4)$$

where $p(\theta'|s', \theta) \sim p(s'|\theta')p(\theta'|\theta)$ is the approximation of $b_H(s')$ through our parametrization.

E. Belief evaluation and situation awareness estimation

The inferred human belief can be used to enhance cooperation in various ways. In Eq. (2) we already compute a probabilistic action prediction. But we want not only to predict but to understand the human behavior to improve the cooperation on another level, e.g. providing him necessary information. Therefore we use the belief to evaluate the situation awareness of the human. We call a human situation aware, if the optimal action cost based on his belief is not significantly worse than the best action cost given perfect knowledge. If this is the case, the human belief contains all relevant information for optimal action selection. To evaluate it, the robot itself must be aware of the situation, since we need this knowledge to estimate the human belief.

The approach for filtering the human belief was introduced on a general level. We will now move to a concrete situation to apply and test it. In the following section, we present our human-robot cooperation experiment and the application of the human belief inference.

III. Experiments

We designed a cooperative experiment with the objective of generating useful data to evaluate our approach. This has several requirements: The final target is an application to a real-world task, e.g. a cooperative manufacturing task, which we want to represent in an abstracted way. We need interaction in the form of interdependence between robot and human. Further, we require a planning process by the participants to include temporal aspects and raise the complexity, which we need to achieve variability in the human performance. It should also be possible to evaluate the performance of the agents. Inspired by cooperative manufacturing,

we selected a sequential task, where a human and a robot have to press buttons in a specific order. Each button press can represent a processing step of the manufacturing task. The abstraction of using buttons has the advantage that we avoid domain specific difficulties for the robot, like complex controllers to grasp tools. Further, the workspace provides the opportunity to realize other application scenarios.

A. Setup

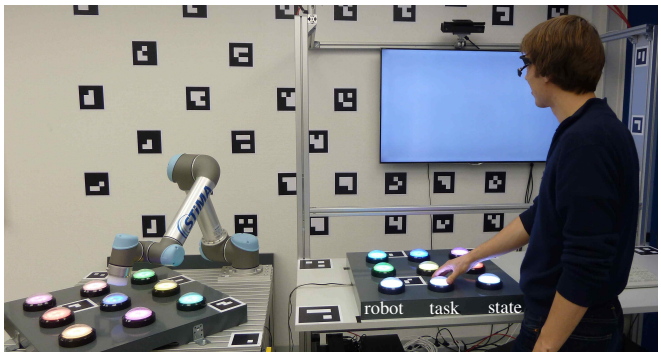


Fig. 2: Experimental setup for human robot cooperation experiment, gathering buttons on human board in white.

Our setup consists of a cooperative workspace for human and robot, as shown in figure 2. The robot is a one arm industrial UR5 robot, designed for operating with humans in a common workspace. Both agents, human and robot have a board on the table with 9 buttons in a 3 times 3 grid in front of them. The buttons can be set to different colors which are used to distinguish different actions. A screen on the wall is used to display task relevant information, e.g. task sequence or state. Human pose and gaze can be measured, but are not used so far. Instead, gaze as active observation is abstracted by discrete gathering actions on a subset of the buttons, as explained below. Currently, we do not use direct physical intersection, the agents stay in separated working areas. Interdependence is achieved through the task and the need for respecting the other's actions in planning.

B. Task design

The task consists of pressing colored buttons in a given order. It is automatically generated following specified rules to respect our demands. An example structure is shown in Fig. 3. Distinct actions are represented by button colors. The shape serves to distinguish the actor, squares for robot actions and circles for human actions. The nodes between the actions are used as system states and pressing the button marked on an outgoing edge leads to a transition to the linked node. The task starts at the most left node, ends at the goal node on the right, and is shared between both agents. Different paths are possible to achieve the goal, since there are states with multiple outgoing edges, leading to different branches of the graph. Both actors can be in situations, where they select a branch by taking an action. Starting on the left-most node of Fig. 3, the robot can decide between upper path (pressing yellow) and lower path (pressing red). The human has to track the robot's decision to determine subsequent actions. If the robot presses the yellow button, the human has to press the purple one subsequently. We introduced the branching to force planning over purely reactive behavior and to achieve the required interaction and complexity. Taking an action in a state where it is not a specified transition is a mistake and leads to a negative reward, while the state remains unchanged. To achieve the desired task complexity, we generate tasks with a size of 8 states.

The explicit task is generated with randomly varying branching and merging points and button attributions. The mapping of colors to the buttons is changed between tasks to avoid habituation.

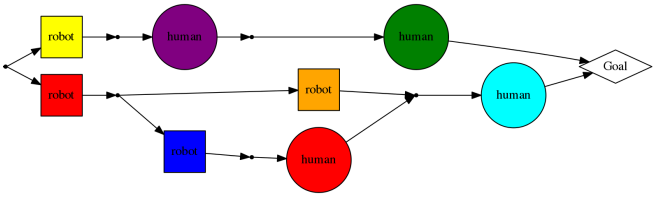


Fig. 3: Cooperative task plan example

In addition to the colored buttons, we take human information gathering into account. To focus on the actual inference process, we introduce discrete gathering actions as temporary simplification. Therefore we use three buttons, illuminated in white (Fig. 2). By holding them, the human obtains a certain piece of information on the screen. "Task gathering" displays the whole task as in Fig. 3. "State gathering" shows the actual state of the environment as node in the graph including last and next action(s). "Robot gathering" visualizes the robot button approach to distinguish between different robot actions. Due to this explicit robot gathering action, we want to avoid direct visual information and the robot is simulated during the main part of the experiment.

We introduce a gathering cost through a time delay for displaying the information. The human should always track the environmental state, therefore we use the highest delay of one second for that. Gathering the task is delayed by a half second, the task is already shown at the start of each run but difficult to remember. Since the robot moves and acts, the human needs to dynamically adapt to what it does, so we assign no further delay for robot gathering.

In our current experiments, the robot randomly presses one of the next buttons from the task. The intention behind this is that the human has to actively track the robot's actual action. When the robot chooses a path, the human has to be aware of the actual decision, which he can achieve through robot gathering. This robot selection strategy is communicated to the participant before an experiment. The actual as well as simulated button press produces a sound, from which the human can infer the end of the robot action execution and the related state transition. After the sound, he can continue with the next action. The robot needs about 2 to 6 seconds for pressing a button, depending on the button location. This

is normally sufficient for the human to gather information towards the upcoming parts. Varying the robot's speed would be a way to adapt the human stress level.

The reward function contains two parts, the number of false button presses and the execution time. The human has to balance between performing fast with the risk of mistakes and performing safely with the cost of time through information gathering. He is told about both aspects without the explicit weighting to avoid further confusion. At the end of the task, we provide feedback in form of a score. From these it should be possible to get at least an intuition on the weights, however they may have their individual weighting, which we represent as static hyperparameter w.

At the beginning of each task run, called episode, the corresponding graph is displayed on the screen for three seconds. This time is designed to be insufficient for the human to memorize every aspect so the human has to use the task gathering button during execution. The experimental procedure of several tasks starts with a habituation phase, where the human has the chance to learn the overall principle, how to progress, and to get used to the gathering buttons. After this habituation phase of 5 episodes, we recorded the execution of 20 tasks.

We performed experiments with 9 participants. 21 times, that is 5% of the task actions, a false button was pressed by the human. The execution time as second measure varied a lot (STD 1.8 s, Mean 2.2 s) between different episodes and participants. From the frequent usage of the gathering actions, we conclude that memory is a significant limitation in our task. For future work it will be interesting to deal with other (e.g physical) limitations of the human.

In average, the participants took 2.4 task actions and 3 gatherings per episode. We observed differing strategies by the participants regarding task gathering versus state gathering. Indeed, there is information overlap, because gathering the state also provides the information over the immediate next action(s). However, the strategy of relying mainly on state gathering is not globally optimal, because it neither allows to choose the best path nor to prepare a sequence of consecutive human actions. Instead, it seems to be a local optimum in the human learning process. This strategy was observed mainly for participants 2, 3 and 4 and partly for 5.

C. Filter specification

We now apply the general belief inference concept from section II. As environmental state s we refer to the current node number of the task graph. The transition function T is specified by the task.

1) Representation: We are interested in three aspects of the human belief b_H, namely the state, the transition function and the belief over the next robot action. We parametrize and update these independently (but with the same observations).

For the state belief we use one parameter describing the most probable state $\mu_s = \text{argmax}_s(b_H(s))$ and its probability $\alpha_s = b_H(\mu_s)$. As prior we use the start state, $p(\mu_s = 0) = 1$.

For the transition dynamics we represent the structure as adjacency matrix T_H as part of the true transition function T, together with several probability parameters α_T. We further divide the human task belief in two parts, the overall structure respectively connectivity and the actual buttons connecting the nodes. Since we designed the task such that the human is unable to memorize it completely, we model two sequential steps by the human. The first is the selection of one single path through the graph whereupon he tries to capture the corresponding buttons. This structure consisting of one possible paths is represented as connectivity matrix C_T together with a probability $\alpha_{T,C}$. Further, the human needs to know the required actions to proceed on the selected path. To respect the human memory limitation, we introduce a hyperparameter n_{mem} for the memory size, characterizing the number of buttons, the human can remember. Based on short term memory research, we select four as typical number [15]. Within this assumption, we have n_{mem} many parameters $\mu_{T,i}$ of most probable buttons along the path and the corresponding probabilities $\alpha_{T,i}$ ($i = 1 \dots n_{mem}$). The connectivity matrix C_T and the colors $\mu_{T,i}$ are combined in the adjacency matrix T_H. Together with the probabilities α_T, it forms the task parametrization, for which we use a uniform prior.

In situations, where the robot has multiple action options, the human needs to know, what action the robot performs. Therefore, we represent the belief for the current robot action $b_H(a_R)$. Because we informed the human about the random action selection by the robot, we assume, that he has a uniform prior over robot actions. When the human gathers information on the robots movement, his belief will update to the true action, $b(a_R = a_{R,true}) = 1$. Since the robot action belief is clear and only important for the current node, we do not use a probabilistic representation for it. The used belief representation combines all three aspects, the parameters are $\theta = \begin{pmatrix}\mu_s & \alpha_s & T_H & \alpha_T & \mu_R\end{pmatrix}$.

2) Observations: Each gathering action is assumed to lead to an observation and a corresponding belief update by the human. Therefore, we have to specify the observation probabilities $p(o|s,T)$ from Eq. (1). With the discrete information buttons in our experiment, the observations are clearly related to these and assumed to provide reliable information.

Accordingly, task gathering sets the human task probabilities α_T to one and the estimates to the true values. Gathering the state will display the actual state together with the immediate next action(s). Thus state μ_s and and the next button color $\mu_{T,1}$ are updated to the true values and the corresponding certainties α_s and $\alpha_{T,0}$ becomes one. Lastly, information of the robot's movement will lead to certain human knowledge, regarding the robot action.

For another observation type, the sound triggered by a robot action, we use Eq. (1), where we have to specify the observation probability $p(o_{aR}|s,T)$. It results from the probability for any robot action in the current state, $p(o_{aR}|s,T) = \sum_{a_R} p(a_R|s,T)$. A robot action leads to a (possibly uncertain) state transition, changing the human belief over state. The parametrization updates to $p(\theta') =$

$\sum_{a_R} p(\theta'|a_R)p(a_R|\mu_R)$, where $p(\theta'|a_R)$ results from Eq. (4), replacing human with robot action.

3) Human action: For the update of $p(\theta)$ according to human actions, we need to compute the action value function $Q(a, \theta)$. For our relatively simple task, we can declare an action value function by hand and do not need to use a POMDP solver. The expected value of the action $\mu_{T,1}$ depends on the probability for state and task, $Q(\mu_{T,1}) = \alpha_{T,1}\alpha_s$, describing the risk of failure. Gathering actions augment the human corresponding probabilities α and increase the following success probability. However, if the task progress depends on the human action, gathering will delay the task progress and therefore result in a gathering cost c. The resulting action values are $Q(state) = 1 - c_{state}$, $Q(task) = \alpha_s(1 + \gamma\alpha_{T,2:}) - c_{task}$, $Q(robot) = \alpha_{T,1}\alpha_s + c_{state}$ with the corresponding gathering costs for states μ_s, where the human can progress.

Therefore we used -1 as reward for a false action and the gathering costs from time delay, multiplied by the false/time ratio, $w = 3/s$. Choosing a longer path augments the time needed and results in an additional cost for those actions. Gathering the task reduces the human's uncertainty not only for the next action but also for future actions. The benefit of future uncertainty reduction may be discounted by some factor γ, modeled as hyper parameter. We set it to $\gamma = 0.5$ and use a rationality parameter $\tau = 3$.

D. Results

The principle of belief inference is demonstrated in Fig. 4 for a part of a recorded run. The relevant part of the task structure is shown in a). The human presses the purple button in b), transitioning to the next task state. In c), the result of the inference process after this action is visualized. The inferred state belief $p(\mu_s)$ is represented through the width of the pentagon on the node. The probability for the human state estimate μ_s is concentrated on the first state. One part of the task belief is shown by the empty symbols, representing buttons probably unknown for the human. In this state, the robot can press dark blue or rose, leading to different branches. Since the robot chooses the action randomly, the human should gather the robot's movement to track the state transition. Our filter predicts this action as most likely by the human. In this recording however, the human does not follow our prediction, while the robot presses dark blue in d). In the updated belief estimate e), multiple states regarding human belief μ_s are possible. Additionally a low human state certainty α_s (not illustrated) is estimated. Due to that, the filter predicts state gathering by the human, which he actually does in the next step f). Consequently, the new state belief of the human is concentrated on the true state and the next action options should be known by the human g). Latest after the next human action (purple or light blue) we expect him to gather for task, since we estimate the human to be uncertain about the subsequent colors (represented as white circles).

When introducing interaction, e.g. providing specific missing information to the human, an online calculation is necessary. Each of our belief update steps is computed sufficiently fast, it takes less than 5 ms on a standard desktop PC with 3.2 GHz single core computation.

For a quantitative evaluation, we compute the hit rate of the action prediction. According to Eq. (2), we calculate action probabilities and compare the most probable with the actual human action. Since there are cases with multiple appropriate action (e.g. branching over comparable paths), we use as second measure the likelihood, the probability our filter attributes to the actual human action.

The hit rate of the prediction is shown in Fig. 5a. The averaged rate is 56%, but varies between the participants. We compute a statistical baseline over all participants, that always predicts the most frequent action (task gathering, 26%). For comparison, we let an expert (first author) predict the human actions using the same information (action history and task knowledge) as our system (59%). The low success rate for participants 2 and 3 might result from a suboptimal strategy that they use (which violates the rationality assumption). Focusing only on predicting human task actions, the success rate is significantly higher at 82%. It is most difficult to predict the gathering actions for task and state correctly. Looking at the experts performance, our task seems to have some general problems regarding predictability.

The average likelihoods, assigned to the performed human actions, are shown in Fig. 5b, the mean inference result is 51%. Again, we calculate a statistical baseline (relative action frequencies). The expert was restricted to spread his prediction uniformly over (multiple) actions. We observe several cases, where two actions or three gathering actions of the human are reasonable and the action prediction distributed.

Finally, we analyze the influence of the hyperparameters. Once selected, they are held constant for the evaluation of all participants, to achieve better comparisons. The memory parameter is varied between 2 and 8 with little effects regarding the hit rate, shifting towards prediction of task progressing actions. In fact, our memory model is simple and e.g. the existing dependence on the task complexity (number of branches) is not respected. The rationality parameter has almost no influence on the hit rate. The action distribution becomes sharper which leads to more peaked probability predictions. Most influence results from the weighting parameter of the human reward function, the ratio error to time (in seconds). An increased time cost leads to better prediction results. Over the ratios 0.1 to 100 the prediction rate rises from 52 to 59 %.

The belief inference produces plausible qualitative behavior and the prediction results are promising. Further analysis is necessary to ascertain the performance of our approach and the specific parametrization.

IV. CONCLUSIONS

We presented an approach for inferring human belief to estimate his situation awareness, relying on human information gathering and human actions. We proposed to approximate the human belief distribution with a parametrized distribution, allowing the inference of these parameters without need for double inference.

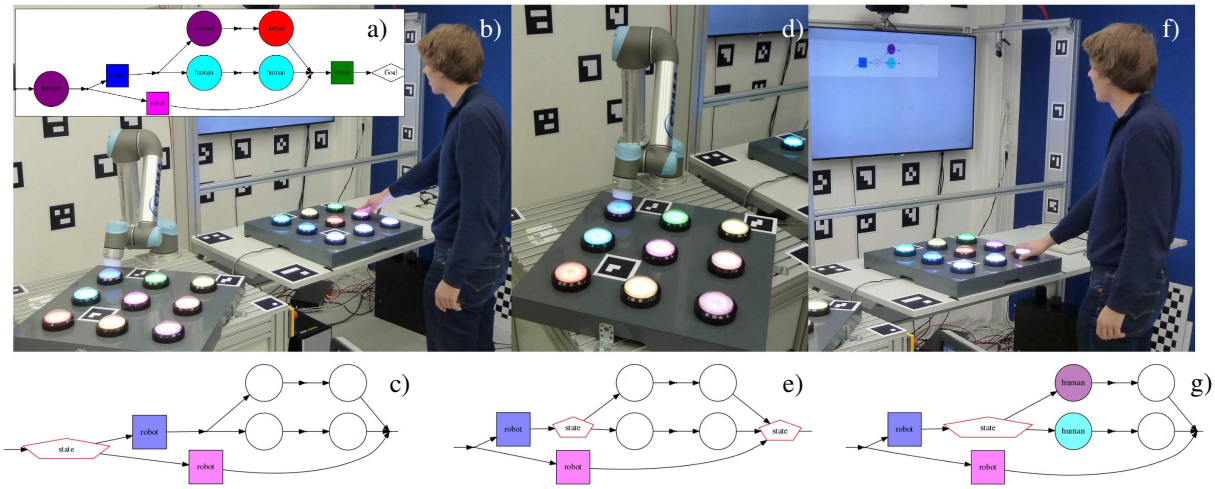
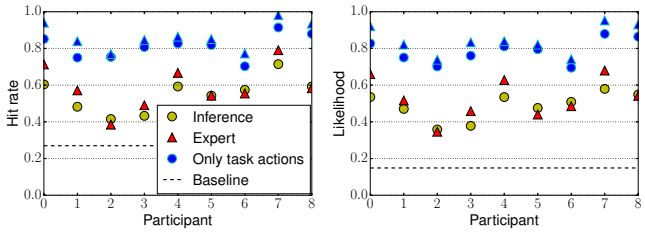


Fig. 4: Example case, a) describes the task, b), d) and f) show three consecutive button presses, c), e) and g) represent the estimated human belief after the corresponding action.



(a) Hit rate (b) Mean predictive likelihood

Fig. 5: Evaluation of action prediction

Inspired by the needs and structure of a cooperative manufacturing task, we designed a human-robot experiment, where we applied the belief inference method. We looked at different cases of the experiment and could verify the qualitative consistence with human observers. The action prediction based on the inferred belief produced promising results.

We will analyze the capabilities and limits of the proposed inference method, regarding parametrization as well as other application scenarios. One extension would be the use of Inverse Reinforcement Learning to infer the actual human reward function instead of static modeling.

Future research will further be directed towards enhancing cooperation of human and robot with the knowledge of internal human beliefs. Therefore we want to evaluate the human's situation awareness and adapt the robot behavior to warn or inform the human and to optimize the joint behavior. We think that we need the knowledge of human cognitive states to achieve the goal of making human-robot cooperation more efficient and intuitive.

REFERENCES

[1] M. R. Endsley, "Toward a Theory of Situation Awareness in Dynamic Systems," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 37, no. 1, pp. 32–64, 1995.

[2] B. Liddle and D. Nettle, "Higher-order theory of mind and social competence in school-age children," *Journal of Cultural and Evolutionary Psychology*, vol. 4, no. 3, pp. 231–244, 2006.

[3] A. Y. Ng and S. J. Russell, "Algorithms for Inverse Reinforcement Learning," in *Proceedings of the Seventeenth International Conference on Machine Learning (ICML)*, 2000.

[4] D. Hadfield-Menell, A. Dragan, P. Abbeel, and S. Russell, "Cooperative inverse reinforcement learning," in *Advances in Neural Information Processing Systems*, 2016, pp. 3916–3924.

[5] D. Sadigh, S. Sastry, S. A. Seshia, and A. D. Dragan, "Planning for Autonomous Cars that Leverage Effects on Human Actions," in *Robotics: Science and Systems XII*, 2016.

[6] L. Paletta, A. Dini, C. Murko, S. Yahyanejad, M. Schwarz, G. Lodron, S. Ladstätter, G. Paar, and R. Velik, "Towards Real-time Probabilistic Evaluation of Situation Awareness from Human Gaze in Human-Robot Interaction," in *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction - HRI*, 2017.

[7] T. Baer, D. Linke, D. Nienhuser, and J. M. Zollner, "Seen and missed traffic objects: A traffic object-specific awareness estimation," in *IEEE Intelligent Vehicles Symposium Workshops (IV Workshops)*, 2013, pp. 31–36.

[8] J. Schwehr and V. Willert, "Driver's gaze prediction in dynamic automotive scenes," in *IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*, 2017, pp. 1–8.

[9] O. Palinko, F. Rea, G. Sandini, and A. Sciutti, "Robot reading human gaze: Why eye tracking is better than head tracking for human-robot collaboration," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2016, pp. 5048–5054.

[10] C. A. Rothkopf and D. H. Ballard, "Modular inverse reinforcement learning for visuomotor behavior." *Biological cybernetics*, vol. 107, no. 4, pp. 477–90, 2013.

[11] L. Zettlemoyer, B. Milch, and L. Kaelbling, "Multi-agent filtering with infinitely nested beliefs," *Advances in Neural Information Processing Systems*, pp. 1–8, 2009.

[12] C. L. Baker, J. Jara-Ettinger, R. Saxe, and J. B. Tenenbaum, "Rational quantitative attribution of beliefs, desires and percepts in human mentalizing," *Nature Human Behaviour*, vol. 1, no. 4, p. 0064, 2017.

[13] L. P. Kaelbling, M. L. Littman, and A. R. Cassandra, "Planning and acting in partially observable stochastic domains," *Artificial Intelligence*, vol. 101, no. 1-2, pp. 99–134, 1998.

[14] A. Somani, N. Ye, D. Hsu, and W. S. Lee, "DESPOT: Online POMDP Planning with Regularization," in *Advances in Neural Information Processing Systems*, 2013, pp. 1772–1780.

[15] N. Cowan, "The magical number 4 in short-term memory: a reconsideration of mental storage capacity." *The Behavioral and brain sciences*, vol. 24, no. 1, pp. 87–114, 2001.