HHS Public Access

Author manuscript

Proc ACM SIGCHI. Author manuscript; available in PMC 2019 January 11.

Published in final edited form as:

Proc ACM SIGCHI. 2016 March ; 2016: 75–82. doi:10.1109/HRI.2016.7451736.

Formalizing Human-Robot Mutual Adaptation: A Bounded Memory Model

Stefanos Nikolaidis[*], **Anton Kuznetsov**[*], **David Hsu**[†], and **Siddhartha Srinivasa**[*]

[*]The Robotics Institute, Carnegie Mellon University

[†]Department of Computer Science, National University of Singapore

Abstract

Mutual adaptation is critical for effective team collaboration. This paper presents a formalism for human-robot mutual adaptation in collaborative tasks. We propose the *bounded-memory adaptation model* (BAM), which captures human adaptive behaviors based on a bounded memory assumption. We integrate BAM into a partially observable stochastic model, which enables robot adaptation to the human. When the human is adaptive, the robot will guide the human towards a new, optimal collaborative strategy unknown to the human in advance. When the human is not willing to change their strategy, the robot adapts to the human in order to retain human trust. Human subject experiments indicate that the proposed formalism can significantly improve the effectiveness of human-robot teams, while human subject ratings on the robot performance and trust are comparable to those achieved by cross training, a state-of-the-art human-robot team training practice.

I. Introduction

The development of new robotic systems that operate in the same physical space as people highlights the emerging need for robots that can integrate into human teams. Such systems can achieve significant economic and ergonomic benefits in manufacturing, as well as improve the quality of life of people at home. Previous work in human teaming has shown that mutual adaptation can significantly improve team performance [1]; we believe that the same holds for human-robot teams in collaborative tasks.

In previous work, human-robot cross-training has been shown to significantly improve subjective measures of team performance and metrics of team fluency [2]. The focus has been the computation of a robot policy aligned with the human preference, without taking into account the quality of that preference. This can result in the team executing sub-optimal policies, if for instance the human has an inaccurate model of the robot capabilities. On the other hand, given a known optimal way to execute the task, one could simply program the robot to always follow the optimal path, ignoring the actions of the human teammate. This, however, can have a negative effect on the human trust in the robot, affecting the willingness

snikolai@andrew.cmu.edu, antonk@andrew.cmu.edu, dyhsu@comp.nus.edu.sg, siddh@cs.cmu.edu

of people to work with their robotic teammates and ultimately damaging the overall team performance [3]–[5].

For instance, Fig. 1 illustrates a collaborative task, where human and robot are carrying a table outside of the room. There are two ways to finish the task, one with the robot facing the door (Goal 1, Fig. 1a-top) and one with the robot facing the room (Goal 2, Fig. 1a-bottom). We assume that Goal 1 is better, since the robot has a clear view of the door using its on-board sensor and the team is more likely to succeed in executing the task. The human does not have this information and may prefer to rotate the table towards Goal 2. Intuitively, if the human insists on the suboptimal goal, the robot should comply in order to finish the task. If the human is willing to adapt, the robot should guide them towards the optimal goal.

In this paper, we describe a formalism for human and robot mutual adaptation, where the robot builds a model of human adaptation to guide the human teammate towards more efficient strategies, while maintaining human trust to the robot. We first present Bounded memory Adaptation Model (BAM), a model based on a *bounded memory* assumption which limits the history length that the human team member considers in their decision making. BAM additionally assumes that each human teammate has an a priori willingness to adapt to the robot, which we define as *adaptability*. The adaptability of a participant is unknown beforehand and cannot be directly observed. Therefore, we denote it as a partially observable variable in a mixed-observability Markov decision process (MOMDP) [6]. The MOMDP formulation enables the robot to infer the adaptability of a human teammate through interaction and observation, and reason in a probabilistic sense over the ways the human can change their strategy.

We conducted a human subject experiment ($n = 69$) on a simulated table carrying task (Fig. 1). Fig. 2 shows different human and robot behaviors. If human and robot disagree on their strategies within an interaction history of 3 time-steps and the human insists in their strategy in the next time-step, then the MOMDP belief is updated so that smaller values of adaptability α have higher probability (lower adaptability). If the human switches to the robot strategy, larger values become more likely. The belief remains the same once human and robot agree on their strategies. If the robot infers the human to be non-adaptive, it complies to the human strategy. Otherwise, it guides them towards the optimal goal.

In the experiment, participants were significantly more likely to adapt to the robot strategy when working with a robot utilizing the proposed formalism ($p = 0.036$), compared to cross-training with the robot. Additionally, participants found the performance as a teammate of the robot executing the learned MOMDP policy to be not worse than the performance of a robot that cross-trained with the participants. Finally, the robot was found to be more trustworthy with the learned policy, compared with executing an optimal strategy while ignoring the adaptability of the human teammate ($p = 0.048$).

II. Relevant Work

There has been extensive work on one-way robot adaptation to the human. Approaches involve a human expert providing demonstrations to teach the robot a skill or a specific task

[7]–[12]. Robots have also been able to infer the human preference online through interaction. In particular, partially observable Markov decision process (POMDP) models have allowed reasoning over the uncertainty on the human intention [13], [14]. The MOMDP formulation [6] has been shown to achieve significant computational efficiency, and has been used in motion planning applications [15]. Recent work has also inferred human intention through decomposition of a game task into subtasks for game AI applications [16]. Alternatively, Macindoe et al. proposed the partially observable Monte-Carlo cooperative planning system, in which human intention is inferred for a turn-based game [17]. Nikolaidis et al. proposed a formalism to learn human types from joint-action demonstrations, infer online the type of a new user and compute a robot policy aligned to their preference [18]. Simultaneous intent inference and robot adaptation has also been achieved through propagation of state and temporal constraints [19]. Another approach has been the human-robot cross-training algorithm, where the human demonstrates their preference by switching roles with the robot, shaping the robot reward function [2]. Although it is possible that the human changes strategies during the training, the algorithm does not use a model of human adaptation that can enable the robot to actively influence the actions of its human partner.

There have also been studies in human adaptation to the robot. Previous work has focused on operator training for military, space and search-and-rescue applications, with the goal of reducing the operator workload and operational risk [20]. Additionally, researchers have studied the effects of repeated interactions with a humanoid robot on the interaction skills of children with autism [21], on language skills of elementary school students [22], as well as on users' spatial behavior [23]. Human adaptation has also been observed in an assistive walking task, where the robot uses human feedback to improve its behavior, which in turn influences the physical support provided by the human [24]. While the changes in the human behavior are an essential part of the learning process, the system does not explicitly reason over the human adaptation throughout the interaction. On the other hand, Dragan and Srinivasa proposed a probabilistic model of the inference made by a human observer over the robot goals, and introduced a motion generating algorithm to maximize this inference towards a predefined goal [25].

We believe that the proposed formalism for human-robot mutual adaptation closes the loop between the two streams of research. The robot reasons in a probabilistic sense over the different ways that the human may change their strategy, based on a model of human adaptation parameterized by the participant's willingness to adapt. It updates the model through interaction and guides participants towards more efficient strategies, while maintaining human trust to the robot.

Mutual adaptation between two agents has been extensively explored in the field of game theory [26]. Economic theory relies significantly on on strong assumptions about the rationality of the agents and the knowledge of the payoff functions. Such assumptions are not necessary applicable in settings where the players are not involved in a full computation of optimal strategies for themselves and the others [27]. We believe that this is particularly true in a human-robot team setting, where the human is uncertain on how the robot will act and has little time to respond. Therefore, we propose a model of human adaptive behavior

based on a *bounded memory* assumption [28]–[30] and integrate it into robot decision making.

III. Problem Setting

We formally describe the evolution of the human-robot collaborative task as a Multi-Agent Markov Decision Process (MMDP) [17] with a set of states Q: $X_{world} \times H_t$, robot actions $a_r \in A_r(x_{world})$ and human actions $a_h \in A_h(x_{world})$. X_{world} is a set of world states and H_t is the set of all possible histories of interactions until time t: $h_t = \{x_{world}(0), a_r(0), a_h(0), \ldots, x_{world}(t-1), a_r(t-1), a_h(t-1)\}$. The MMDP has a state transition function $T: Q \times A_r \times A_h \rightarrow \Pi(Q)$ and an immediate reward function $R: R(x_{world}, a_r, a_h) \mapsto r \in \mathbb{R}^+$.

We assume that the human is enacting a stochastic policy π_h unknown to the robot. The human policy can be arbitrarily nuanced: $\pi_h: X_{world} \times H_t \rightarrow \Pi(A_h)$.

The robot's goal is to compute its optimal policy $\pi_r: X_{world} \times H_t \rightarrow A_r$ that maximizes the expected discounted accumulated reward:

$$\pi_r^* = \arg\max_{\pi_r} \mathbb{E}_{\pi_h}\left[\sum_{t=0}^{\infty} \gamma^t r(t)\right] \quad (1)$$

where $\gamma \in [0, 1)$ is a discount factor that downweighs future rewards. Note here that because the human policy is unknown to the robot, it has no choice but to reason (and take expectations over) all possible human policies π_h.

In this work, we present BAM, a model of human adaptation which specifies a parameterization of the human policy π_h. We define a set of *modal policies* or *modes* M, where $m \in M$ is a deterministic policy mapping states and histories to joint human-robot actions: $m: X_{world} \times H_t \times A_r \times A_h \rightarrow \{0, 1\}$. At the time-step, the human has a mode $m_h \in M$ and perceives the robot as following a mode $m_r \in M$. Then, in the next time-step the human may switch to m_r with some probability α. If m_h maximizes the expected accumulated reward, the robot optimal policy would be to take actions a_r specified by m_h. If m_h is suboptimal and $\alpha = 1$, the robot optimal policy would be to follow m_r, expecting the human to adapt. In the general case of an unknown α, how can we compute π_r^* in Eq. (1)?

We approach this problem using a MOMDP formulation, wherein α is an unobserved variable. This formulation allows us to estimate α through interaction and integrate predictions of the human actions into robot action selection (Fig 3).

IV. The Bounded Memory Adaptation Model

We model the human policy π_h as a probabilistic finite-state automaton (PFA), with a set of states $Q: X_{world} \times H_t$. A joint human-robot action a_h, a_r triggers an emission of a human and robot modal policy $f: Q \times M \times M \rightarrow \{0, 1\}$, as well as a transition to a new state $P: Q \rightarrow \Pi(Q)$.

A. Bounded Memory Assumption

Herbert Simon proposed that people often do not have the time and cognitive capabilities to make perfectly rational decisions, in what he described as “bounded rationality” [31]. This idea has been supported by studies in psychology and economics [32]. In game theory, bounded rationality has been modeled by assuming that players have a “bounded memory” or “bounded recall” and base their decisions on recent observations [28]–[30]. In this work, we introduce the bounded memory assumption in a human-robot collaboration setting. Under this assumption, humans will choose their action based on a history of k-steps in the past, so that $Q: X_{world} \times H_k$.

B. Feature Selection

The size of the state-space in the PFA can be quite large ($|X_{world}|^{k+1}|A_r|^k|A_h|^k|$). Therefore, we approximate it using a set of features, so that $\phi(q) = \{\phi_1(q), \phi_2(q), ..., \phi_N(q)\}$. We can choose as features the frequency counts ϕ_m^h, ϕ_m^r of the modal policies followed in interaction history, so that:

$$\phi_m^h = \sum_{i=1}^{k} \left[m_h^i = m\right] \quad \phi_m^r = \sum_{i=1}^{k} \left[m_r^i = m\right] \quad \forall m \in M \qquad (2)$$

m_h^i and m_r^i is the modal policy of the human and the robot i time-steps in the past. We note that k defines the history length, with $k = 1$ implying that the human will act based only on the previous interaction. Drawing upon insights from previous work which assumes maximum likelihood observations for policy computation in belief-space [33], we used as features the modal policies with the maximum frequency count: $m_h = \arg\max_m \phi_m^h$, $m_r = \arg\max_m \phi_m^r$.

The proposed model does not require a specific feature representation. For instance, we could construct features by combining modal policies m_h^i, m_r^i using an arbitration function [34]. Additionally, rather than using frequency counts, we could maintain a probability distribution over human and robot modes given the history, but we leave this for future work.

C. Human Adaptability

We define the adaptability as the probability of the human switching from their mode to the robot mode. It would be unrealistic to assume that all users are equally likely to adapt to the robot. Instead, we account for individual differences by parameterizing the transition function P by the *adaptability* α of an individual. Then, at state q the human will transition to a new state by choosing an action specified by m_r with probability α, or an action specified by m_h with probability $1 - \alpha$ (Fig. 4). We include noise in the model, by assuming that the human can take any other action uniformly with some probability ϵ.

V. Robot Planning

In this section we describe the integration of BAM in the robot decision making process using a MOMDP formulation. A MOMDP uses proper factorization of the observable and unobservable state variables $S: X \times Y$ with transition functions $\mathcal{T}_x$ and $\mathcal{T}_y$, reducing the computational load [6]. The set of observable state variables is $X: X_{world} \times M^k \times M^k$, where X_{world} is the finite set of task-steps that signify progress towards task completion and M is the set of modal policies followed by the human and the robot in a history length k. The partially observable variable y is identical to the human adaptability α. We assume finite sets of human and robot actions A_h and A_r, and we denote as π_h the stochastic human policy. The latter gives the probability of a human action a_h at state s, based on the BAM human adaptation model.

Given $a_r \in A_r$ and $a_h \in A_h$, the belief update becomes:

$$b'(y') = \eta O(s', a_r, o) \sum_{y \in Y} \mathcal{T}_x(s, a_r, a_h, x') \mathcal{T}_y(s, a_r, a_h, s') \pi_h(s, a_h) b(y) \quad (3)$$

We use a point-based approximation algorithm to solve the MOMDP for a robot policy π_r that takes into account the robot belief on the human adaptability, while maximizing the agent's expected total reward.

The policy execution is performed online in real time and consists of two steps (Fig. 3). First, the robot uses the current belief to select the action a_r specified by the policy. Second, it uses the human action a_h to update the belief on α (Eq. 3). Fig. 2 shows different user behaviors in the human subject experiment described in Sec. VI. Fig. 5 shows the corresponding paths on the MOMDP policy tree.

VI. Human Subject Experiment

We conducted a human subject experiment on a simulated table-carrying task (Fig. 1) to evaluate the proposed formalism. We were interested in showing that integrating BAM into the robot decision making can lead to more efficient policies than state-of-the-art human-robot team training practices, while maintaining human satisfaction and trust.

On one extreme, we can “fix” the robot policy so that the robot always moves towards the optimal goal, ignoring human adaptability. This will force all users to adapt, since this is the only way to complete the task. However, we hypothesize that this will significantly impact human satisfaction and trust in the robot. On the other extreme, we can efficiently learn the human preference [2]. This can lead to the human-robot team following a sub-optimal policy, if the human has an inaccurate model of the robot capabilities. We show that the proposed formalism achieves a trade-off between the two: When the human is non-adaptive, the robot follows the human strategy. Otherwise, the robot insists on the optimal way of

completing the task, leading to significantly better policies compared to learning the human preference, while maintaining human trust.

A. Independent Variables

We had three experimental conditions, which we refer to as “Fixed,” “Mutual-adaptation” and “Cross-training.”

Fixed session—The robot executes a fixed policy, always acting towards the optimal goal. In the table-carrying scenario, the robot keeps rotating the table in the clockwise direction towards Goal 1, which we assume to be optimal (Fig. 1). The only way to finish the task is for the human to rotate the table in the same direction as the robot, until it is brought to the horizontal configuration of Fig. 1a-top.

Mutual-adaptation session—The robot executes the MOMDP policy computed using the proposed formalism. The robot starts by rotating the table towards the optimal goal (Goal 1). Therefore, adapting to the robot strategy corresponds to rotating the table to the optimal configuration.

Cross-training session—Human and robot train together using the human-robot cross-training algorithm [2]. The algorithm consists of a forward phase and a rotation phase. In the forward phase, the robot executes an initial policy, which we choose to be the one that leads to the optimal goal. Therefore, in the table-carrying scenario, the robot rotates the table in the clockwise direction towards Goal 1. In the rotation phase, human and robot switch roles, and the human inputs are used to update the robot reward function. After the two phases, the robot policy is recomputed.

B. Hypotheses

H1—Participants will agree more strongly that the robot is trustworthy, and will be more satisfied with the team performance in the Mutual-adaptation condition, compared to working with the robot in the Fixed condition. We expected users to trust more the robot with the learned MOMDP policy, compared with the robot that executes a fixed strategy ignoring the user's willingness to adapt. In prior work, a task-level executive that adapted to the human partner significantly improved perceived robot trustworthiness [35]. Additionally, working with a human-aware robot that adapted its motions had a significant impact on human satisfaction [36].

H2—Participants are more likely to adapt to the robot strategy towards the optimal goal in the Mutual-adaptation condition, compared to working with the robot in the Cross-training condition. The computed MOMDP policy enables the robot to infer online the adaptability of the human and guides adaptive users towards more effective strategies. Therefore, we posited that more subjects would change their strategy when working with the robot in the Mutual-adaptation condition, compared with cross-training with the robot. We note that in the Fixed condition all participants ended up changing to the robot strategy, as this was the only way to complete the task.

H3—The robot performance as a teammate, as perceived by the participants in the Mutual-adaptation condition, will not be worse than in the Cross-training condition. The learned MOMDP policy enables the robot to follow the preference of participants that are less adaptive, while guiding towards the optimal goal participants that are willing to change their strategy. Therefore, we posited that this behavior would result on a perceived robot performance not inferior to that achieved in the Cross-training condition.

C. Experiment Setting: A Table Carrying Task

We first instructed participants in the task, and asked them to choose one of the two goal configurations (Fig. 1a), as their preferred way of accomplishing the task. To prompt users to prefer the sub-optimal goal, we informed them about the starting state of the task, where the table was slightly rotated in the counter-clockwise direction, making the suboptimal Goal 2 appear closer. Once the task started, the user chose the rotation actions by clicking on buttons on a user interface (Fig. 1b). If the robot executed the same action, a video played showing the table rotation. Otherwise, the table did not move and a message appeared on the screen notifying the user that they tried to rotate the table in a different direction than the robot. In the Mutual-adaptation and Fixed conditions participants executed the task twice. Each round ended when the team reached one of the two goal configurations. In the Cross-training condition, participants executed the forward phase of the algorithm in the first round and the rotation phase, where human and robot switched roles, in the second round. We found that in this task one rotation phase was enough for users to successfully demonstrate their preference to the robot. Following [2], the robot executed the updated policy with the participant in a task-execution phase that succeeded the rotation phase. We asked all participants to answer a post-experimental questionnaire that used a five-point Likert scale to assess their responses to working with the robot. They also responded to open-ended questions about their experience.

D. Subject Allocation

We chose a between-subjects design in order to not bias the users with policies from previous conditions. We recruited participants through Amazon's Mechanical Turk service. Since we are interested in exploring human-robot mutual adaptation, we disregarded participants that had as initial preference the robot goal. To ensure reliability of the results, we asked all participants a control question that tested their attention to the task and eliminated data associated with wrong answers to this question, as well as incomplete data. To test their attention to the Likert questionnaire, we included a negative statement with the opposite meaning to its positive counterpart and eliminated data associated with positive or negative ratings to both statements, resulting in a total of 69 samples.

E. MOMDP Model

The observable state variables x of the MOMDP formulation were the discretized table orientation and the human and robot modes for each of the three previous time-steps. We specified two modal policies, each deterministically selecting rotation actions towards each goal. The size of the observable state-space X was 734 states. We set a history length $k = 3$ in BAM. We additionally assumed a discrete set of values of the adaptability α : {0.0, 0.25,

0.5, 0.75, 1.0}. Therefore, the total size of the MOMDP state-space was 5 × 734 = 3670 states. The human and robot actions a_h, a_r were deterministic discrete table rotations. We set the reward function R to be positive at the two goal configurations based on their relative cost, and 0 elsewhere. We computed the robot policy using the SARSOP solver [37], a point-based approximation algorithm which, combined with the MOMDP formulation, can scale up to hundreds of thousands of states [15].

VII. Results and Discussion

A. Subjective Measures

We consider hypothesis **H1**, that participants will agree more strongly that the robot is trustworthy, and will be more satisfied with the team performance in the Mutual-adaptation condition, compared to working with the robot in the Fixed condition. A two-tailed Mann-Whitney-Wilcoxon test showed that participants indeed agreed more strongly that the robot utilizing the proposed formalism is trustworthy ($U = 180$, $p = 0.048$). No statistically significant differences were found for responses to statements eliciting human satisfaction: "I was satisfied with the robot and my performance" and "The robot and I collaborated well together." One possible explanation is that participants interacted with the robot through a user interface for a short period of time, therefore the impact of the interaction on user satisfaction was limited.

We were also interested in observing how the ratings in the first two conditions varied, depending on the participants' willingness to change their strategy. Therefore, we conducted a post-hoc experimental analysis of the data, grouping the participants based on their adaptability. Since the true adaptability of each participant is unknown, we estimated it by the mode of the belief formed by the robot at the end of the task on the adaptability α:

$$\hat{\alpha} = \arg\max_{\alpha} b(\alpha) \quad (4)$$

We considered only users whose mode was larger than a confidence threshold and grouped them as *very adaptive* if $\hat{\alpha} > 0.75$, *moderately adaptive* if $0.5 < \hat{\alpha} \leq 0.75$ and *non-adaptive* if $\hat{\alpha} \leq 0.5$. Fig. 7b shows the participants' rating of their agreement on the robot trustworthiness, as a function of the participants' group for the two conditions. In the Fixed condition there was a trend towards positive correlation between the annotated robot trustworthiness and participants' inferred adaptability (Pearson's $r = 0.452$, $p = 0.091$), whereas there was no correlation between the two for participants in the Mutual-adaptation condition ($r = -0.066$). We attribute this to the MOMDP formulation allowing the robot to reason over its estimate on the adaptability of its teammate and change its own strategy when interacting with non-adaptive participants, therefore maintaining human trust.

Interestingly, when asked to comment on the robot behavior, several adaptive participants in both conditions attempted to justify the robot actions, stating that "probably there was no room to rotate [counter-clockwise]," and that "maybe the robot could not move backwards." Some non-adaptive participants in the Fixed condition used stronger language, noting that

"the robot is incapable of adapting to my efforts," and that it was "stubborn and would not let us turn in the direction that would make me do the least amount of work." On the other hand, non-adaptive participants in the Mutual-adaptation condition mentioned that the robot "attempted to anticipate my moves" and "understood which way I wanted to go."

Recall hypothesis **H3:** that the robot performance as a teammate in the Mutual-adaptation condition, as perceived by the participants, would not be worse than in the Cross-training condition. We define "not worse than" similarly to [38] using the concept of "non-inferiority" [39]. An one-tailed unpaired t-test for a non-inferiority margin $\Delta = 0.5$ and a level of statistical significance $\alpha = 0.025$ showed that participants in the Mutual-adaptation condition rated their satisfaction on robot performance ($p = 0.006$), robot intelligence ($p = 0.024$), robot trustworthiness ($p < 0.001$), quality of robot actions ($p < 0.001$) and quality of collaboration ($p = 0.002$) not worse than participants in the Cross-training condition. This supports hypothesis **H3** of Sec. VI-B.

B. Quantitative Measures

To test hypothesis **H2**, we consider the ratio of participants that changed their strategy to the robot strategy towards the optimal goal in the Mutual-adaptation and Cross-training conditions. A change was detected when the participant stated as preferred strategy a table rotation towards Goal 2 (Fig. 1a-bottom), but completed the task in the configuration of Goal 1 (Fig. 1a-top) in the final round of the Mutual-adaptation session, or in the task-execution phase of the Cross-training session. As Fig. 7a shows, 57% of participants adapted to the robot in the Mutual-adaptation condition, whereas 26% adapted to the robot in the Cross-training condition. A Pearson's chi-square test showed that the difference is statistically significant ($\chi^2(1, N = 46) = 4.39$, $p = 0.036$). Therefore, participants that interacted with the robot of the proposed formalism were more likely to switch to the robot strategy towards the optimal goal, than participants that cross-trained with the robot, which supports our hypothesis.

In Sec. VII-C, we discuss the robot behavior for different values of history length k in BAM.

C. Selection of History Length

The value of k in BAM indicates the number of time-steps in the past that we assume humans consider in their decision making on a particular task, ignoring all other history. Increasing k results in an exponential increase of the state space size, with large values reducing the robot responsiveness to changes in the human behavior. On the other hand, very small values result in unrealistic assumptions on the human decision making process.

To illustrate this, we set $k = 1$ and ran a pilot study of 30 participants through Amazon-Turk. Whereas most users rated highly their agreement to questions assessing their satisfaction and trust to the robot, some participants expressed their strong dissatisfaction with the robot behavior. This occurred when human and robot oscillated back and forth between modes, similarly to when two pedestrians on a narrow street face each other and switch sides simultaneously until they reach an agreement. In this case, which occurred in 23% of the samples, when the human switched back to their initial mode, which was also the robot

mode of the previous time-step, the robot incorrectly inferred them as adaptive. However, the user in fact resumed their initial mode followed before two time-steps, implying a tendency for non-adaptation. This is a case where the 1-step bounded memory assumption did not hold.

In the human subject experiment of Sec VI, we used $k = 3$, since we found this to describe accurately the human behavior in this task. Fig. 6 shows the belief update and robot behavior for $k = 1$ and $k = 3$, in the case of mode oscillation.

D. Discussion

These results show that the proposed formalism enables a human-robot team to achieve more effective policies, compared to state-of-the-art human-robot team training practices, while achieving subjective ratings on robot performance and trust that are comparable to those achieved by these practices. It is important to note that the comparison with the human-robot cross-training algorithm is done in the context of human adaptation. Previous work [2] has shown that switching roles can result in significant benefits in team fluency metrics, such as human idle time and concurrent motion [40], when a human executes the task with an actual robot. Additionally, the proposed formalism assumes as input a set of modal policies, as well as a quality measure associated with each policy. On the other hand, cross-training requires only an initialization of a reward function of the state space, which is then updated in the rotation phase through interaction. It would be very interesting to explore a hybrid approach between learning the reward function and guiding the human towards an optimal policy, but we leave this for future work.

E. Generalization to Complex Tasks

The presented table-carrying task can be generalized without any modifications in the proposed mathematical model, with the cost of increasing the size of the state-space and action-space. In particular, we made the assumptions: (1) discrete time-steps, where human and robot apply torques causing a fixed table-rotation. (2) binary human-robot actions. We discuss how we can relax these assumptions:

1) We can approximate a continuous-time setting by increasing the resolution of the time discretization. Assuming a constant displacement per unit time v and a time-step dt, the size of the state-space increases linearly with $(1/dt)$: $O(|X_{world}| |M|^2 k) = O((\theta_{max} - \theta_{min}) * (1/v) * (1/dt) * |M|^2 k)$, where θ is the rotation angle of the table.

2) The proposed formalism is not limited to binary actions. For instance, we can allow torque inputs of different magnitudes. The action-space of the MOMDP increases linearly with the number of possible inputs.

Finally, we note that the presented formalism does not rely on any features of the table-carrying task. For instance, we could apply our formalism in the case where human and robot cross a hallway and coordinate to avoid collision, and the robot guides the human towards the right side of the corridor. Alternatively, in an assembly manufacturing task the robot could lead the human to strategies that require less time or resources.

VIII. Conclusion

We presented a formalism for human-robot mutual adaptation, which enables guiding the human teammate towards more efficient strategies, while maintaining human trust in the robot. First, we proposed BAM, a model of human adaptation based on a bounded memory assumption. The model is parameterized by the adaptability of the human teammate, which takes into account individual differences in people's willingness to adapt to the robot. We then integrated BAM into a MOMDP formulation, wherein the adaptability was a partially observable variable. In a human subject experiment ($n = 69$), participants were significantly more likely to adapt to the robot strategy towards the optimal goal when working with a robot utilizing our formalism ($p = 0.036$), compared to cross-training with the robot. Additionally, participants found the performance as a teammate of the robot executing the learned MOMDP policy to be not worse than the performance of the robot that cross-trained with the participants. Finally, the robot was found to be more trustworthy with the learned policy, compared with executing an optimal strategy while ignoring human adaptability ($p = 0.048$). These results indicate that the proposed formalism can significantly improve the effectiveness of human-robot teams, while achieving subjective ratings on robot performance and trust comparable to those of state-of-the-art human-robot team training strategies.

We have shown that BAM can adequately capture human behavior in a collaborative task with well-defined task-steps on a relatively fast-paced domain. However, in domains where people typically reflect on a long history of interactions, or on the beliefs of the other agents, such as in a Poker game [41], people are likely to demonstrate much more complex adaptive behavior. Developing sophisticated predictive models for such domains and integrating them into robot decision making in a principled way, while maintaining computational tractability, is an exciting area for future work.

ACknowledgments

We thank Michael Koval and Shervin Javdani for very helpful discussion and advice. This work was funded by the DARPA SIMPLEX program through ARO contract number 67904LSDRP, National Institute of Health R01 (#R01EB019335), National Science Foundation CPS (#1544797), and the Office of Naval Research. We also acknowledge the Onassis Foundation as a sponsor.

REFERENCES

1. Mathieu JE, et al. The influence of shared mental models on team process and performance. Journal of applied psychology. 2000
2. Nikolaidis S, Shah J. Human-robot cross-training: computational formulation, modeling and evaluation of a human team training strategy. HRI. 2013
3. Hancock PA, Billings DR, Schaefer KE, Chen JY, De Visser EJ, Parasuraman R. A meta-analysis of factors affecting trust in human-robot interaction. Human Factors. 2011
4. Salem M, Lakatos G, Amirabdollahian F, Dautenhahn K. Would you trust a (faulty) robot?: Effects of error, task type and personality on human-robot cooperation and trust. HRI. 2015
5. Lee JJ, Knox WB, Wormwood JB, Breazeal C, DeSteno D. Computationally modeling interpersonal trust. Front. Psychol. 2013
6. Ong SC, Png SW, Hsu D, Lee WS. Planning under uncertainty for robotic tasks with mixed observability. IJRR. 2010

7. Argall BD, Chernova S, Veloso M, Browning B. A survey of robot learning from demonstration. Robot. Auton. Syst. 2009
8. Atkeson CG, Schaal S. Robot learning from demonstration. ICML. 1997
9. Abbeel P, Ng AY. Apprenticeship learning via inverse reinforcement learning. ICML. 2004
10. Nicolescu MN, Mataric MJ. Natural methods for robot task learning: Instructive demonstrations, generalization and practice. AAMAS. 2003
11. Chernova S, Veloso M. Teaching multi-robot coordination using demonstration of communication and state sharing. AAMAS. 2008
12. Akgun B, Cakmak M, Yoo JW, Thomaz AL. Trajectories and keyframes for kinesthetic teaching: a human-robot interaction perspective. HRI. 2012
13. Lemon, O, Pietquin, O. Data-Driven Methods for Adaptive Spoken Dialogue Systems: Computational Learning for Conversational Interfaces. Springer Publishing Company, Incorporated; 2012.
14. Broz F, Nourbakhsh I, Simmons R. Designing pomdp models of socially situated tasks. RO-MAN. 2011
15. Bandyopadhyay, T, Won, KS, Frazzoli, E, Hsu, D, Lee, WS, Rus, D. WAFR. Springer; 2013. Intention-aware motion planning.
16. Nguyen T-HD, Hsu D, Lee WS, Leong T-Y, Kaelbling LP, Lozano-Perez T, Grant AH. Capir: Collaborative action planning with intention recognition. AIIDE. 2011
17. Macindoe O, Kaelbling LP, Lozano-Pérez T. Pomcop: Belief space planning for sidekicks in cooperative games. AIIDE. 2012
18. Nikolaidis S, Ramakrishnan R, Gu K, Shah J. Efficient model learning from joint-action demonstrations for human-robot collaborative tasks. HRI. 2015
19. Karpas E, Levine SJ, Yu P, Williams BC. Robust execution of plans for human-robot teams. ICAPS. 2015
20. Goodrich MA, Schultz AC. Human-robot interaction: a survey. Foundations and trends in human-computer interaction. 2007
21. Robins B, Dautenhahn K, Te Boekhorst R, Billard A. Effects of repeated exposure to a humanoid robot on children with autism. Designing a more inclusive world. 2004
22. Kanda T, Hirano T, Eaton D, Ishiguro H. Interactive robots as social partners and peer tutors for children: A field trial. Human-computer interaction. 2004
23. Green A, Httenrauch H. Making a case for spatial prompting in human-robot communication, in multimodal corpora: From multimodal behaviour theories to usable models. workshop at LREC. 2006
24. Ikemoto S, Amor HB, Minato T, Jung B, Ishiguro H. Physical human-robot interaction: Mutual learning and adaptation. IEEE Robot. Autom. Mag. 2012
25. Dragan A, Srinivasa S. Generating legible motion. RSS. 2013
26. Fudenberg D, Tirole J. Game theory mit press. 1991
27. Fudenberg, D, Levine, DK. The theory of learning in games. MIT press; 1998.
28. Powers R, Shoham Y. Learning against opponents with bounded memory. IJCAI. 2005
29. Monte D. Learning with bounded memory in games. GEB. 2014
30. Aumann RJ, Sorin S. Cooperation and bounded recall. GEB. 1989
31. Simon HA. Rational decision making in business organizations. The American economic review. 1979:493–513.
32. Kahneman D. Maps of bounded rationality: Psychology for behavioral economics. American economic review. 2003:1449–1475.
33. Platt R, Tedrake R, Kaelbling L, Lozano-Perez T. Belief space planning assuming maximum likelihood observations. RSS. 2010
34. Dragan A, Srinivasa S. Formalizing assistive teleoperation. RSS. 2012
35. Shah J, Wiken J, Williams B, Breazeal C. Improved human-robot team performance using chaski, a human-inspired plan execution system. HRI. 2011

36. Lasota PA, Shah JA. Analyzing the effects of human-aware motion planning on close-proximity human–robot collaboration. Hum. Factors. 2015

37. Kurniawati H, Hsu D, Lee WS. Sarsop: Efficient point-based pomdp planning by approximating optimally reachable belief spaces. RSS. 2008

38. Dragan AD, Srinivasa SS, Lee KC. Teleoperation with intelligent and customizable interfaces. JHRI. 2013

39. Lesaffre E. Superiority, equivalence, and non-inferiority trials. Bulletin of the NYU hospital for joint diseases. 2008

40. Hoffman G, Breazeal C. Effects of anticipatory action on human-robot teamwork efficiency, fluency, and perception of team. HRI. 2007

41. Von Neumann, J, Morgenstern, O. Theory of games and economic behavior. Princeton university press; 2007.

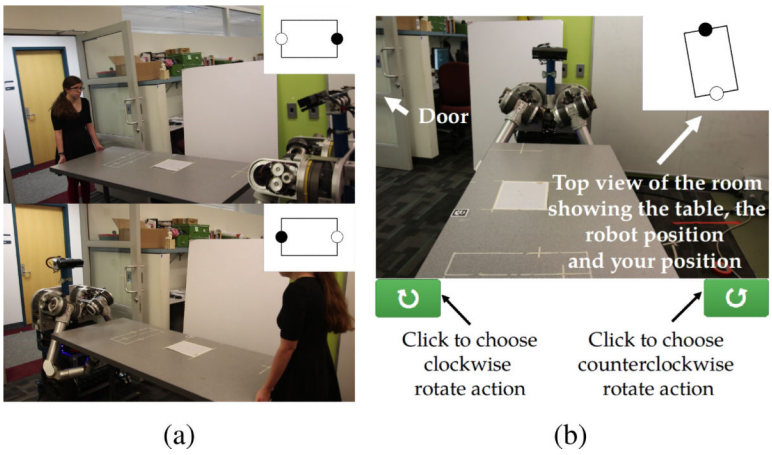

Fig. 1.

(a) Human-robot table carrying task. Rotating the table so that the robot is facing the door (top, Goal 1) is better than the other direction (bottom, Goal 2), since the exit is included in the robot's field of view and the robot can avoid collisions. (b) UI with instructions.

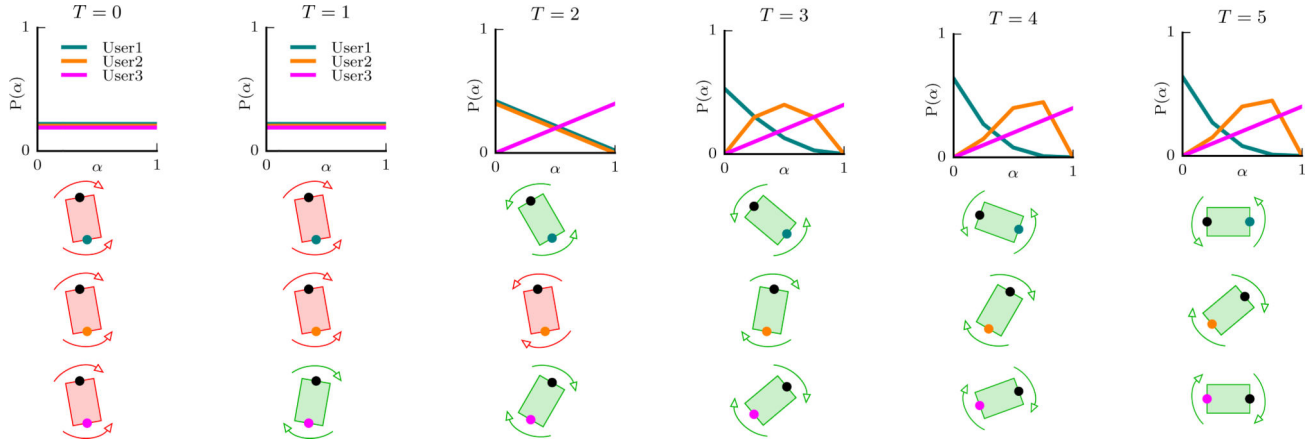

Fig. 2.

(Top row) MOMDP belief update on human adaptability $a \in \{0, 0.25, 0.5, 0.75, 1.0\}$ for three different users in the human subject experiment of Sec. VI. Larger values of a indicate higher adaptability. (Second, third and bottom row) The rows correspond to Users 1, 2 and 3 and show the table configuration at each time-step of task execution. Columns indicate different time-steps. Red color indicates human and robot disagreement in their actions, in which case the table does not rotate. User 1 (teal dot) insists on their initial strategy throughout the task and the robot (black dot) complies, whereas Users 2 and 3 (orange and magenta dot) adapt to the robot.

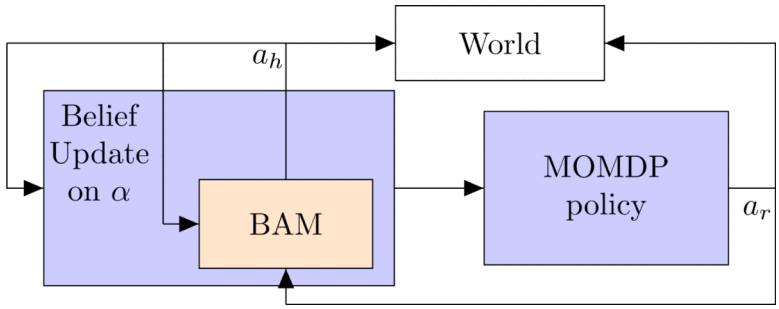

Fig. 3.
Integration of BAM into MOMDP formulation.

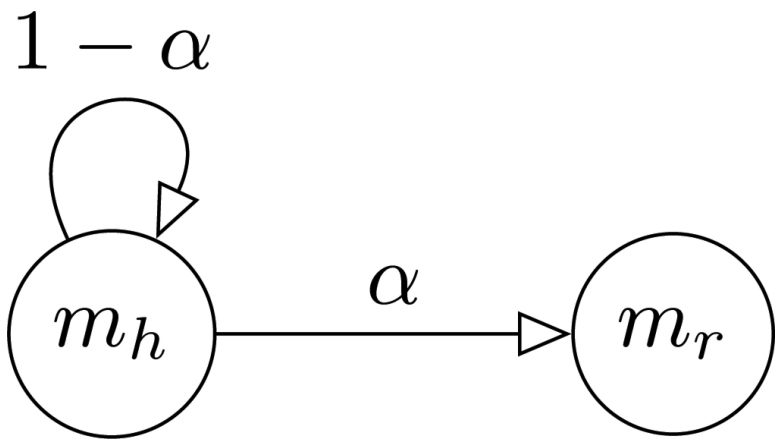

Fig. 4.
The BAM human adaptation model.

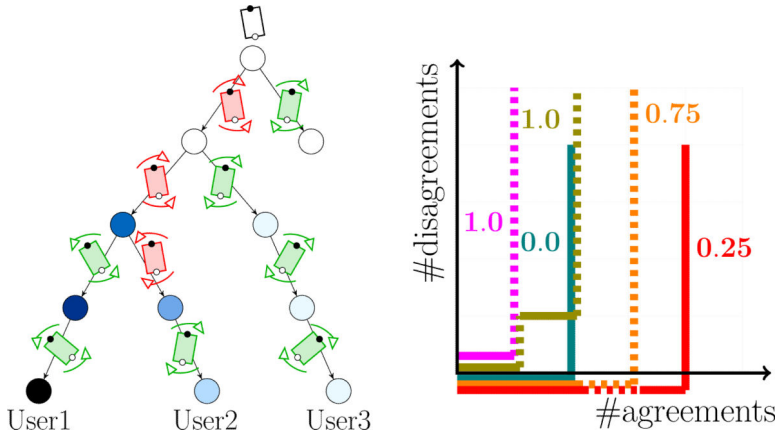

Fig. 5.

(Left) Different paths on MOMDP policy tree for human-robot (white/black dot) table-carrying task. The circle color represents the belief on a, with darker shades indicating higher probability for smaller values (less adaptability). The white circles denote a uniform distribution over a. User 1 is non-adaptive, whereas Users 2 and 3 are adaptive. (Right) Instances of different user behaviors in the first round of the Mutual-adaptation session. A horizontal/vertical line segment indicates human and robot disagreement/agreement on their actions. A solid/dashed line indicates a human rotation towards the sub-optimal/optimal goal. The numbers denote the most likely estimated value of a.

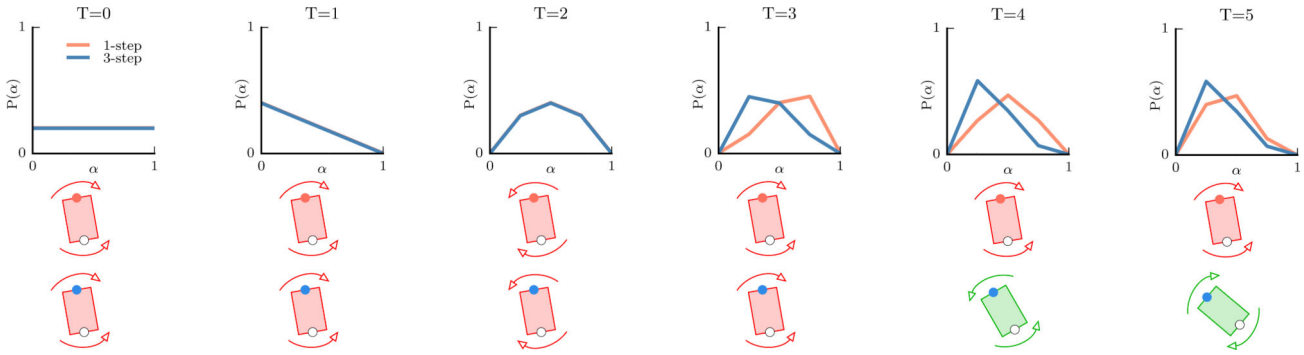

Fig. 6.

(Top row) Belief update for the 1-step and 3-step bounded memory models at successive time-steps. (Middle/bottom row) Table configuration in the 1-step/3-step trial. ($T = 1$) After the first disagreement and in the absence of any previous history, the belief remains uniform over a. The human (white dot) follows their modal policy from the previous time-step, therefore at $T = 2$ the belief becomes higher for smaller values of a in both models (lower adaptability). ($T = 2$) The robot (orange dot for 1-step, blue dot for 3-step) adapts to the human and executes the human modal policy. At the same time, the human switches to the robot mode, therefore at $T = 3$ the probability mass moves to the right. ($T = 3$) The human switches back to their initial mode. In the 3-step model the resulting distribution at $T = 4$ has a positive skewness: the robot estimates the human to be not adaptive. In the 1-step model the robot incorrectly infers that the human adapted to the robot mode of the previous time-step, and the probability distribution has a negative skewness. ($T = 4, 5$) The robot in the 3-step trial switches to the human modal policy, whereas in the 1-step trial it does not adapt to the human, who insists on their mode.

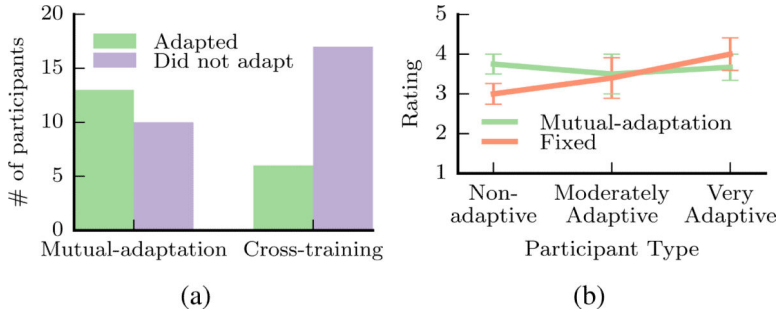

Fig. 7.

(a) Number of participants that adapted to the robot for the Mutual-adaptation and Cross-training conditions. (b) Rating of agreement to statement “The robot is trustworthy.” Note that the figure does not include participants, whose mode of the belief on their adaptability was below a confidence threshold and therefore were not clustered into any of the three groups.