FABRIC: A Framework for the Design and Evaluation of Collaborative Robots with Extended Human Adaptation

O. CAN GÖRÜR, DAI-Labor, Technische Universität Berlin
BENJAMIN ROSMAN, University of the Witwatersrand
FIKRET SIVRIKAYA, GT-ARC gemeinnützige GmbH
SAHIN ALBAYRAK, DAI-Labor, Technische Universität Berlin

A limitation for collaborative robots (cobots) is their lack of ability to adapt to human partners, who typically exhibit an immense diversity of behaviors. We present an autonomous framework as a cobot's real-time decision-making mechanism to anticipate a variety of human characteristics and behaviors, including human errors, toward a personalized collaboration. Our framework handles such behaviors in two levels: (1) short-term human behaviors are adapted through our novel Anticipatory Partially Observable Markov Decision Process (A-POMDP) models, covering a human's changing intent (motivation), availability, and capability; (2) long-term changing human characteristics are adapted by our novel Adaptive Bayesian Policy Selection (ABPS) mechanism that selects a short-term decision model, e.g., an A-POMDP, according to an estimate of a human's workplace characteristics, such as her expertise and collaboration preferences. To design and evaluate our framework over a diversity of human behaviors, we propose a pipeline where we first train and rigorously test the framework in simulation over novel human models. Then, we deploy and evaluate it on our novel physical experiment setup that induces cognitive load on humans to observe their dynamic behaviors, including their mistakes, and their changing characteristics such as their expertise. We conduct user studies and show that our framework effectively collaborates non-stop for hours and adapts to various changing human behaviors and characteristics in real-time. That increases the efficiency and naturalness of the collaboration with a higher perceived collaboration, positive teammate traits, and human trust. We believe that such an extended human-adaptation is a key to the long-term use of cobots.

CCS Concepts: • **Computing methodologies** → **Robotic planning**; *Planning under uncertainty*; *Modeling and simulation*; • **Human-centered computing** → **HCI design and evaluation methods;**

Additional Key Words and Phrases: Collaborative robots, human-robot collaboration, anticipatory decision-making, user studies, evaluating human adaptation

This work is a revised and expanded version of our articles "Social Cobots: Anticipatory Decision-Making for Collaborative Robots Incorporating Unexpected Human Behaviors" published in the Proceedings of the 11th ACM/IEEE International Conference on Human–Robot Interaction (HRI'18) and "Anticipatory Bayesian Policy Selection for Online Adaptation of Collaborative Robots to Unknown Human Types" published in the Proceedings of the 18th International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS'19).
Authors' addresses: O. C. Görür was with DAI-Labor, Technische Universität Berlin when the work was conducted. He is currently affiliated with Micropsi-Industries GmbH, Möckernstrasse 120, 10963 Berlin, Germany; email: cangorur88@gmail.com; B. Rosman, University of the Witwatersrand, 1 Jan Smuts Avenue, Braamfontein 2000, Johannesburg, South Africa; email: benjamin.rosman1@wits.ac.za; F. Sivrikaya, GT-ARC gemeinnützige GmbH, Ernst-Reuter-Platz 7, 10587 Berlin, Germany; email: fikret.sivrikaya@gt-arc.com; S. Albayrak, DAI-Labor, Technische Universität Berlin, Ernst-Reuter-Platz 7, 10587 Berlin, Germany; email: sahin.albayrak@dai-labor.de.



This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike International 4.0 License.

© 2023 Copyright held by the owner/author(s).
2573-9522/2023/05-ART38 $15.00
https://doi.org/10.1145/3585276

ACM Reference format:
O. Can Görür, Benjamin Rosman, Fikret Sivrikaya, and Sahin Albayrak. 2023. FABRIC: A Framework for the Design and Evaluation of Collaborative Robots with Extended Human Adaptation. *ACM Trans. Hum.-Robot Interact.* 12, 3, Article 38 (May 2023), 54 pages.
https://doi.org/10.1145/3585276

1 INTRODUCTION

An efficient **human-robot collaboration** (**HRC**) on production lines involves robots monitoring a human partner's actions and processing them to anticipate the human's plans and goals in a collaboration task [23, 29, 31, 51]. This anticipation requires nonverbal communication to increase the efficiency in industrial environments, e.g., due to hearing limitations [9, 32]. The use of such anticipated knowledge in a robot's decision-making mechanism is dependent on the adaptation of these collaborative robots (cobots) to various types of humans, their dynamic behaviors, needs, and preferences [19, 25, 43]. Our motivation is to ensure such autonomous anticipatory adaptation of robots and their fluent coordination, i.e., well-synchronized coordination of robot behaviors with their human partners through mutual adaptation [22]. We refer to such robots as *social cobots*.

From the existing studies on interactive robots, we deduce that to ensure the long-term usability of collaborative robots, a robot should adapt to both short-term changes for individual differences (e.g., a tough day at work may cause a dynamic human performance) and long-term personal habits, preferences and trust [26, 34, 54]. Hence, we propose a framework that can handle both short-term and long-term robot adaptation. For fluent coordination, we require that short-term adaptation should handle the following settings: *(i)* when the tasks are short or come with changing requirements, e.g., flexible task allocation for mass customization; *(ii)* when the collaboration partner exhibits dynamic behaviors due to, e.g., their changing mental states like intentions and emotions that may cause dynamic human performance and preferences [4, 22, 50]. We refer to the collection of such short-term changing human behaviors that may eventually result in a human error, a task failure, or dropped efficiency in a collaboration task as "unanticipated" human behaviors, as they are unreasonably uncommon to observe compared to expected human performance. These are largely overlooked by existing studies. To address this, we propose a **partially observable Markov decision process** (**POMDP**) that models such unanticipated human behaviors as a latent variable. For example, a human may show signs of distraction and struggle with a task, or may not want the robot's assistance due to, e.g., mistrust. Our POMDP model detects and further reasons about such changes in human behaviors to maintain a fluent and efficient collaboration. The model may anticipate, for instance, that the human is distracted and may need assistance, but the human would not approve the robot's intervention for the specific task. In such a case, the robot should point to remind the task instead of proactively taking over. To achieve this, we do not enforce a turn-taking collaboration but rather allow for flexible collaboration. A robot deploying this POMDP optimizes the plan for improved efficiency and naturalness of the collaboration, ensuring the safety and the autonomy of the human partner. We name our novel approach **Anticipatory Partially-Observable Markov Decision Process** (**A-POMDP**).

Most HRC studies focus on short-term collaboration due to practical reasons, leaving long-term HRC that sustains a rapport and efficient collaboration as a largely unexplored area. Moving from the relevant literature in [4, 26, 33, 34, 50], we consider the following target settings for long-term adaptation: *(i)* when collaboration is required for a repetitive task; *(ii)* when long-term collaborations with different people are expected (e.g., HRC in a factory environment with many work shifts). Our interpretation of long-term adaptation is that a robot needs to distinguish and adapt to

various characteristics of the collaborated humans, i.e., personalization, that eventually affect their short-term behaviors. We call each different combination of such characteristics a unique *human type*. We believe that it is already very difficult to design or learn a single intention-aware model for a person that a robot is collaborating with, let alone for a globally optimal model for different human types, due to the growing complexity of such models. Parallel to this, our early experiments show that learning one decision-making model covering all such diversity leads to less accurate responses with larger delays, causing unreliable real-time interaction with humans. Until the model adapts to a certain human type and changing human behaviors during a task, we observe high frustration in collaborated humans and very low task efficiency, resulting in inconclusive experiments. Our intuition is that rather than a single adaptive model, sometimes a robot may need to follow completely different decision-making strategies, i.e., policies, to enable fast and reliable online adaptation to various human types. For instance, if a robot has two different assistant modes, a human with a lower expertise is more likely to need the robot to act on a task whereas an expert may prefer to be reminded of the instructions. The robot should pick the right policy, accordingly. For this purpose, we present a novel Bayesian policy selection mechanism built on top of existing intention- and situation-aware models (i.e., our A-POMDP models) for an extended adaptation of robots to various and possibly unknown human types. The selection is based on an estimate of a human's long-term workplace characteristics that correlate to the policy performance; hence, we name this mechanism **Adaptive Bayesian Policy Selection** (**ABPS**).

In this article, our first goal is to extend and further improve our previous works on short- and long-term adaptation models and mechanisms, i.e., Görür et al. [16] and [15], respectively, and the second goal is to deploy and evaluate these approaches through user studies on real world scenarios. Our first contribution is a novel framework for collaborative robots that integrates our strategies for short- and long-term extended human adaptation, i.e., A-POMDP models and ABPS mechanism, respectively, into a holistic approach. This is an anticipatory architecture that ensures a robot's fully autonomous human adaptation by integrating the short-term adaptation goals and deploying long-term strategies to regulate and personalize short-term robot behaviors accordingly. That is, a decisional level categorizes a human's characteristics as her level of *expertise, stamina (or fatigue), attention,* and *collaborativeness*,[1] and selects the best short-term strategy, e.g., a A-POMDP model, toward a personalized collaboration. Then, the selected strategy is executed to recognize the human's changing *availability, intent (motivation)*, and *capability* as observations, and to estimate further if the human wants/needs help to guide robot behaviors accordingly. The real-time execution of the selected strategy and sensing and actuating skills of a robot take place in the functional level of the architecture.

Our second goal is to train and validate this framework; however, benchmarking interactive robots, in general, is a difficult task due to the lack of availability of the whole range of human behavior dynamics during their training and evaluation process [57]. As a result, there exists almost no HRC research that considers a vast range of human intentions and actions to the best of our knowledge. Subsequently, such robots show very limited adaptation capabilities as they face a greater diversity of previously unanticipated human behaviors when they are deployed in the wild [21, 55]. In a recent survey on shared autonomy, the authors conclude that the evaluation of interactive robots is highly subjective and the approaches require benchmarking methods and validations in simulation and/or real-world scenarios to speed up their deployment in human environments [51]. As our second contribution, we address this issue by devising a pipeline for

[1]We use this term to indicate the willingness of a human to collaborate that may change due to, for example, task-relevant distrust of the human to the robot.

systematically benchmarking collaborative robots. The pipeline offers a way of training any such system toward various adaptation goals and to test them under these and many more diverse conditions. Similar to the development cycles of autonomous systems, our pipeline rigorously tests the framework first in simulation with accurate human models to train the model with a greater diversity of human behaviors. Then, it deploys the framework in a real environment through user studies to calibrate the framework and to validate it. The calibration phase is similar to any deployment phase existing HRC studies use. It consists of framework adjustments to ensure reliable physical manipulation on the specific application setup and an additional training phase for the fine-tuning of the decision models. The latter step is crucial to accommodate possible differences in the application environment into the models, e.g., various task lengths and operational delays may impact the state transitions in a POMDP running in real-time. Since the pipeline follows a continuous development and integration practice, both of these processes are repeated and continuously provide feedback while the collaborative robot encounters a vast range of short- and long-term dynamics of human behaviors.

We follow the pipeline to evaluate our framework's extended adaptation goals. In order to exemplify our framework, we focus on a nonverbal HRC scenario at a conveyor belt for the task of inspecting and storing various products. We devise a novel collaboration setup that provides a rather unconstrained human intention space by running a cognitively exhaustive task and by not enforcing a turn-taking collaboration, unlike most HRC experiments [10, 22]. The task allows the robot to observe various human characteristics, e.g., a defeatist person, and "unanticipated" human behaviors, e.g., a human who tires easily. Moreover, we develop the same scenario in a factory simulation with our novel human decision models as **Markov Decision Processes** (**MDPs**). The models provide reliable responses with a greater diversity through sampling than the ones observed in a real setup. This allows the framework to train on large-scale data and to be rigorously tested under greater uncertainty, providing training data with higher quality. For example, we could observe and train our decision models against a competitive person with bad skills who is constantly rejecting the robot's assistance. The simulation environment, including human and robot decision models, and our integrated system are available as an open source project[2] To the best of our knowledge, this is the first time an anticipatory robot decision-making mechanism has been tested on such a greater diversity of human behaviors and characteristics for a more realistic evaluation of its adaptation skills.

In the remainder of the article, we first give details of our framework in Section 3 and how it incorporates our A-POMDP models for short-term adaptation (in Section 3.2.1) and ABPS mechanism for long-term adaptation (in Section 3.2.2). Then, we describe our continuous development and integration pipeline for human-aware systems and show how we use it to evaluate our framework in Section 4. This is followed by our novel experiment setup and the collaboration scenario, detailed along with our system's real-time human interaction, in Section 5. Finally in Section 6, we provide our results by first demonstrating that our collaboration experiments are able to create a cognitive load on humans and that we observe "unanticipated" human behaviors. Then, we validate that our A-POMDP model design provides a more efficient and natural collaboration compared to anticipatory models that do not cover "unanticipated" human behaviors. In the final experiments, we show that the complete framework, handling such human variability, provides a fast and reliable adaptation to both short- and long-term changing human behaviors, while being perceived to have high collaboration skills, positive teammate traits and trust. To the best of our knowledge, this is the first time a systematic approach has been applied to developing and testing social collaborative robots, that takes into account such a large diversity of human behaviors and characteristics.

[2]https://github.com/cangorur/human_robot_collaboration.

2 RELATED WORK

Existing intention-aware planning approaches mostly introduce human intentions as a latent variable in a decision-making model, such as in POMDPs [8, 11, 44]. Such a modeling scheme has proven to be efficient in HRC scenarios; however, this scales poorly with an increasing number of human intentions modeled. Hence, such a design conventionally has to limit the human intention space and systematic errors a human can make [20]. Therefore, the studies implicitly make the assumption that either a human's intention (or goal) is constant or it is changing in a limited intention space known to the robot [1, 5, 12, 17, 24, 25, 38, 45]. Moreover, they further assume that humans accept the robot's assistance when offered. It has been stated that such assumptions limit a robot's anticipation of a human's changing behaviors and goals that are mostly observed over the course of a repeated collaboration [3, 34].

Contingencies in human actions have been partly considered [19, 30]; however, all observed actions are still assumed to be toward fulfilling a task, possibly in a way that differs from the expected plan. In a repeated HRC over some tedious tasks, it is more likely that the human performs behaviors that are not even related to the task itself but implicitly affects her performance, e.g., due to fatigue [27]. The robots should be aware of and adapt to such unanticipated behaviors of humans, which to the best of our knowledge, represents a largely unexplored area of research. For this purpose, we propose our A-POMDP mechanism as an alternative to the existing high-level human intention-aware decision models. It removes the assumptions mentioned here, and additionally incorporates such unanticipated human behaviors, e.g., getting tired, not wanting the robot's assistance. With such an extended short-term adaptation, it is also able to coordinate with humans on changing plans and roles for the collaboration task.

Long-term adaptation of robots is still largely unexplored in the HRI domain due to practical reasons [26]. In the long-term, humans exert even more variety of behaviors that require further strategies for the robots to adapt. In Nikolaidis et al. [43], humans are clustered from observations during a training phase into a finite number of human types. The estimated human type is again used as a latent variable in a **mixed observability MDP** (**MOMDP**) model; hence, the authors limit the number of different types due to the increasing complexity with the models. Similarly, Nemlekar et al. [40] show the effectiveness of proactive robot actions based on a prediction of human collaborator's preferences; however, humans are again clustered into a limited number of task-relevant preferences. It has been recently stated that POMDPs require accurate system models that are often unavailable or fail to adapt to various conditions in long-term missions [37]. Bandyopadhyay et al. [6] build several such MDPs with varying reward and transition functions to handle different tasks. In other words, the robots are given the ability to explore different policies and tradeoff between interaction and task quality. However, the study is limited to analyzing different policies to govern such varieties in humans, leaving out the autonomous selection of an optimal one. Our approach to long-term adaptation brings together the idea of generating many such reliable Markov models in Bandyopadhyay et al. [6] to construct a policy library, and the idea of estimating human types on a meta-level as a complementary solution to the intention-aware models in Nikolaidis et al. [43], and goes beyond them to offer a fast and reliable policy selection mechanism as part of a closed-loop robot system, covering a greater diversity of human behaviors and intentions.

For policy selection in the context of social agents, the exploration factor would be very dangerous and frustrating for a human collaborator in the real world; for instance, when using a **contextual multi-arm bandit** (**CMAB**) in an assistant selection for collaborated humans [37]. In addition, in policy or reward learning algorithms, the learning rate is very difficult to tune and the response time is relatively high for any interaction in real-time [47]. There are studies that

efficiently handle the online adaptation of robots to task-relevant preferences of human collaborators using interactive learning [1, 39]. However, when human workers have their shift changes or when a human drastically exerts different behaviors and preferences that may also be irrelevant to a collaboration task (e.g., loss of attention), learning a new policy would take time which is very costly in collaboration scenarios. In such cases, it is better to reuse an already trained reliable model rather than training a new one. In our solution to the long-term adaptation, the ABPS mechanism, we have incorporated **Bayesian Policy Reuse** (**BPR**) in Rosman et al. [49], which has been shown to perform faster and more reliable in online adaptation tasks with greater uncertainty. Instead of modeling human types as a latent variable, ABPS complements existing intention-aware planning solutions, e.g., our A-POMDP, by selecting a policy from a library of such models, each of which already handles the short-term adaptation, based on an estimate of a human's long-term workplace characteristics. This allows us to handle more human diversity in a more computationally efficient way, and to handle unknown human types, while mitigating the need to learn response policies on the fly. Even though the ABPS is agnostic to any labels of human types and robot policies, we generalize some characteristic features of humans in workplaces, inspired from previous studies [13, 27, 37], that are crucial for a collaborative robot to know. These are a human's *expertise, attention, stamina-level,* and *collaborativeness* and they are used to describe a human type.

Regarding the evaluation of adaptation, benchmarking interactive robots is very difficult due to the lack of availability of the complete dynamics of human behaviors. It is hard to expect humans in real user studies to evaluate the system objectively and to convey diverse behaviors, for example, due to the novelty effect they face or due to the constrained environments [34, 51]. Li et al. [35] offer an adaptive policy selection mechanism to pick an optimal strategy against estimated human types; however, the system is trained and evaluated on a constrained task in real user studies with limited and known human actions and intentions. This leaves the reliable adaptation of the framework to diverse and unconstrained human intentions on a more complex task unknown. That said, we believe that there is a lack of human simulations that the anticipatory robots can use as ground truth to train and test on human diversity. Hence, we devise a simulation environment with crafted human models sampling this diversity. The tested solutions can then be brought to the real user-studies, providing more effective adaptation to humans. However, it is an open task to transfer the results from such a simulation into a real-world experiment and validate it by means of user studies. Our intuition is that an evaluation setup should emulate real conditions instead of constraining human intentions so that we observe more dynamics, including unanticipated human behaviors. Additionally, most of the existing HRC solutions are structured around command and response patterns or turn-taking with previously set roles, which limits the fluency, i.e., a key to a satisfying collaboration [22]. We believe that a collaboration setup also needs to enable flexible planning instead of preassigned plans and roles so that the collaborating partners can compensate for possible unexpected situations. For that, following the existing research on human work environments, e.g., [20], we design an experiment setup that induces a cognitive load on humans to invoke a larger diversity of human intentions and behaviors. Then, our robot is faced with greater uncertainty and should naturally coordinate with the human to compensate for possible human errors and to contribute positively to the task.

3 ANTICIPATORY DECISION-MAKING FRAMEWORK

Our framework is designed as a human-aware system to anticipate human behaviors and respond with appropriate robot decisions. It models humans with their long-term characteristics and short-term changing behaviors and utilizes these models to generate high-level strategies and regulate the low-level functional processes of a robot. We deploy this framework to improve a collaborative robot's long-term assistance during an extended collaboration with humans. We

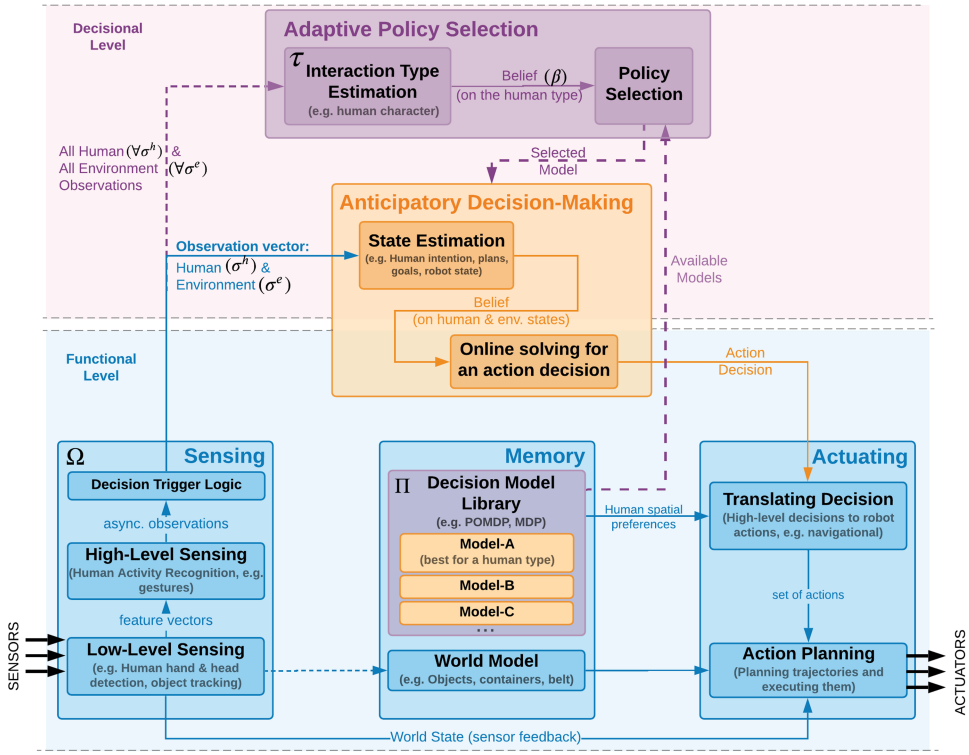


Fig. 1. Our **Framework for Anticipatory Behaviors in Robots toward Interactive human Collaborations** (**FABRIC**) provides improved short- and long-term adaptation to humans and the environment for a robot's personalized and safe collaboration.

name the architecture FABRIC, as a "Framework for Anticipatory Behaviors in Robots toward Interactive human Collaborations" (shown in Figure 1). It is a fully autonomous lightweight system with human-in-the-loop decision-making in real-time. Our application domain is HRC; however, broader use of FABRIC is possible for personalized robot assistants adapting to cared-for individuals, for example, social robots as home assistants.

A common practice in developing robotic system architectures is to consider autonomy in two levels, i.e., functional and decisional autonomy [2]. It is also stated that self-regulated learning is crucial in open-ended tasks, which is the case for cobots in their long-term operations, and it requires both the functional process for sensing, acting, and constructing knowledge and the decisional process for monitoring and regulating the learned knowledge, i.e., high-level reasoning [14, 53]. Moving from this, we design FABRIC with two main parts: *Decisional level* and *functional level*. The *functional level* is mostly reactive and it integrates sensory-motor skills of a robot that is still able to function without the *decisional level*. It comprises of the *sensing*, *memory*, and *actuating* components. The framework recognizes external stimuli (mostly human observations) and actuates action decisions for executing plans in the form of robot actions. The *decisional level*, on the other hand, further processes the observations to infer the characteristics of the human (long-term behaviors), current human mental state (short-term behaviors), and the state of the environment, and generates high-level decisions in the form of goals, plans, and actions. It continuously evaluates the success of the current strategy and, if necessary, interrupts to change it;

however, in general, it is a non-blocking process for the *functional level* that runs in real-time. Our short- and long-term adaptation goals for the cobot are realized in the *decisional level* through the *anticipatory decision-making* (in Section 3.2.1) and the *adaptive policy selection* (in Section 3.2.2) components, respectively. where the former acts as a mediator between the two levels as shown in Figure 1. The following sections detail the framework by running the collaboration task we used in our experiments as an example. That is a robot and a human work on the same conveyor belt to pick and place the objects based on their colors.

3.1 Functional Level

The functionality of this level starts with the *sensing* component that consists of *low- and high-level sensing* and *decision trigger logic*. *Low-level sensing* maps the environment, detects the objects and the presence of a human, and extracts signals from the human. Afterward, the *high-level sensing* block takes the detected observables, and with the help of recognition algorithms, it semantically describes the interaction environment. Finally, an observation vector is generated and forwarded to the *decisional level* through a *decision trigger logic* to update the decision-making (in Figure 1). The logic synchronizes various observations and integrates rules to dynamically call for a decision update (instead of a constant frequency). This is to be able to respond reliably to irregular timings of human behaviors so that the application of FABRIC is not limited to turn-taking collaboration (our applied logic is detailed in Section 5.2.2). In our case, we use a **human activity recognition** (**HAR**) algorithm to recognize a human's hand and head gestures. A generated observation would be, e.g., the human grasped the red-colored object. If the next observation is that the human is looking around, a decision trigger is immediately generated. After a decision is generated from the *decisional level*, the *actuating* component processes it to generate motion commands for the robot actuators. It first semantically describes the abstracted decisions, e.g., the robot arm should point to the conveyor belt as a reminder for the distracted human, as a related action or a set of actions and goals. Finally, the *action planning* mechanism buffers the semantically described actions coming after each new decision and prioritizes them to execute. It plans trajectories and generates motion commands for the robot actuators. It runs a control loop that receives the sensor feedback from the *sensing* component until the actuators reach the goal stated by the current action command. As the design of sensing and actuating components are system- and environment-specific, we detail their implementation in Section 5.

The *memory* component stores the learned models, i.e., the experience, of different contexts for a cobot's future operations. The *world model* component of the *memory*, as shown in Figure 1, contains information on the environment the interaction takes place. The *action planner* retrieves the knowledge of an interaction space for a collision-free trajectory. For example, in our context, the environment model contains information on the location of all the detected static objects like the containers and the conveyor belt in our collaboration setup. Most importantly, the *memory* block stores our novel intention-aware cobot decision-making models in the *decision model library* as different decision strategies, which are encoded as POMDPs and MDPs. The reason for having multiple decision models follows our discussions that rather than a single adaptive decision model, sometimes a cobot may need to follow completely different decision-making strategies, i.e., policies, for a more reliable and faster adaptation, especially when adapting to humans. For that, we design several A-POMDP models with distinct intrinsic parameters (in Section 3.2.1). Each of these models generates a unique set of policies that are stored in the library and may be selected as the best strategy for a unique interaction type, which is assumed initially unknown to the cobot but to be estimated (in Section 3.2.2). In conclusion, the *sensing, actuating, and memory* blocks are crucial for the deployment of our framework to satisfy a cobot's autonomous, reliable, human-in-the-loop, and long-term collaboration.

3.2 Decisional Level: Anticipatory HRC

Our primary novelty lies in the *decisional level* of our framework. The decision-making solutions applied at this level are detailed in the following subsections. Here, we provide the entire anticipatory decision-making process of FABRIC, summarized in Algorithm 1. The algorithm implements a nested loop at its core: The outer loop (starting at *line 3*) iterates between the tasks and runs the *adaptive policy selection* process of FABRIC before a new task starts; the inner loop (starting at *line 6*) iterates during a task and runs the selected policy over the task (i.e., the *decision-making* component of FABRIC). In our application, we select a new policy once at the beginning of each collaboration task. That is, our decision models (A-POMDPs) are designed with a state machine that terminates after the task ends and another policy is selected at the outer loop for the next task. Hence, we refer to the outer loop as *long-term adaptation* since it assumes the context might change from one task to another when compared to changes during a task, i.e., the inner loop, which is rather short-term in our case.

ALGORITHM 1: Anticipatory Decision-Making of FABRIC

Require: (Refering to FABRIC in Figure 1)

- Interaction type space $\mathcal{T}$ (i.e., human types, in our case),
- *Decision model library* of FABRIC with policies Π,
- Observation vector for the observed human behaviors σ^h and the observed environment σ^e in an observation space Ω,
- Observation model to match observables to known interaction types $P(\Omega|\mathcal{T},\Pi)$,
- Utility as accumulated discounted reward obtained from running a policy U,
- Performance model, $P(U|\mathcal{T},\Pi)$
- Exploration heuristics $\mathcal{U}$,
- Number of collaboration tasks K.

1: Train offline for performance and observation models. For that, we run several collaboration tasks between all types in $\mathcal{T}$ and all robot policies in Π.
2: Initialize a belief: β^0 uniform distribution from the prior $\mathcal{T}$. (β is the belief on the current interaction type, i.e., the output of the *interaction type estimation* component of FABRIC.)
3: **for** task IDs $t = 1...K$ **do**
4: The *policy selection* component selects a policy $\pi^t \in \Pi$ using β^{t-1} and performance model, $P(U|\mathcal{T},\Pi)$, using $\mathcal{U}$ in Equation (3).
5: Forward the decision model generated the selected policy, π^t, to the *anticipatory decision-making* component of FABRIC to solve and execute it online over the task.
6: **while** task t continues **do**
7: **wait** for the new σ^h and σ^e from the *sensing* component.
8: The *state estimation* component updates the belief state, i.e., a state is $s \in S$ of our A-POMDP model design in Figure 2, using σ^h and σ^e.
9: Solve the A-POMDP for an action decision. In our case, an action is $a \in A$ of our A-POMDP model design in Figure 2.
10: Forward the action decision to the *actuating* component and execute it on the world.
11: **end while**
12: Obtain all of the collected observation vectors, $\sigma^t = (\forall\sigma^h) + (\forall\sigma^e)$, emitted during the task t.
13: The *interaction type estimation* component updates belief β^t using σ^t by belief update function in Equation (1).
14: **end for**

Starting with the outer loop, a policy is selected according to the current interaction type estimated (*line 4* of Algorithm 1). In FABRIC, the type is an abstract term to represent the long-term

dynamics of collaboration. It identifies the changing requirements or the context of a collaboration task, which may require a different strategy for the cobot. For the type estimation, an interaction type space τ, an observation model to match observables to known types, and a belief distribution β^0 from the prior τ are initialized. The current belief on the type (β^t, where t defines the current task ID) is always updated using the observation model and the collected sets of observations from the previous interactions (at *lines 12 and 13* of Algorithm 1). In our case, τ refers to the unknown type of a human collaborator, defined by the level of her expertise, stamina, collaborativeness, and attention. For example, after a task completion at *line 13*, FABRIC may estimate β^t as a more beginner and pensive human type (in the *adaptive policy selection* component in Figure 1) from the collected observations in the *sensing* component, σ^t, where the human has often placed the objects to the wrong containers (i.e., task failure) and has been detected to be looking around (i.e., distracted). In the next iteration (at *line 4*), the most suitable policy is then selected for the current task from the *decision model library* Π according to the current belief on the type $\beta^t - 1$ and a performance model that matches types to policies. In our case, the cobot may select a policy that more likely assists the collaborating human during a task if her type is estimated to have beginner-level expertise for the task. The entire policy selection process, *the adaptive policy selection* component, is detailed in Section 3.2.2.

After the policy selection process at the outer loop, the selected decision model, e.g., an A-POMDP model in Figure 2 that favors assisting human with the task, is forwarded to the *anticipatory decision-making* component of FABRIC to solve and execute it online during the task. Then, the cobot is ready for the new collaboration task, i.e., the inner loop starting at *line 6* of Algorithm 1. We note that we use an online POMDP solver on the selected A-POMDP model to find the best robot action to execute at each time step (detailed in Section 3.2.1). This provides a faster and more reliable adaptation to the dynamics during the task, e.g., to the human collaborator. The online planning starts at *line 7*. After the *anticipatory decision-making* component receives the observations, at *line 8*, it estimates the current state of the collaborating human, the task, and the environment (i.e., the belief on the current A-POMDP state in Figure 2). According to this estimation, the online solver runs the lookahead search on the A-POMDP model to find the best action decision to execute at each time (at *line 9*). From the same collaboration example with a beginner and pensive human type, the selected policy may estimate the human to have lost her attention during the task at *line 8*. Since the policy favors task efficiency and is aware of the human type, it may decide to act on the task to assist the human at *line 9* (see the possible action decisions in Figure 2). After an action decision is generated, the *actuating* component of FABRIC processes it to generate motion commands for the cobot actuators to realize it in the environment (at *line 10*), which ends one iteration of the inner loop. For example, the cobot picks the object on the conveyor belt and places it in the right container. The next iteration starts at *line 7* after receiving the new observations that are generated in the *sensing* component of FABRIC. Since the human responses are dynamic and their timing is arbitrary, we do not follow a constant frequency to iterate the inner loop. Instead, we design a logic that triggers a new decision for the cobot in response to human actions. This logic (in Figure 9) is detailed in Section 5.2.2. After the task ends, the inner loop ends, another policy selection takes place, and the algorithm waits until a new task starts. The entire decision-making process at the inner loop, i.e., the short-term adaptation, is detailed next.

3.2.1 Anticipatory Decision-making Component. This component implements the short-term adaptation of a cobot, i.e., the inner loop between the *lines 6–11* of the Algorithm 1. The adaptation focuses on unanticipated behaviors of humans and their changing preferences during a collaboration task in real-time. We propose a novel stochastic robot decision-making model, a POMDP, which we call an A-POMDP. This anticipates a human's state of mind in two-stages (see

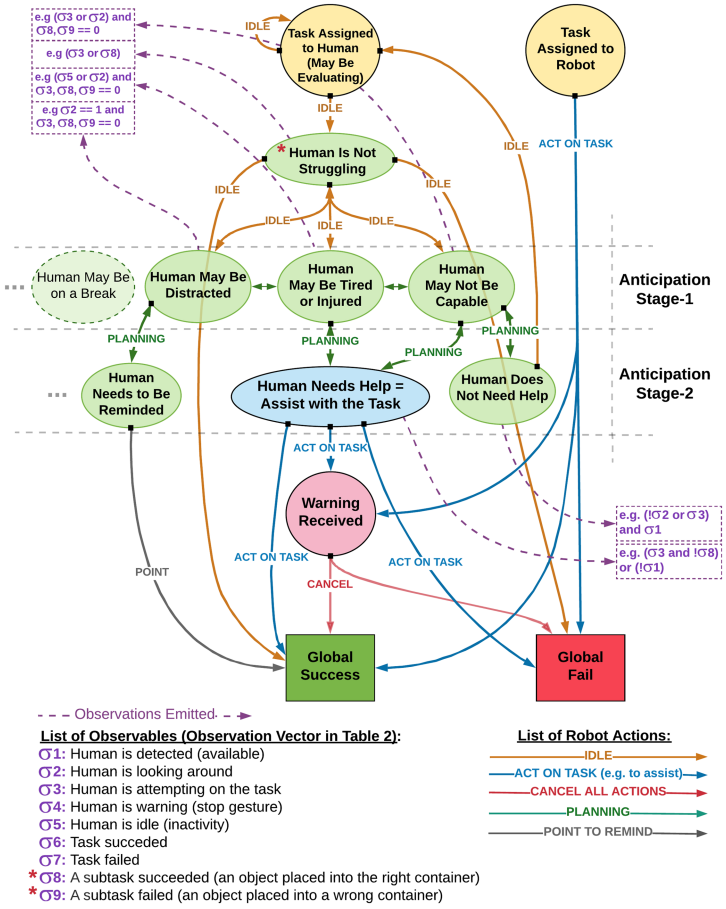


Fig. 2. Anticipatory robot decision-making model, A-POMDP, adapted from [16] for the real user studies and the collaboration environment. * marks the new additions; a new state added to track if human is not struggling with a task and the new observables, i.e., $\sigma 8$ and $\sigma 9$, are introduced to inform about the current subtask status.

in Figure 2). In the first stage, it anticipates the human's task-related changing *availability, intent (motivation),* and *capability* during the collaboration. This includes, but is not limited to, anticipating the human's changing tiredness, attention, and task success (capability) from the observed human actions. In the second, it further reasons about these states to anticipate the human's true need for assistance. It is anticipatory as the need for assistance may become critical or may be requested by the human in the future. The robot generates proactive decisions to effectively assist before the task performance is negatively affected by it. Our contribution lies in the ability of our model to handle these unanticipated conditions: (1) when the human's intention is estimated to be irrelevant to the assigned task and may be unknown to the cobot, e.g., motivation is getting lost, another assignment is received, the onset of tiredness; (2) when the human's intention is relevant but the human may not want the cobot's assistance in the given context, e.g., because of the human's changing emotional states or the human's task-relevant distrust for the cobot.

We have previously implemented a basic version of the model and evaluated it in simulation to show that integrating this model into a cobot's decision-making process to handle unanticipated

human behaviors increases the efficiency and naturalness of the collaboration [16]. In this article, we improve the model design to ensure its real-time interaction capability with real humans in a more realistic collaboration scenario. These improvements are mostly through the insights we obtain and the data collected from the previous simulation tests (see our real-world integration pipeline in Section 4). The biggest challenge in designing such a decision model in a real-world setting is the greater uncertainty in the human states, their changes (i.e., state transitions), and the timing of these transitions. Since our scenarios are not limited to turn-taking collaboration and the environment is rather unconstrained, a human collaborator is left with flexible strategy-making and arbitrary behaviors. In a real-world setting, the time to finish executing an action for humans, including thinking, resting, observing, and picking and placing, would greatly differ from one person to another, and even for an individual during multiple tasks running consecutively for a long-time. As a result, our decision models cannot follow a constant update frequency for the decision-making and it should be aware that the transition probabilities may greatly differ from one collaboration to another.

Our A-POMDP model is defined as a tuple $\{S, A, T, R, \Omega, O, \gamma\}$. S comprises the mental states of the human collaborator from the cobot's perspective, the global success and failure states that define the result of a task (terminal states), the states of a new task assigned to the human or to the cobot (initial states), and a state when the cobot receives a warning from the human for any reason; A is the set of cobot actions and Ω is the set of cobot observations with the observation vector $\sigma \in \Omega$ from the human and the environment (listed in Figure 2); T is the state transition probabilities where $T(s'|s, a)$ gives the probability of a transition from state $s \in S$ to $s' \in S$ for a given robot action $a \in A$; O is the observation probabilities where $O(\sigma|s', a)$ gives the probability of observing σ in a new state s' after taking action a; R is the immediate reward the robot receives and $R(s, a)$ gives the expected reward for taking action a while in state s; γ is the discount factor for delayed rewards. We solve the POMDP model for an optimal robot policy, π. We generalize the states and the robot actions to comply with several collaboration tasks. The cobot actions are to wait for the human (*idle*), plan for assisting action (*planning*), and *assist* the human, e.g., by directly acting on the task. Further details of our design strategies are given in [16]. In the previous work, our goal was to prove whether our models could cover the unanticipated human states, and show the importance of doing so. In this article, our goal is to extend the approach to more realistic real-world scenarios.

For that, we consider a collaboration task with multiple steps (in Section 5.2.1), which we call subtasks, as opposed to each successful attempt completing a task in our previous work. This suggests longer iteration in our POMDP state machines before reaching a terminal state (e.g., *Global Success*). To better track the state of a human collaborator during such a task, we introduce new states and observations to our A-POMDP tuple (see the red marks in Figure 2). A partially observable state called *Human Is Not Struggling* is added to indicate that the human is successfully achieving the subtasks (i.e., each successful placement of an object in our case), and the cobot most likely estimates and stays at this belief state unless the human repeatedly executes any unanticipated behaviors. In this case, the cobot may transition to any state in the *anticipation stage-1*. The new model also favors a transition from this stage back to the *Human Is Not Struggling* state if the human could keep up with the task again. Finally, the newly added observables reflect the success status of a subtask, as detailed in Section 5.2.2.

The A-POMDP receives a negative reward R if the model reaches the *Warning Received* state, that is when the cobot receives a warning from its human partner for any reason. This mostly happens when the cobot misjudges the human's need for assistance and wrongly takes over a task. Other than that, the immediate rewards are assigned for the terminal states of *Global Success* and *Global Fail*. Finally, we set a discount factor to avoid longer wait times before an object placement takes

place. This encourages a cobot takeover if the human delays. We set a cooperative reward for the human-robot team to evaluate the performance of the collaboration during a task. Whenever the system detects a subtask success or a subtask failure, an immediate cooperative reward is assigned. It should be noted that the same rewards are assigned after a subtask success or failure regardless of who completes it.

We use an updated version of DESPOT solver in [58] to solve for and generate a policy online from our A-POMDP models. Our version is also capable of executing the policies in real-time against real humans. The online solver constructs a belief tree with the belief on the current state at the root of the tree and performs a lookahead search on the tree for a policy [58]. This alleviates the computational complexity by computing good local policies at each decision step, which also allows for an online adaptation to the uncertainties in the environment or in the collaborating human's behaviors. The setting of the transition and observation probabilities are realized in multiple steps that involve simulation runs and training with real humans (following our continuous integration pipeline in Figure 3, detailed in Section 5.4).

3.2.2 Adaptive Policy Selection Component. Our goal is to extend the adaptation of a single intention-aware robot decision-making model, such as A-POMDP, to various and changing human characteristics in long-term for personalized collaboration. Here, we present a novel policy selection mechanism, which we call ABPS, that builds on top of existing intention- and situation-aware robot decision-making models for an extended adaptation of cobots to various human types. It handles the long-term adaptation of FABRIC, satisfying a personalized collaboration for a cobot. We have previously implemented a basic version of ABPS and evaluated it in simulation to show that such a mechanism extends a cobot's human adaptation and provides a more efficient and natural collaboration in long-term [15]. In this article, we improve ABPS to ensure its real human collaboration and its system integration in a more realistic scenario. ABPS is equipped with a policy library (i.e., the *decision model library* of FABRIC in Figure 1) to act appropriately in the context of some human types and tasks in the HRC domain. The selection process is handled in the *adaptive policy selection* component of FABRIC. It is presented with an unknown interaction type, which must be solved within a limited time and small number of trials. As mentioned, we define an interaction type in our case as a human collaborator having an unknown type (i.e., characteristics) in a known task. The goal of ABPS is to select policies from the library for the new and possibly unknown human type, over which it has a belief distribution, while minimizing the total regret in a limited time. Minimizing the regret in our domain is defined as increasing the task success rate and decreasing the number of warnings received from the human collaborator, relative to the best alternative from the library in hindsight.

At the core lies the *decision model library* Π that holds all robot decision models. Our goal is to limit the arbitrary generation of the policies to avoid overloading the library with unreliable candidates [3]. For that, we generate new models moving from a base model design, which is an A-POMDP model with the connection scheme as in Figure 2. The intrinsic parameters of the base model are trained after the calibration experiments described in Section 5.4. The offline generation of different robot models is done by adjusting T and O of the base model: the state and observation probabilities of the A-POMDP model corresponding to different human types (see Section 3.2.1). Changes in T correspond to different transitions of a human's internal states, e.g., a model assumes the human tires faster, or the human needs more assistance when she is not capable. Changes in O then define the observations emitted by the human as a function of her internal states. For example, a human not being able to handle the task could indicate that she is tired, or she is an inexperienced, both of which should be handled differently by the cobot. Additionally, by adjusting O, we are able to make the model a partially, mixed or fully observable

Markov decision process (POMDP, MOMDP, or MDP, respectively). We randomly adjust the probabilities to generate various models, each of which handles a unique human type, and solve for their optimal policies to construct our policy library Π. Through this random generation, we are agnostic to specific human types and behaviors, which also allows ABPS to integrate into any existing intention-aware model. We are aware that this random generation may still produce unreliable policies for the cobot. Nonetheless, reliability and usability of such policies become prominent after the training process (in Section 5.4).

ABPS measures the similarity between an unknown interaction type and previously known types, to identify which policies may be the best to reuse (in the *interaction type estimation* component in Figure 1. In this case, a collaborated human's type is latent and the human type space is not fully known. Therefore, a correlation between policies and a bounded set of human types is not possible. In fact, the space of human types is in general infinite, but we limit this to control complexity. Therefore, the construction of a type space τ is a crucial process. For this purpose, we train an estimation model from a set of known types and use it online to estimate a new unknown type as a belief distribution over the known ones, $\beta(.)$ (see Section 5.4). In order to train such a model, we generalize some characteristic human features to approximate a human type. These are a human's changing levels of expertise, attention, stamina, and collaborativeness. These features are not exhaustive but they are inspired by [13, 27, 37] and are stated to be crucial to be known by a cobot. The type space consists of many human types by adjusting the level of these features, e.g., a human with beginner skills, pensive, low stamina and non-collaborative behaviors (e.g., always rejecting a robot's assistance due to distrust). We argue that any human worker can be represented as a distribution of such features in our experiments. The interaction type estimation model is used by ABPS as *a priori* information, which we call the *observation model*.

Definition 1 (Observation Model). For a robot policy π, an interaction (i.e., human) type τ and an observation vector σ obtained from the human actions and the environment, the observation model $P(\sigma|\tau, \pi)$ is a probability distribution over the observation signals $\sigma \in \Omega$ that results by applying the policy π to the type τ.

All combinations of known human types in τ and the robot policies in the library are run against each other offline several times to generate our *observation model* (in *line 1* of Algorithm 1, detailed in Section 5.4). The observation signals are emitted by the collaborated human and the environment, reflecting a human's actions and their impact on the task and the environment. In our experiments, an observation vector, $\sigma \in \Omega$, is a 9-D boolean vector as listed in Figure 2 and detailed in Table 2. The ABPS agent receives these observables at every episode of a task and accumulates them to update its belief on the human type after a task terminates (see *line 12, 13* of Algorithm 1). Finally, the type belief update is Bayesian, given by

$$\beta^t(\tau) = \frac{P(\sigma^t|\tau, \pi^t)\beta^{t-1}(\tau)}{\sum_{\tau' \in \mathcal{T}} P(\sigma^t|\tau', \pi^t)\beta^{t-1}(\tau')}, \quad \forall \tau \in \mathcal{T}, \tag{1}$$

where β^{t-1} is the previous belief and $P(\sigma^t|\tau, \pi^t)$ is the probability of observing σ^t after applying π^t in an interaction with any human type τ. This distribution is directly retrieved from the *observation model* for each requested type and policy.

The policy selection process of the robot is based on an exploration heuristic called ***expected improvement (EI)*** [49]. As stated in *line 4* of Algorithm 1, this algorithm runs on another trained *a priori* model called the *performance model*.

Definition 2 (Performance Model). The performance model, $P(U|\tau, \pi)$, is a probability distribution over the utility, U, of a policy π when applied to interaction (i.e., human) type $\tau \in \mathcal{T}$.

The system utility, U, is the accumulated discounted reward received after a policy is run (see Section 3.2.1 for the immediate rewards a robot obtains during a task). All the combinations of known human types $\tau \in \mathcal{T}$ and the robot policies $\pi \in \Pi$ are repeatedly run against each other offline to generate our *performance model* (see Section 5.4). Then, this model is used by the policy selection heuristic. The heuristic assumes that there is a U^+ in reward space which is larger than the best estimate under the current type belief, U^β. A probability improvement algorithm can be defined to choose the policy that maximizes Equation (2) and achieves the utility U^+.

$$\pi' = \arg\max_{\pi \in \Pi} \sum_{\tau \in \mathcal{T}} \beta(\tau) P(U^+ | \tau, \pi). \tag{2}$$

Because the choice of U^+ directly affects the performance of the exploration, its selection is crucial to the performance of this exploration. The *EI* approach instead addresses this nontrivial selection of U^+. The algorithm iterates through all the possible improvements on an existing U^β of the current belief, which satisfies $U^\beta < U^+ < U^{max}$. The policy with the best potential is then chosen, as given in Equation (3).

$$\pi' = \arg\max_{\pi \in \Pi} \int_{U^\beta}^{U^{max}} \sum_{\tau \in \mathcal{T}} \beta(\tau) P(U^+ | \tau, \pi) dU^+, \tag{3}$$

$$= \arg\max_{\pi \in \Pi} \sum_{\tau \in \mathcal{T}} \beta(\tau)(1 - F(U^\beta | \tau, \pi)), \tag{4}$$

where $F(U^\beta | \tau, \pi) = \int_{-\infty}^{U^\beta} P(u | \tau, \pi) du$ is the cumulative distribution function of U^β for a τ and π. The algorithm, therefore, selects the robot policy with the most likely improvement on the expected utility. Finally, once a policy is selected, the A-POMDP model that has generated it is forwarded to the *anticipatory decision-making* component of FABRIC to be executed online (see *line 5* of Algorithm 1). The process is repeated until all the scheduled tasks end.

4 A PIPELINE FOR CONTINUOUS DEVELOPMENT AND INTEGRATION OF COBOTS

When deployed for interaction with humans, robots face a great deal of uncertainty in the long-term. This is majorly due to the lack of diversity of human behaviors available during the training and validation processes of the interactive robots. To decrease the uncertainty, we follow the conventional way of developing autonomous systems, i.e., rigorously testing a system in simulation, and then deploying and validating it in a real environment. To do so with interactive robots, we propose a pipeline in Figure 3, to systematically train and evaluate a robotic framework with a vast range of short- and long-term dynamics of humans. The pipeline takes an interactive robot framework, runs it through simulations, designs a dynamic and unconstrained environment, deploys it for user studies and follows a continuous development and integration practice by iteratively training on various human behaviors.

First, a simulation environment with simulated humans running realistic decision models is needed to start the training and testing processes of an interactive robotic framework (*step 2* of the pipeline in Figure 3). Interactive and collaborative robots mostly rely on the data collected from people participating in user studies, which are mostly in a confined space with limited observable human intentions, e.g., due to constrained environments. This makes it challenging to reuse and benchmark such robotic solutions [46]. Through sampling methods, e.g., Monte Carlo, running on simulated human models, we can generate a diversity of human behaviors including the unanticipated ones that are hard to observe in lab environments (see our simulation in Section 5.1.1). In *step 3* of the pipeline, the simulations enable a robot framework to train on this large-scale human

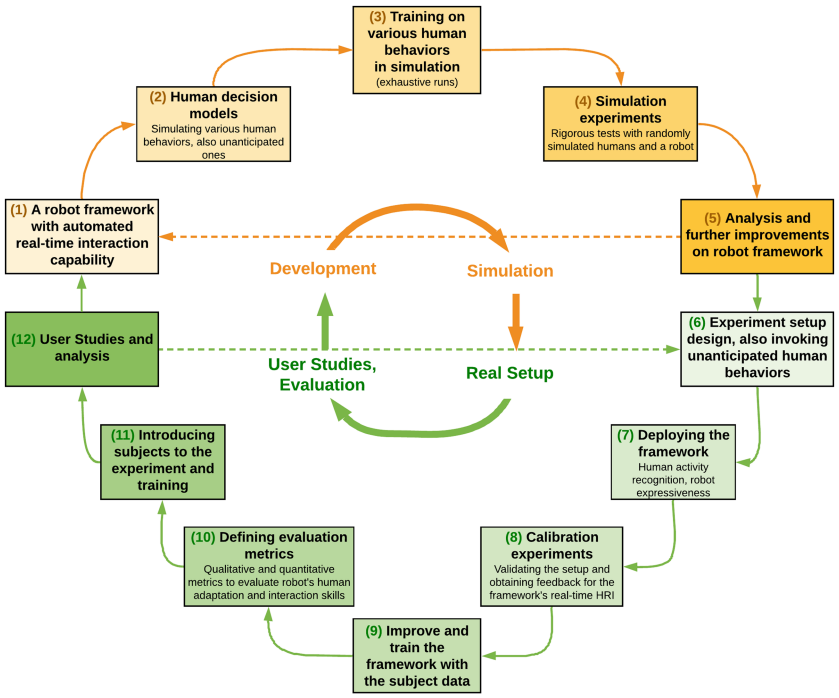


Fig. 3. Our pipeline for integration, deployment, and evaluation of an anticipatory collaborative robot with extended human adaptation. The development, deployment, simulation experiments, and user studies are iterated while the framework trains on and is tested against a greater diversity of human behaviors.

data. Then, in *step 4*, the same simulation environment is used to rigorously test and evaluate the framework.

Even though we use the same simulation environment for training and testing, the randomly generated human models and the sampling of human behaviors at every run avoids overfitting of the trained models and provides enough diversity for their evaluation. Also, we highlight the importance of designing a 3D simulation environment with real-world physics to introduce changing dynamics into the manipulation tasks. e.g., dynamic performance of a simulated human with the changing weight of a package, unknown delays due to the mechanics of the conveyor belt, dynamic HAR accuracy due to various lighting or occlusions, and dynamic timing and strategy for a human to complete an action. In general, we expect epistemic and aleatoric uncertainties in the robot decision models due to the lack of training data that incorporates sensor noise, mechanical drifts and delays, and a variety of human intentions and responses. This greater uncertainty and diversity provides higher quality data for the construction of the policy library. It also provides feedback during the evaluations, which is both qualitative with observations from the simulations and quantitative like collected rewards for a robot's decision-making. In *step 5* of the pipeline, these are analyzed to identify and handle unnatural (e.g., unreliable or unexpected) or inefficient robot interactions before the framework's real-world deployment. We would then repeat *steps 3–5* of the pipeline. The details of our simulation environment are in Section 5.1.

For validation, a user study experiment should be realistic in the sense that human participants should not be confined to a structured environment with limited interactions. However, it is also difficult to have humans in real user studies convey a full range of possible diverse behaviors.

Step 6 of the pipeline, therefore, focuses on the open task of transferring the results from the simulation targeting a wide range of human behaviors into a real-world experiment and validate it by means of user studies. Additionally, most of the existing cobot solutions are structured around command and response patterns or turn-taking with previously set roles, which limits the fluency of a collaboration [22]. Hence, our design goals in *step 6* are twofold: (1) A setup should foster unanticipated and uncommon human behaviors; (2) It should let a human and the cobot freely act on the task and flexibly replan to compensate for such behaviors. For that, our experiment setup design must place a cognitive load on humans to invoke such human behaviors (detailed in Section 5.2).

Physically interacting with humans in real-time in such an unconstrained environment brings about a number of challenges for an autonomous robot. In HRC, poor coordination between the human and the cobot is likely to result in a collaboration that is neither effective nor perceived as natural. In *step 7*, the pipeline deals with these challenges to ensure a reliable coordination between the cobot and human in real-time. There are a variety of different human characteristics with various preferences; as a result, recognizing human actions and understanding human behaviors show a great diversity due to their unique reaction times and different ways of reflecting the same intentional and emotional behaviors, i.e., hidden intentions. For example, one person may idle too long when evaluating a subtask whereas for another this would mean they need assistance. This is especially the case when an interaction does not follow a turn-taking approach. Thus, different reaction times and strategies need to be considered in the decision models for reliable responses. In addition, effective collaboration means good coordination of intentions and actions [56], and communication plays a decisive role in this. The cobot should then be equipped with intention expressive gestures that effectively communicate its decisions. Hence, in *step 7*, we need to develop: (1) HAR solutions, (2) adaptive decision update frequencies for the robot to respond timely to dynamic human behaviors in real-time, and (3) expressive robot motions to effectively communicate the robot's action decisions to a human.

The rest of the pipeline focuses on systematic user studies. In *step 8*, we conduct calibration experiments with real humans to validate *steps 6* and *7* and to provide feedback for further improvements and training in the next step. In these studies, a first batch of participants interact with the cobot running the framework to evaluate the activities under *step 7* and, in general, the reliability of the cobot's interaction. In the second batch, we evaluate the experiment setup and our design goals to see whether we could invoke the unanticipated human behaviors by comparatively testing several interaction scenarios through within-subject experiments. These experiments are briefly detailed in Section 5.3. In *step 9*, we first process the results of *step 8* to improve the setup and the framework for its real-time interaction capability, then we train the decision-making models with the real human data obtained from *step 8* along with the simulation data.

The next steps are conventional steps in most of the user studies. In *step 10*, the performance metrics need to be defined to evaluate the framework's interaction and human adaptation goals. In general, the participants are asked to evaluate a robot's adaptation skills, whereas in the HRC context, this can also be objectively analyzed from the efficiency metrics defined for the collaboration. In our case, we have designed a warning gesture for the participants to reflect their dissatisfaction with the cobot responses, which also numerically reflects the reliability of the cobot adaptation. Finally, in *step 11*, we introduce the participants to the experiment setup and do training runs to familiarize them with the environment. This way, we avoid the practice effect in our comparative analysis and partially let the novelty effect fade away. We believe that this step is crucial to more realistically evaluate human adaptation skills of a robotic framework. Then, we conduct user studies in *step 12*. The user study and simulation loops are iterated with different participants and various human models until the robot framework performance converges to steady state.



Fig. 4. Simulation of an HRC at a conveyor belt for the task of product inspection and storing.

5 APPLYING OUR FRAMEWORK THROUGH THE PIPELINE

In this section, we train and deploy our framework given in Figure 1 following the pipeline in Figure 3 both to give an example of how the pipeline works and to evaluate and validate the framework and our discussions on short- and long-term human adaptation of cobots. As in the pipeline, we divide this section into simulation and real setup applications. A figure summarizing and listing all of our activities following the pipeline is given in the appendix.

5.1 Simulation Environment

This section follows *steps 2–5* of the pipeline in Figure 3. For this purpose, we devise a 3D simulation environment to simulate our HRC scenario in a factory environment on a conveyor belt. This environment has already been used to train and evaluate both *anticipatory decision-making* (i.e., A-POMDP models) and *adaptive policy selection* components (i.e., ABPS) in Görür et al. [2018] and in Görür et al. [2019], respectively. In this section, we detail the environment and our novel simulated human models. The simulation, in Figure 4, consists of existing human and robot models (PR2 is selected but the framework is indifferent to the robot hardware), a conveyor system, produced packages, two containers for processed products, a container for unprocessed products and a restroom. All of our scenarios consist of several sequential task assignments to simulate long-term collaboration. A task in the simulation is a product inspection and storing job. It starts with a user-defined task assignment. A task is successful when the product is inspected and put into the processed-product containers either by the human or by the robot. The conveyor belt waits for a certain time for a package to be processed, and then runs and the product falls into the uninspected-product container leading to a task failure.

Both the robot and the human are controlled by their decision models whereas the generated action decisions are executed in the MORSE simulator. The robot's decisions are generated by our anticipatory decision-making approaches. The observations the robot receives (in Section 5.2.2) are the 3D human body joints that are always available directly from the simulated human model and the proximity sensors placed inside the containers to monitor the task status as succeeded, failed, or ongoing. A state-of-the-art HAR module, inspired by Roitberg et al. [2014], has been implemented to recognize the constrained and distinct simulated human gestures from the body joints available, constructing the *sensing* component of FABRIC. We have designed certain human actions that are required for our scenario, that are also observations for the robot. Based on the available list of actions, we have designed human models that simulate a variety of human decisions. Just like the robot, the human decisions are executed in MORSE, and the human decision

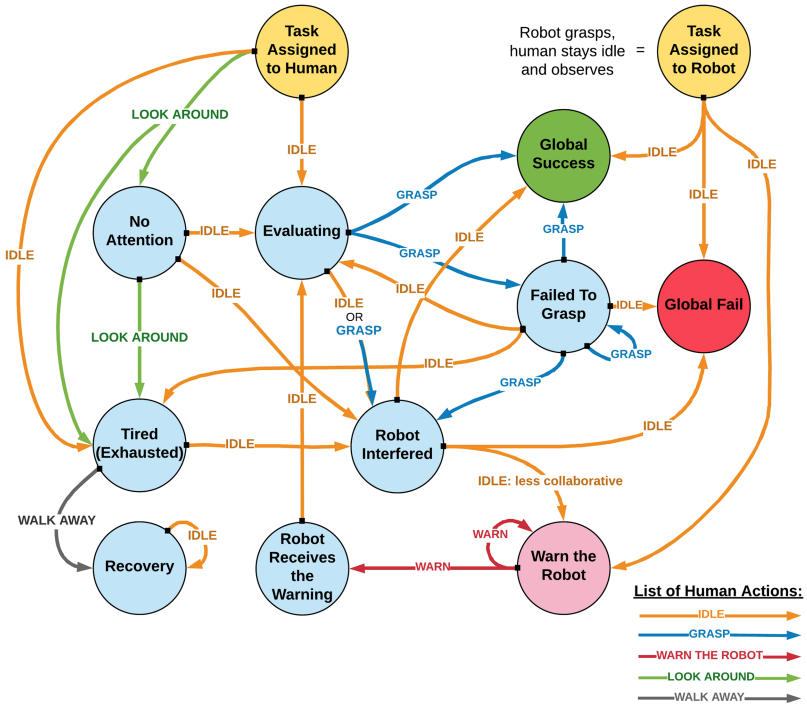


Fig. 5. Human MDP model in detail, showing state-action connections for the most probable transitions.

model receives observations for the current state of the environment (including the robot motions) as feedback. Hence, simulated humans are also automated agents. The timing and the duration of human actions are dynamic with random factors, in order to contribute to the uncertainty of the environment. In summary, this environment allows for a fully automated long-term HRC to train and test robot decision-making solutions under various conditions.

5.1.1 Human Simulation: Human Decision Models. Simulating humans allows us to scale the experiments to emulate many different combinations of human behaviors, including unanticipated ones. Our goal is to create use-cases where a human worker follows the aforementioned unanticipated conditions and occasionally performs behaviors like stubbornly rejecting the robot's help, tiring quickly, being easily distracted, and being distrustful of the robot. A representative proof of concept human decision model is built using an MDP as shown in Figure 5. We note that only the transitions with non-negligible probabilities are shown in the figure. Our model design is inspired by available studies analyzing human workers operating on repeated tedious tasks in a workplace [13, 27, 37]. We assume that a human worker optimizes an objective function to reach her goal. However, following our statement, this may also be an internal goal irrelevant to the assigned task, e.g., leaving their place for a short break. We also assume that any human actions may be imperfect [20]. Simulating such a human has been shown to be accurate using an MDP to generate a policy for a human agent [6].

Our human MDP is a tuple $\{S, A, T, R, \gamma\}$ where S is the human states of mind, A is the human actions, T is the state transition probabilities, γ is the discount factor, and R is the immediate rewards received based on the result of a task and the type of the human to encourage that type of behavior, e.g., a distracted person receives positive rewards in the *Global Success* and *No Attention* states. This model is inspired by our expectation that a human chooses an action based on the

collaborated robot's action, the state of a task, the human's internal states, and her internal goals. Additionally, we govern a human's responsiveness to the interacted robot actions. Such responsiveness is handled through a transition function $T(s, a, s') = P(s'|s, a, n_r, k_t)$ for $s, s' \in S$, $a \in A$, the number of times the robot interfered in a task n_r and the number of tasks handled so far k_t. That means we have dynamic transition probabilities changing over the course of the interactions, leading to updates on human models after each task, and so updated human behaviors. An example of such responsive behaviors is that a human becomes less collaborative as her robot partner selects wrong policies, e.g., the robot takes over a task (depicted as n_r) when the human was already planning to handle it. A decrease in collaborativeness is handled with an increased transition probability of the human to *Warn the Robot* when a robot interferes with a task. Another example is that a transition to the state of being tired depends on the number of tasks already handled, k_t.

The model samples random but goal-oriented dynamics for the human collaborator using **Markov Chain Monte Carlo** (**MCMC**). In the end, the MCMC sampling and the responsive transition function lead to human simulations exerting dynamic behaviors changing in response to the robot decisions and with a small random factor. Through changing T and R and solving the model for a policy, π, we create various human types with changing characteristics (i.e., long-term behaviors). This modeling scheme is then used to automate the training and testing of our cobot's reasoning under various hard-to-predict conditions. We discuss our training process in Section 5.4. Our simulation experiments, following *steps 3–5* of the pipeline, for A-POMDP model (i.e., short-term adaptation) are in [16] and for the integrated anticipatory decision-making with ABPS (i.e., long-term adaptation) are in [15]. We show our analysis on the reliability of our human modeling scheme and the diversity of behaviors toward large scale training and tests in the appendix.

5.2 Real Environment and Setup Design

This section details our activities on *steps 6* and 7 of the pipeline (see Figure 3). We devise a collaboration setup, shown in Figure 6, that provides a rather unconstrained human intention space. Moreover, we design a cognitively exhaustive task of sorting colored cubes continuously flowing on a conveyor belt and placing them into relevant colored containers according to complex rules. Through inducing a cognitive challenge, our goal is to observe various human characteristics, e.g., a competitive person with a bad memory, and invoke unanticipated behaviors, e.g., constantly rejecting the robot's assistance, lost attention and lost motivation. Our goal is to show that the environment is realistic and able to induce a desired cognitive load to invoke such diverse human behaviors. Additionally, we do not enforce turn-taking collaboration in the task, allowing a human and the robot to flexibly replan task allocations and freely act based on their estimate of the partner's behavior. As a result, this allows for evaluating a broader range of adaptation skills leading to the validation of our framework.

5.2.1 Collaboration Scenario and the Task Design. Our collaboration scenario starts after a human participant sits on the chair and wears the yellow gloves and the safety helmet shown in Figure 6. These are used to recognize human activities. The experimenter selects a task and the task rules are displayed on the *task monitor* as shown in Figure 6. Every task consists of several object sortings and placements. The task description is visible only for a certain amount of time and then it disappears as the participant's job is to memorize it (i.e., another cognitive challenge). The selected task starts right after the rules disappear. The conveyor belt then starts to move and transports a wooden cube towards the human and the robot. Our robot arm, we call it the cobot for simplicity, picks the cubes with suction yet for simplicity we also refer it as grasping. A cube is available to be grasped once it stops in front of an infrared sensor on the belt. The participant and the cobot might now decide to grasp the object and place it in one of the containers according to

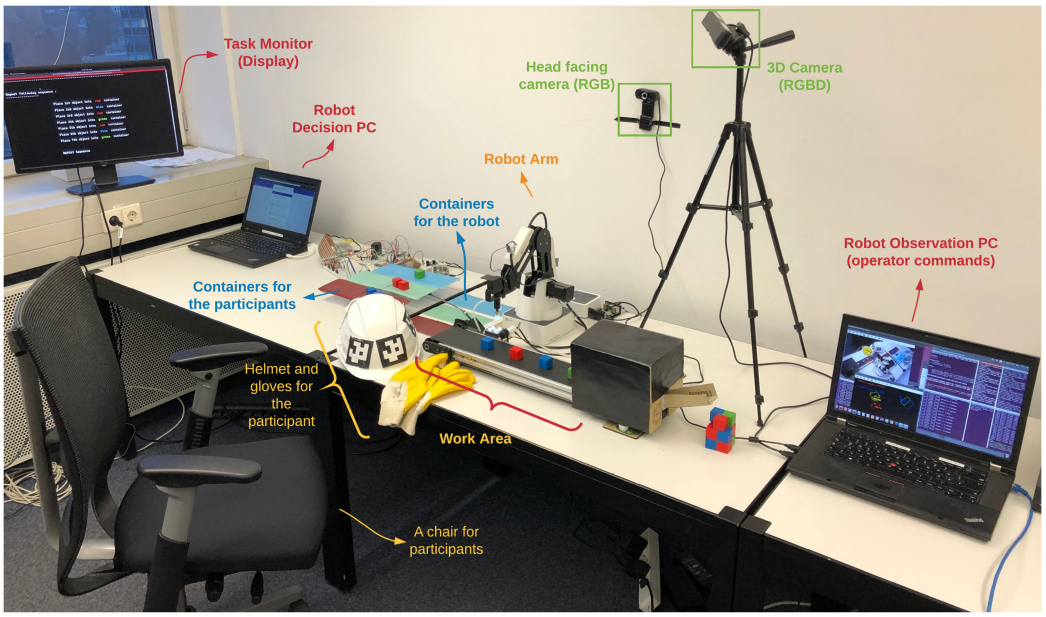


Fig. 6. Real-world setup: HRC on a conveyor belt.

the task description. After each placement, either by a human or by the cobot, the status of which cube is placed on which container is displayed live on the *task monitor* without the information on whether the placement is correct or not. This is only for the participants to keep a track of the task with many cube placements yet introduces a distraction on a continuous task flow. Once a maximum number of cubes are placed in a task, the cobot's decision-making terminates (see *Global Success/Fail* terminal states that are fully observable in A-POMDPs in Figure 2). The task results are then shown on the monitor, which are the success rate, task duration and a score. A score is to motivate the participants to achieve the goals during an experiment, which is detailed later.

Cognitively challenging task design: A task in our experiments allows for a fluent collaboration, where a human and a cobot should take the initiative to change and adapt to task allocations on-the-fly. In simulation, we model a task to be physically demanding, and a cobot would track a human's physical abilities and conditions. In user studies with real people, such a setup would be difficult to prepare in a lab environment. For this purpose, we still focus on a task of pick and place. But for the simplicity, we create a cognitive load for humans instead of a physical one. A cognitive load replaces physical exhaustion and incapability by introducing cognitively demanding memory and coordination exercises. We argue that human states like tiredness, distraction, lost motivation, and so on can also result from a difficult cognitive task, which is an easier and a safer option in a lab environment.

A larger number of placement rules to remember yields a more difficult task since they are only displayed to a participant for a short amount of time. Therefore, a task is cognitively demanding to varying extents according to the complexity of the rules and the amount of time they are displayed. In total, five different task types have been implemented (shown in Figure 7). Each type is inspired from mind games to sort colors with confusing rules, e.g., using Stroop effect [52]. Our goal is to select tasks that are both challenging but with a difficulty still achievable by the people, on average, when they receive no help from a cobot. In this way we ensure a collaboration. Hence, we also evaluate and compare the cognitive loads these tasks induce on people during the calibration



Fig. 7. Displaying task rules for five different task types. The cubes flow on the conveyor with a random order, unknown to the participants.

experiments (in *step 8* of the pipeline), which is summarized in the appendix. In addition to the color rules and their limited display time, after each cube placement (subtask), the participants are informed to wait for audible feedback before they move to the next one. In the scenario, this corresponds to a supervisor check. The timing of the sound is random (usually around 2 seconds); hence, it introduces another cognitive challenge to remembering the rules.

How to induce the unanticipated human behaviors? The collaboration between a human and the cobot is motivated by a scoring system that punishes wrong placements with negative rewards and assigns positive rewards for any correctly sorted objects (subtasks), whether it is by a participant or by the cobot. As the tasks are designed to require high cognitive load from the participants, such a rewarding system is expected to incentivize them to accept help from the cobot, especially when they are struggling with a task. We also have initial task assignments to be able to better evaluate the cobot's assessments over the human's progress when, e.g., the task is assigned to her. In our scenarios, more rewards are received when a subtask is achieved by whomever the task is assigned to. This is to favor the assignee to do the job. The same rewarding mechanism is also used in the cobot decision models, i.e., A-POMDP (in Section 3.2.1). Hence, it is a shared goal of the team to maximize the collaborative score. Since the color of the cubes can be misidentified due to the changing lighting conditions and other classification errors, the cobot may also make a mistake. This is informed to the participants without the success rates of the cobot, which is above 95%, to also evaluate trust. After a task ends, the final score is displayed to the participant to allow her to draw conclusions from it and create strategies for the next task. For instance, she may decide to trust the cobot more, which we categorize as the long-term changing human characteristics.

Table 1. Our Experiment Design Choices to Invoke Unanticipated Human Behaviors

Human Behaviors	Invoked through
Failures	cognitively demanding task rules, rules being visible to a participant for seconds, and audible feedback before moving to the next subtask.
Distraction	the *task monitor* that participants sometimes check to see the current status of a task and do not pay attention to the task itself.
Tiredness	the cognitive load that accumulates through multiple tasks and a long experiment that takes approximately 1.5 hours.
Motivation to work	scoring system that drives competitive behaviors and the training phase in which the participants practice tasks and the collaboration with the cobot before the experiments.
Lost motivation	the same task type repeated several times, the difficulty of a task, and the length of an experiment that all require a constant attention and memory.
Willingness to collaborate with the cobot	the trust in the cobot's success in a task, the scoring system, and the difficulty of an experiment.
Not wanting the cobot to assist (non-collaborative behaviors)	decrease in trust due to unwanted behaviors of the cobot (e.g., misinterpreting a human's assistance need, when the cobot makes a mistake.), that the tasks are initially assigned to the participants where they get more rewards if the person achieves a task. The participants can also warn the cobot to stop, indicating competitive behaviors in humans.

One goal of our cobot is to adapt to a human's changing willingness to collaborate. Therefore, we initially assign all of the subtasks to the participants, and the cobot is there to assist if needed. Our intention is to reduce the situations where a participant lets the cobot permanently take over a task and to create a more dynamic collaboration by giving her the autonomy to decide if the cobot should assist or not. For example, if a human is not sure about a subtask, she could let the cobot take over, thus avoiding a negative reward while sacrificing the larger reward. All of this is told to the participants before the experiments with an analogy from a real factory that an assigned task is better achieved by the assignee for an increased quality. To describe this analogy better, we give an example that the objects to be placed are fragile and so a human would handle the placement better than a cobot. In order to maximize the collaborative reward, we expect this to encourage the participants to take over as many of the placement tasks as possible. This helps us observe more human dynamics during an experiment. Finally, the goal of this environment setup and the scenario is to invoke the unanticipated human conditions in a work environment, such as tiring, losing motivation, failures, and so on that are expected to be observed during monotonous working conditions. In Table 1, we summarize all of our strategies to induce such conditions (referring to Figure 6). The actual effect of these strategies depends on a participant's characteristics (e.g., memory, stubbornness, and competitiveness), the time of the day in which the experiment is conducted and a participant's background (e.g., field of study, previous experience with robots). This gives us more variety and less control, which is a better test environment for evaluating adaptation.

5.2.2 Real-time Human Interaction. This section details our activities on *step (7)* of the pipeline (see Figure 3) to successfully deploy the *sensing and actuating* components of FABRIC in Figure 1. In order to allow for a real-time human interaction, the cobot needs to constantly perceive its environment and the human actions. For this purpose, we gather visual information about the scene using two cameras. A human worker interfaces the system through a yellow glove and a safety helmet (as shown in Figure 6). From an RGBD camera, we recognize hand gestures and the

Table 2. Observation Vector with the Descriptions and How They Are Derived

	ID	Observation	Description	Means of detection
Observed human behaviors (σ^h)	$\sigma 1$	Is human detected?	yellow glove detected by RGBD camera	object tracking
	$\sigma 2$	Is human looking around?	side marker of helmet is detected	AruCo marker tracking
	$\sigma 3$	Is human attempting to grasp?	human grasps a detected object and/or moves it	GMMHMM gesture classification
	$\sigma 4$	Is human warning the robot?	warning gesture by human is detected	GMMHMM gesture classification
	$\sigma 5$	Is human idle (inactive)?	human hands are not moving	GMMHMM gesture classification
Observed environment (σ^e)	$\sigma 6$	task success	counting succeeded subtasks	successful subtasks ($\sigma 8$)
	$\sigma 7$	task failure	counting failed subtasks	failed subtasks ($\sigma 9$)
	$\sigma 8$	has human succeeded in a subtask?	human grasped and placed an object onto a correct container tray	human grasp recognition ($\sigma 3$) and load sensors detect the correct tray
	$\sigma 9$	has human failed in a subtask?	human grasped and placed an object onto a wrong container tray	human grasp recognition ($\sigma 3$) and load sensors detect a wrong tray

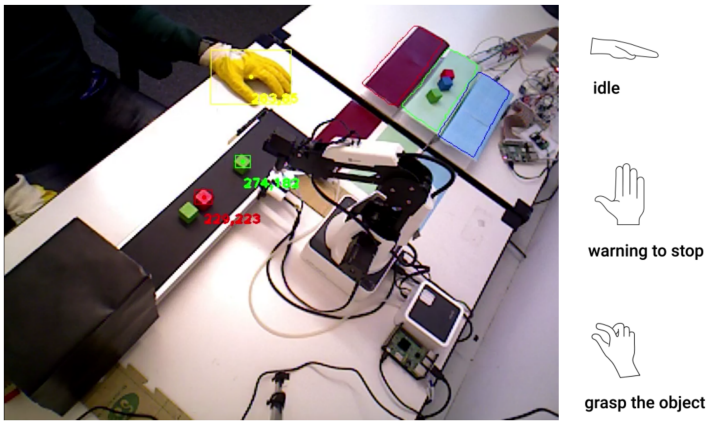


Fig. 8. The cubes, the containers, and the glove are detected in the work space even when the cubes are grasped.

objects to interpret human actions and interactions with the objects. The other camera on the wall tracks human attention through the head gestures recognized. In addition, there are load sensors located under the container trays that detect the objects the participants put onto them. The human activities and environment state are processed under the *sensing* component of our framework (in Figure 1 to generate the observation vector for the cobot's decision-making.

Observation Vector. The observation vector is the feature vector for the cobot decision-making, i.e., observables for A-POMDP models during a task and collectively used to estimate the interaction type in ABPS before a new task (in Section 3.2). We define the vector with nine different observables as shown in Figure 2. In Table 2, we list these observables and how they are generated by the *sensing* component of FABRIC. Among the observables, there are direct observations, such as, a human is detected ($\sigma 1$) when a hand glove is visible for the HAR system, the human attempts to grasp an object of interest ($\sigma 3$), the human warns the cobot to stop ($\sigma 4$), and the human stays idle ($\sigma 5$).

The last three, $\sigma 3$, $\sigma 4$, and $\sigma 5$ are obtained by a state-of-the-art HAR module that recognizes the constrained and distinct human gestures from the movements of the glove (see Figure 8).

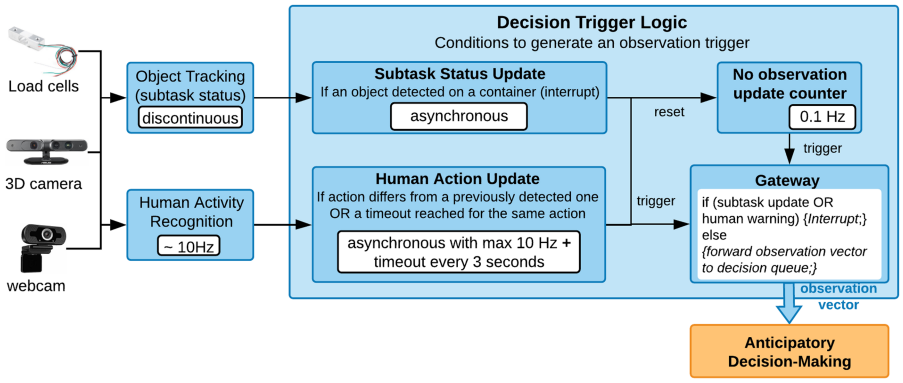


Fig. 9. Decision trigger logic of our framework in Figure 1.

Moving from the 3D tracked location of the glove, the hand's velocity, and the spatial relation between the tracked objects and the hand, human actions of *idling (inactive), attempting to grasp*, and *warning sign* are recognized by a **Gaussian Mixture Model (GMM) Hidden Markov Model (HMM)** classifier (inspired by [36, 48]). An additional class of "undefined action" is also added with some random irrelevant gestures in order to avoid false positives on the three important gestures. A GMM operation creates discrete clusters of these continuous variables that are then classified by HMM. We create a separate GMM-HMM for every action of interest. Then, every time a feature vector arrives, all models generate a confidence value and the one with the highest confidence is taken as the current most probable action. After training the system with various people, we reached over 90% accuracy with $10Hz(fps)$ on a continuous video stream.

The observable $\sigma 2$, looking around action, is detected by tracking the markers on the safety helmet (in Figure 6). The posture of the participant's head is used to detect if the person is looking around but not gazing somewhere in the work area. For this purpose, we ran several tests during the calibration experiments to define margins of tracked markers to interpret if a human is looking around. The accuracy of this observable is very high; however, a misdetection is compensated through the need to sequentially observe this gesture to estimate the state of a human who has lost her attention in our A-POMDP (in Figure 2). Additionally, there are some observables that are derived from the others. According to the assigned task rules, we check the status of all of the load sensors under the container trays and the grasped object to output a subtask success or failure ($\sigma 8$ and $\sigma 9$). Finally, a task's success and failure are calculated by counting the amount of successful and failed subtasks. When the total number reaches a value defined by a task definition (mostly 10 in our experiments), either $\sigma 6$ or $\sigma 7$ becomes true so that a cobot decision-making model (i.e., A-POMDP) terminates. In our experiments, the task status flags are just for the cobot models to terminate. They are not used to measure the final task performance, which is calculated using the subtask results and who achieved it.

Decision Trigger Logic. The stochastic nature of human behaviors makes it difficult for a robot to follow the decision-making with a constant update frequency, especially when the scenario does not follow turn-taking collaboration. For this reason, our cobot should balance timely decision-making with natural interaction speeds, catching human actions and the important changes in the environment to respond reliably. For fluent coordination, our cobot implements a decision trigger logic as represented in Figure 9. In our setup, a cobot decision can be triggered in three different conditions: (1) when a container update is recognized (a subtask is completed), (2) when a change in human actions is recognized, (3) when a timeout is reached in case of no change in human

actions or a long-time absence of an observation update. The *decision trigger logic* component of FABRIC handles these conditions to synchronize the sensory inputs to create the observation vector, analyze it in case a decision should be triggered, and forward it to the *anticipatory decision-making* component accordingly (at the *line 7* of Algorithm 1). Starting with the discontinuous events, as shown in the figure, both a subtask update (i.e., when a cube is detected on a container) and a new human action can happen at any time. We generate interrupts for an immediate decision-making whenever a cube is detected on a container, and when a warning action is recognized. If a decision of cancelling an ongoing action is made from the warning gesture, this also interrupts the existing cobot action under the *actuating* component of FABRIC.

For all of the other cases, we catch and respond to every change in recognized human actions. A currently recognized human action is compared with the previously detected one and a decision is triggered whenever they differ. Nevertheless, a new decision is triggered regardless of this comparison after a timeout of 3 seconds. Through the calibration experiments with the participants, we have calculated the time it takes for a human to start and finish an action to be approximately 3 seconds, taking into account the actions of pick and place, warning the robot, and idling (in the appendix). Therefore, an update frequency of 10 Hz would be unnecessarily fast for the cobot decision-making and would even be unreliable as the cobot may think a human is taking the same action repeatedly even if it is still one action in progress. Even though we use an online POMDP solver with a fast real-time response, generating a decision is the limiting factor for the update frequency of the cobot. Our experiments have shown that generating an optimal decision can take up to around 1 second on an average PC (with 8 GB of RAM and Intel I5 processor), whereas we may generate and forward a new observation vector with a frequency of about 10 Hz. For that, we queue the observation vectors, except for an interrupt case, not to overload and crash the *decision-making* component with many requests (see the gateway in Figure 9). A new observation from the queue is forwarded as soon as a currently running decision process ends. Thanks to the multi-agent architecture of FABRIC, a decision-making process runs in parallel with an ongoing cobot action being executed.

5.3 Calibration Experiments

This section summarizes our activities under *step (8)* of the pipeline in Figure 3. Since the experiments conducted here are not the main focus of the article, their details are provided in the appendix. Calibration experiments involve an initial case study with real humans to validate and fine-tune our novel interaction setup and the cobot system. This experiment acts as prototyping tests on the setup to check and analyze if our setup and the cobot's collaboration is reliable, and whether the setup can induce sufficient cognitive load on the participants to observe unanticipated human behaviors. By doing so, we can better evaluate the adaptation goals of our anticipatory decision-making approaches and the contributions of our framework to the fluency of an HRC. We also utilize the results of these experiments to fine-tune the decision trigger logic, real-time human collaboration capability, and the cobot's reliability in its responses (i.e., the decision-making models).

The first round of experiments evaluated whether the cobot effectively communicates its decisions and internal states to its human partner, i.e., robot expressiveness. We invited 12 people and showed them the cobot actions defined in our A-POMDP design (in Figure 2). Especially the *planning* and *pointing to remind* actions are needed to effectively communicate the cobot's intention as the planning is a precursor for the human to be aware that the cobot may take over the task soon, and the pointing action is to remind the human about the task. In the end, we experimented with different motion designs and chose the most expressive ones, which were correctly understood by the majority of the participants who have never interacted with a robot before.

Our collaboration setup aims at placing a cognitive load on the humans and the degree of cognitive load may differ between the task types in Figure 7. Our goal is to compare the types and choose suitable ones for the final experiments. Our criteria are: (1) The task is cognitively demanding enough to observe unanticipated human behaviors, (2) the task is easy and motivating enough to keep the human engaged in the collaboration. We let eight participants and the cobot collaborate over an extended period (approximately 1.5 hours each) executing our tasks in Figure 7. We compare objective and subjective measures from each task to find out how well the human-cobot team achieved the tasks, the levels of cognitive load induced on the participants (through NASA-TLX metrics [18]), how frequently they exerted unanticipated behaviors, such as, lost attention and failures. More details on the experiments, the metrics used, and the results are given in the appendix.

In summary, we conclude that the setup is effective in invoking unanticipated human behaviors. This is clear in task type-4 (in Figure 7) through the participant responses on not remembering the rules, their lost attention over the course of the experiment, and from very high NASA-TLX scores. These analyses show that the higher the cognitive load is, the more unanticipated human behaviors, followed by more human errors. Since it has led to fluent collaboration with a significant amount of invoked unanticipated behaviors, we choose task type-4 as the main collaboration task to evaluate our cobot's short-term adaptation capabilities in Section 6.1. For the evaluation of long-term adaptation in Section 6.2, we select to work with task type-5, which has the same configuration as type-4 along with an additional Stroop effect (in Figure 7). As mentioned in Section 5.2.1, the Stroop effect was very difficult to master for this experiment that repeated the task only three times for each participant. However, it has proven to be very suitable for a longer collaboration as it leads to a noticeable change in human characteristics, such as expertise through the learning effect.

5.4 Training For Real Anticipatory Decision-making

In this section, we first detail how we improve our framework with the feedback from the calibration experiments, then talk about our training process, following *step (9)* of the pipeline in Figure 3. The parameters of the A-POMDP models have been first tuned in simulation, in various collaboration scenarios. However, a transition to the real-world interaction requires a further adaptation of the models to a real human's decision-making frequency. We are aware that there cannot be one robot decision model optimized for all types of people and interactions. For this purpose, as we discussed in Section 3.2.2, our initial goal is to obtain an A-POMDP model, the "base model", that works reasonably in many interaction scenarios and obtains high rewards. This model is then used to generate several other models for ABPS to select the most suitable one during a specific interaction, as we did for the simulation experiments in [15], toward a personalized human collaboration.

For the base model, we first manually tweaked the probabilities until the new cobot policy reaches an average high reward against many different interactions at the lab. Then, we collect real observations from the calibration experiments (in Section 5.3) and use this data to further train the model's intrinsic parameters. In particular, from an observed sequence of $s_0, a_0, s_1, a_1, \ldots, s_n$, a_n, s_{n+1} we update the transition probabilities, $T(s'|s, a)$, where $s \in S$ and $a \in A$ are the A-POMDP states and actions defined in the model (in Figure 2). The current state information, i.e., s, is subjectively obtained from the participants if it is a human mental state, e.g., the human is tired. Similarly, the sequence of $s_0, a_0, \sigma_1, s_1, a_1, \sigma_2, \ldots, s_n, a_n, \sigma_n, s_{n+1}$ is used to update the observation probabilities, $O(\sigma|s', a)$, where $\sigma \in O$ is the observation vector the cobot receives from the environment after its action. For the purpose of systematically observing and training for distinct human behaviors, most common interaction cases are identified and examined. We also examine

the cobot's reactions in the test runs to filter out the unreliable cases. We list below some of these common interaction scenarios we iteratively tested with different participants and collected the observations at different execution speeds and intervals.

- Human continuously grasps: Human starts in the idle position, grasps the object and places it into a container to grasp the other object right afterward.
- Human idles for long: People idle for changing duration of time as a reflection of their states like evaluating the task or being tired. The cobot may take over or give more time to the human depending on the individual.
- Human warns the robot: The cobot should receive and process the warning on time no matter what the cobot was doing and in whatever way it decides to respond to this warning.
- Human looks around, not attending to the setup for a while: The cobot should estimate that the human may have lost her attention and take an action.

One of the qualitative findings was regarding the duration between multiple subtasks, i.e., the transition times. For example, sequentially recognized idling actions between two subtasks may be interpreted as "the human is tired" by our A-POMDP model if the robot transitions to that state quickly, whereas in reality the human may be waiting for a new object to arrive on our slow conveyor system. Thanks to the calibration experiments, the state transition probabilities are updated to favor the probability of a state persisting rather than transitioning to another one. In the training, we use the subject data to improve the model parameters as mentioned above, which were previously trained on the simulation. The final base model turned out to be a reliable model against various humans. However, we note again that there cannot be a single probability distribution defined for a robot in its interactions with multiple people. The base model is used as the "proactive model", our A-POMDP that handles unanticipated human behaviors, in our short-term adaptation experiments to compare it with a reactive model that discards such behaviors (see Section 6.1).

For the long-term experiments (in Section 6.2, we generate a policy library, being the *decision model library* of FABRIC. For that, we randomly adjust the transition and observation probabilities of the base A-POMDP model to generate various models, each of which handles a unique human type, and solve for their optimal policies to construct our policy library Π. This is to limit the arbitrary generation of the robot policies to avoid overloading the library with unreliable candidates. Then, our ABPS mechanism runs on top of the library to select a policy for an estimated human type (see Algorithm 1). We first train ABPS for the observation and the performance models (in Definitions 1 and 2, respectively) against simulated humans as in [15]. In the end, 20 policies were selected for their use in the experiments based on: (1) how well they performed overall against many different randomly generated human models after discarding the worst ones, (2) how distinct their performance models are from the other policies by grouping the similar ones. The similarity between two policies is measured using the KL-divergence method that calculates the statistical distance between their performance models. We remove the policies that generate a high similarity score to another one in the library. Some policies ignore a human's warnings and try to complete a task, whereas some pay more attention to a human's needs, taking the human as the leader of the collaboration. The tradeoff between these two is clearer when it comes to non-collaborative human types. There are also some policies that prefer to encourage the human to complete the task, e.g., by pointing out to remind the human when distracted instead of directly taking over the task. Which policy is optimal depends on the interacted human type and the task definition.

We are agnostic to the exact type labels of humans in our experiments. As mentioned, we assume each human the robot is interacting with has an unknown type to the robot, which can only be estimated as a distribution over the known types. For that, we have crafted 16 differ-

ent (known) human types using the modeling scheme in Figure 5 with the goal of each generating as a distinct set of human actions as possible in simulation. That means creating human types with the extremes of the four characteristics of our concern, namely the levels of *expertise, stamina, attention*, and *collaborativeness*. Our assumption is that an unknown human type can be approximated as a probability distribution over these extreme types. We note that since each of the 16 human models are stochastic, they still generate a diversity of behaviors after random sampling (see Section 5.1.1). The collaboration of each of the 20 robot policies and each of the 16 human types is repeated for 90 sequential tasks in our simulation environment. In total, we accomplished 28,800 interactions (28,800 task instances and 288,000 subtask instances), which is very difficult to manage in real-world scenarios.

Finally, the human type estimation of ABPS, the observation model, from the simulation experiments is also updated with the real human observations collected from the calibration and the short-term adaptation experiments, where in both ABPS, *the adaptive policy selection* component of FABRIC in Figure 1, is not used. The labeling of the human types is done from the objective and subjective measures collected from the participants, such as their success rates, the warning amounts, exhaustion and attention levels, and their average idling times. The updated observation model provides a more reliable estimation thanks to the real observations and can accurately estimate a wider diversity of the human types than the interacted ones in our calibration experiments thanks to the simulated data (in Section 6.2). Our goal is now to show that the framework is applicable in the real-world for its real-time autonomous human collaboration and to prove that it covers our extended human adaptation goals leading to a more efficient and natural collaboration.

6 EVALUATION AND RESULTS

This section evaluates our framework and its extended human adaptation capabilities covering both our short- and long-term collaboration goals (i.e., the *steps (10), (11), and (12)* of the pipeline). In Section 6.1, our goal is to validate our findings in Section 3.2.1, which is the short-term adaptation capability of our robot handling unexpected human conditions. We repeat the same experiment done in simulation, comparing our A-POMDP robot decision model design with a conventional reactive model in Görür et al. [2018], this in a real-world setting. Then, in Section 6.2, we integrate the full anticipatory decision-making system on the setup to validate the applicability of both our ABPS mechanism from Section 3.2.2 and our overall framework introduced in Section 3. All of the experiments use a within-subject design and we invited different people to each of the experiments to be able to independently evaluate different aspects without any practice effect or prior system knowledge.

6.1 Evaluation of Short-term Adaptation

With this experiment, our goal is to examine the hypotheses below:

Hypothesis 1. *A cobot's fluent collaboration with a human contributes to increased performance in a cognitively challenging task when compared to a human working alone.*

Hypothesis 2. *Our A-POMDP model adapting to a human's unanticipated behaviors (extended short-term adaptation) contributes to more efficient and natural collaboration when compared to a cobot model that does not handle such behaviors.*

Hypothesis 3. *Our A-POMDP model adapting to a human's unanticipated behaviors (short-term adaptation) shows better adaptation skills and it has a higher perceived collaboration, trust, and positive teammate traits, than a cobot model that does not handle such behaviors.*

We let the participants interact autonomously with two different cobot planners. The first one is a *proactive robot* that runs our A-POMDP model in Figure 2. This cobot first anticipates a human's

characteristic, e.g., lost attention, incapability, or tiredness, and then it estimates if the human needs assistance or not (extended short-term adaptation, in Section 3.2.1). On the contrary, the other cobot, the *reactive robot*, does not handle the unanticipated behaviors of a human. It treats a human's need for help as a directly observable (deterministic) state. The reactive robot deterministically decides that a human needs help when (i) a certain time duration has passed without a cube placement (i.e., without a subtask completion), (ii) the human is not detected around the workplace, or (iii) the human fails in a subtask. We design the *reactive robot* by removing the *anticipation stage-1* of our A-POMDP model design in Figure 2. With the *anticipation stage-2* being deterministic, this model is designed as an MDP. Through this comparison, our intention is to show the importance of handling the unanticipated human behaviors (i.e., stochastic interpretation of such human states) for an improved short-term adaptation.

6.1.1 Objective and Subjective Measures. This section details our activities under *step 10* of the pipeline in Figure 3. Our goal is to measure the efficiency and naturalness of the collaboration and the adaptation skills, perceived collaboration, trust, and positive teammate traits of the cobot. In a work environment, in general, efficiency of a task is defined by the quality and the duration of the work done. Hence, work division is crucial for the efficient use of resources that takes into account the availability and the matching skill sets of the workers. For example, in a pick and place task, varying fragility of the products may require a work division, where a human should handle the fragile objects while the cobot should work on bulky and heavy ones. Thus, we consider the highest efficiency when *a task is successfully accomplished by its assigned collaborator*. In our case, the tasks are initially assigned to the people as we focus on the cobot's human anticipation skills and its correct interpretation of a need for assistance. This is motivated by the participants through our scoring system (see Section 5.2.1) and is also known to the cobot by initializing A-POMDP models from the state of "Task is Assigned to Human" (in Figure 2).

The term naturalness defines a natural interaction between the cobot and a human where they achieve a fluent collaboration. In particular, the level of naturalness is defined by the fluency of their communication that is often expected to be nonverbal in a collaboration [22]. A good indicator of the naturalness of collaboration is the level of intrusive behaviors from the cobot, which leads to frustration in the human partner. A cobot should reliably understand the collaborated human's needs and preferences to adapt to a situation and to avoid intrusive behaviors. In our setup, a warning gesture is provided to the human so that she communicates her displeasure with a cobot behavior. Therefore, the number of warnings hints at the naturalness of a collaboration. To conclude, our quantitative measures are the following:

- **Rewards gathered:** We use the same robot reward mechanism as in Section 3.2.1 for both of the reactive and the proactive robot, i.e., punishing each warning received from a human and a subtask failure and rewarding each subtask success.
- **Number of warnings:** The number of warning gestures a participant made during a task.
- **Task success rate (S_{task}):** The overall success rate of a task, calculated by $S_{task} = \frac{n_s}{n_{total}}$, where n_s is the total amount of successful subtasks and n_{total} is the total amount of subtasks.
- **Human success rate (S_{human}):** The rate of the successful placements of a human collaborator out of all of her attempts, calculated by $S_{human} = \frac{n_{s_{human}}}{n_{s_{human}} + n_{f_{human}}}$, where $n_{s_{human}}$ and $n_{f_{human}}$ are the amount of successful and failed subtasks by the human in a task, respectively.
- **Human contribution in a task (C_{human}):** The overall successful contribution of the human to a task is calculated by $C_{human} = \frac{n_{s_{human}}}{n_{total}}$
- **Task efficiency (η_{task}):** The task efficiency depends on the overall task success and how much the assignee (human partner) has contributed. It is calculated by, $\eta_{task} = S_{task} \cdot C_{human}$.

Table 3. Subjective Statements Asked during the Short-term Adaptation Experiments

Questionnaire Statements	Category
"The robot helped me remembering the rules"	Robot's effect on the human cognitive load
"The robot's interference distracted me"	
"Without the robot, I could have achieved a better score."	
"A robot collaboration is beneficial/needed in such tasks." **(General)**	
"The robot was able to understand my assistance needs and behaviors during the task."	Naturalness of the collaboration and the reliability of the robot
"The robot acted as I expected."	
"The robot took over at the right times when I needed assistance."	
"I lost my attention during the task."	
"I had to warn the robot a lot."	
"I became more and more exhausted during the experiment."	
"The robot reacts repetitive rather than responding to my behaviors."	
"The robot was able to adapt to my assistance needs over the course of the experiment." **(General)**	
"I did not want the robot to take over the task."	Trust in the robot, perceived collaboration and teammate traits of the robot
"I trust the robot with the task."	
"The robot seemed competitive in taking over the task."	
"The robot could have assisted me more."	
"I became more comfortable interacting with the robot over the course of the experiment." **(General)**	
"I liked collaborating with the robot." **(General)**	
"I would work with such a robot on such kind of tasks in a work environment." **(General)**	
"Which robot was more reactive (following preset rules) in its takeovers?"	Comparison
"Which robot was learning and adapting better to your assistance needs?"	

We subjectively evaluate the fluency of the collaboration from the perspective of the participants. The subjective measures are collected by means of questionnaire responses, where the participants describe their agreement with a statement through a 5-step Likert scale. The statements are given in Table 3. The category of questions concerning the cobot's effect on the cognitive load is to see if the cobot is perceived as a negative or positive influence on such a challenging task. The remainder of the statements are for measuring the metrics of *perceived naturalness, reliability, trust, positive teammate traits,* and *perceived collaboration of the cobots* that are mostly inspired by the relevant HRC research, in [22, 30, 41, 42]. Even though we categorize the statements for simplicity, they are interchangeably evaluated under the relevant hypotheses. There are also similar statements that are shuffled in the questionnaires to check the consistency of a participant's ratings. The statements categorized as "general" are rated only twice for each participant, one is after the proactive robot experiment and the other is after the reactive one. The comparison statements are only asked once at the very end of the experiment to let the participants compare the two cobots. The other statements are all task-specific and asked after each task completion (see Figure 11 for the experiment protocol).

6.1.2 Experiment Protocol. We invited 14 people (the ages range from 17 to 38) from different ethnic groups and from various backgrounds (computer science, social sciences, law, chemical engineering, tourism, public affairs, and business school) and let them interact with the two cobot models. We considered age and ethnicity as potential covariates but found no significant effect on the results. We designed a within-subject experiment to compare the two cobots. Thus, each of the participants interacted with both of the robot types. We note that the participants knew that there were two types of cobots and that they needed to evaluate them both. We only notified a participant when we switched the robot type; hence, they only knew them as robot type-1 and type-2 for anonymity. Once a participant takes the chair, the experiment starts (see Figure 10).

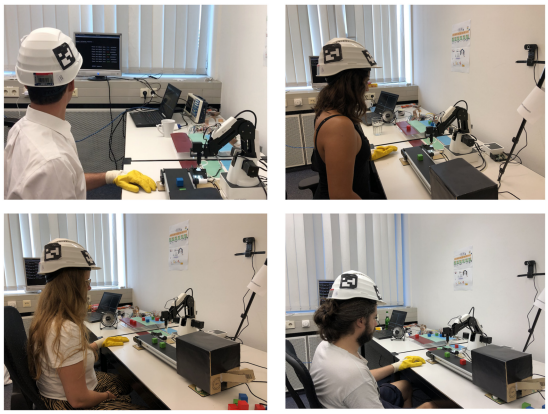


Fig. 10. Participants are collaborating with the cobot.

After the calibration experiments, we chose task type-4 as the most suitable task for collaboration (see Figure 7). We kept the task type the same throughout the whole experiment to remove any effect from changing task difficulties. The experiment procedure is depicted in Figure 11. First, we do a training round with each participant on a simpler task to avoid practice-effect. The operator describes how to complete a task successfully (e.g., how to grasp and place the cube, how to warn the robot) and motivates the participant that the tasks are always assigned to them and for each subtask they achieve the team receives a higher score than the cobot completing it (see the scoring system in Section 5.2.1). The operator also mentions that there is no time limitation and they have to wait for audio feedback after each placement before moving to the new one. Then, the experiment starts. In general, there are three main steps we follow during the experiments: (1) The participant completes one task alone without any robot interaction, (2) the participant collaborates with the reactive robot for three tasks, (3) the participant collaborates with the proactive robot for three tasks (in Figure 11). In total, each participant completes seven tasks. In order to remove any practice-effect over different robot types, seven randomly selected participants interacted first with the reactive robot and the other seven with the proactive one. After each task, a participant needs to fill out a task-specific survey. At the end of each robot type interaction, the participant fills out the general survey questions (in Table 3). In total, an experiment with a participant lasts approximately 1.5 hours.

6.1.3 Results and Discussions. In Figure 12, we give the box and whisker plots of the success and efficiency analysis and in Figure 12(e) are the numerical values and the ANOVA results. As seen in Figure 12(a), the task success rate when the participants worked without a cobot is significantly worse than a collaboration with any of the cobot models ($p < 0.05$). A collaboration with a proactive robot contributes positively to a task success by 41% and with the reactive robot, it is increased by 35% compared to a human working alone. Similarly, a human's success rate has also significantly increased when the human collaborated with either of the cobots (in Figure 12(b) with $p < 0.05$ for both cases), whereas with the proactive one this increase is slightly higher (by 35% increase on the human success rate).

In Table 4, we give the Likert scale (1–5 with increasing agreement) results of the participant's statements throughout this experiment. Table 4(a) shows the statements for analyzing a cobot's impact on such a challenging task. The participants state that both of the cobots helped them remember the task rules and that a robot collaboration is beneficial in such tasks (see high mean rates in the table and the sum across all Likert items significantly favors the impact of the proactive

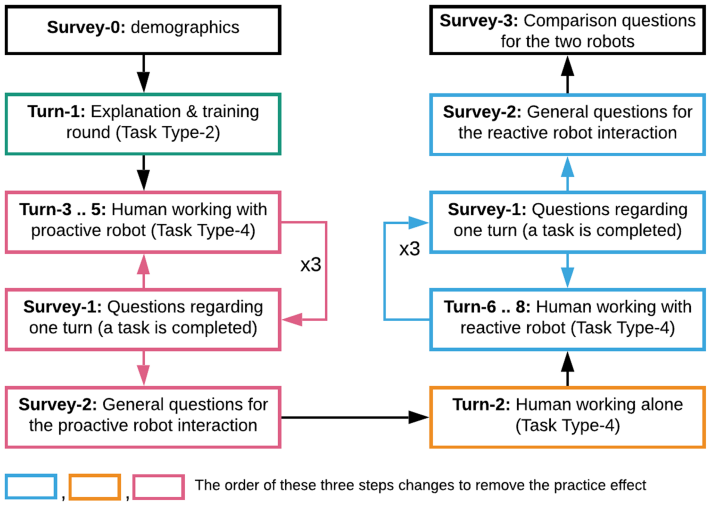
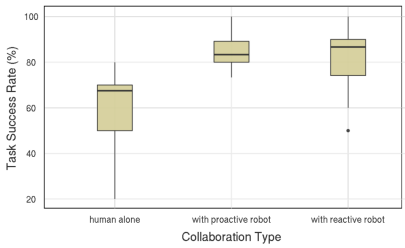


Fig. 11. The protocol of the experiments for each participant.

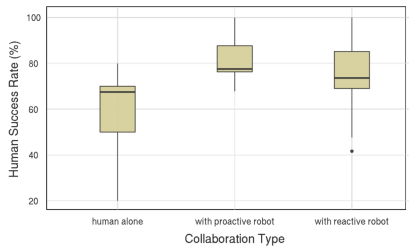
case). Finally, the average task efficiencies are shown in Figure 12(d). Both the reactive and the proactive robot contributed significantly positively to the task efficiency ($p < 0.05$), compared to a human working alone (an increase of 56% for the proactive robot and 21% for the reactive one to the overall task success rates in Figure 12(e)). Thereby, we underscore the importance of such cobot collaborating with humans in challenging tasks. The success rate analysis, the efficiency results, and the subjective ratings of the participants support **Hypothesis 1**.

Even though the proactive robot on average achieved higher success rates than the reactive one (see Figure 12(e)), there is no significant difference between them. The decision models only decide the level of taking over a subtask. After that, both cobots achieve success in placing the cubes (over 95%). Therefore, some participants were comfortable leaving the task to the cobot once they realize its capability. For the reactive robot, this happened significantly more than the proactive one due to its deterministic rules. Still, the proactive robot provides a more stable success than the reactive one by keeping the variance low (see Figure 12(a)). In both cases, our main concern is how much of this success actually comes from the human. Figure 12(c) shows that the proactive robot significantly increased the human's contribution to success when compared to the reactive robot ($p = 0.038$). The reactive robot also respects the initial task assignment; however, it favors taking over a task when, for example, a human idles too long, discarding the unanticipated human behaviors and preferences. This led to a decrease in a human's successful contribution in a task during her collaboration with the reactive robot when compared to her performance alone (as shown in Figure 12(e)). Finally, since a higher efficiency is achieved when a task is successfully accomplished by its assigned collaborator, the proactive robot significantly increased the task efficiency of a human working alone by 57% ($p = 0.0023$) and ruled out the task efficiency achieved with the reactive robot (with $p = 0.0104$ in Figure 12(e)).

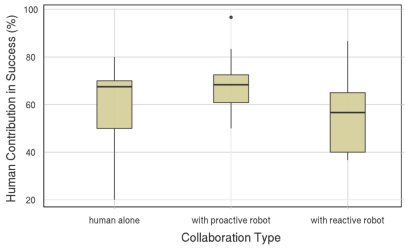
As discussed in Section 6.1.1, the naturalness reflects a fluent communication in that the handovers and turn-taking need to be interpreted correctly by both of the collaborators. Figure 13(a) suggests that the proactive robot could keep the warnings close to zero, suggesting a higher accuracy in estimating a human's unanticipated behaviors and need for help. In Figure 13(c), the ANOVA results point out the significantly higher number of warnings the reactive robot received, which is 3.6 times of the proactive robot's with $p < 0.0001$. The participants also evaluated



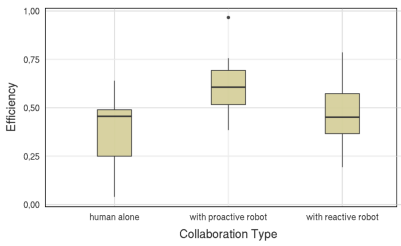
(a) Overall task success rates



(b) Human success rate in a task



(c) Human's successful contribution to the task



(d) Task Efficiencies

Comparisons		Mean rate human alone case	Mean rate with proactive robot	Mean rate with reactive robot	ANOVA
Overall Task Success (%)	human alone vs with proactive robot	60.36	84.78	--	$F(1, 26) = 21.20, p < 0.0001$ **Significant**
	human alone vs with reactive robot	60.36	--	81.19	$F(1, 26) = 11.67, p = 0.0021$ **Significant**
	with proactive robot vs with reactive robot	--	84.78	81.19	$F(1, 82) = 1.05, p = 0.3077$ **Not Significant**
Human Success Rate (%)	human alone vs with proactive robot	60.36	81.29	--	$F(1, 26) = 14.24, p = 0.001$ **Significant**
	human alone vs with reactive robot	60.36	--	74.18	$F(1, 26) = 4.59 p = 0.042$ **Significant**
	with proactive robot vs with reactive robot	--	81.29	74.18	$F(1, 82) = 2.29, p = 0.1343$ **Not Significant**
Human Contribution in Overall Success (%)	human alone vs with proactive robot	60.36	69.52	--	$F(1, 26) = 2.52, p = 0.1244$ **Not Significant**
	human alone vs with reactive robot	60.36	--	55.71	$F(1, 26) = 0.52 p = 0.4753$ **Not Significant**
	with proactive robot vs with reactive robot	--	69.52	55.71	$F(1, 82) = 8.87, p = 0.0038$ **Significant**
Efficiency	human alone vs with proactive robot	0.39	0.61	--	$F(1, 26) = 11.38, p = 0.0023$ **Significant**
	human alone vs with reactive robot	0.39	--	0.47	$F(1, 26) = 1.21, p = 0.2809$ **Not Significant**
	with proactive robot vs with reactive robot	--	0.61	0.47	$F(1, 82) = 6.87, p = 0.0104$ **Significant**

(e) Mean values and the ANOVA results.

Fig. 12. (a)–(d) Box and whisker plots of the overall task success rates, a human's success rate, a human's successful contribution to an overall task, and the efficiency values averaged over 14 participants during their performance alone and their collaboration with the proactive robot and the reactive robot. (e) The ANOVA results of the plotted objective measures to compare the performance of the two cobots.

if a collaboration with a cobot felt comparatively natural to them, i.e., more human-like. The participants stated that the reactive robot's interference distracted them significantly more than the proactive case (with $p = 0.023$ in Table 4(a)). This is largely due to the significantly increased unexpected interferences from the reactive robot (with $p = 0.012$ as shown in Table 4(b)). The "expectation" here is an ambiguous term that might differ from one person to another; however, it is inherently discussed in the literature that an efficient collaboration is achieved when the partners reach a joint intention. Thus, the expectations of the partners are often toward understanding each other and obtaining a joint action on a task [7].

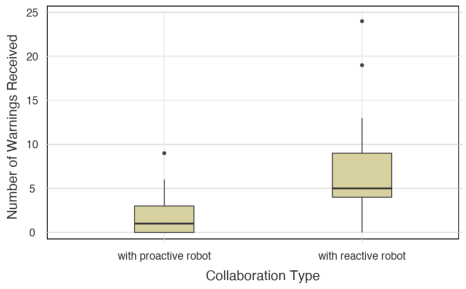
Table 4. The Results of the Subjective Statements Asked during the Short-term Adaptation Experiments

Statements	μ (proactive rates)	μ (reactive rates)	ANOVA
(a) Statements for analyzing a robot's impact on such a challenging task.			
"Robot helped me remembering the rules"	3.45	3.55	$F(1, 26) = 0.104, p = 0.749$ **Not Significant**
"Robot interference distracted me"	1.95	2.81	$F(1, 26) = 5.87, p = 0.023$ **Significant**
"I could have achieved more without the robot "	1.98	2.33	$F(1, 26) = 1.24, p = 0.276$ **Not Significant**
"A robot collaboration is beneficial in such tasks"	4.36	3.50	$F(1, 26) = 4.66, p = 0.040$ **Significant**
(b) Statements for analyzing a robot's reliability, adaptation, and the naturalness of the collaboration.			
"The robot was able to understand (anticipate) my assistance needs and behaviors during the task."	3.69	3.00	$F(1, 26) = 8.78, p = 0.006$ **Significant**
"The robot acted as I expected."	3.74	2.98	$F(1, 26) = 7.24, p = 0.012$ **Significant**
"The robot took over at the right times when I needed assistance."	3.79	2.90	$F(1, 26) = 11.8, p = 0.002$ **Significant**
"I had to warn the robot a lot."	1.86	3.64	$F(1, 26) = 29.6, p < 0.001$ **Significant**
"The robot reacts repetitive rather than responding to my behaviors."	1.74	3.07	$F(1, 26) = 24.1, p < 0.001$ **Significant**
"The robot was able to adapt to my assistance needs over the course of the experiment."	4.00	2.93	$F(1, 26) = 9.98, p = 0.004$ **Significant**
(c) Statements for analyzing the trust in the robots, their perceived collaboration, and teammate traits			
"I did not want the robot to take over the task."	2.64	3.26	$F(1, 26) = 4.86, p = 0.030$ **Significant**
"I trust the robot with the task."	4.43	4.05	$F(1, 26) = 4.29, p = 0.041$ **Significant**
"The robot seemed competitive in taking over the task."	2.10	3.90	$F(1, 26) = 25.8, p < 0.001$ **Significant**
"I became more comfortable interacting with the robot over the course of the experiment."	4.57	3.50	$F(1, 26) = 11.0, p = 0.003$ **Significant**
"The robot could have assisted me more."	2.29	2.17	$F(1, 26) = 0.379, p = 0.543$ **Not Significant**
"I liked collaborating with the robot."	4.50	3.57	$F(1, 26) = 7.5, p = 0.011$ **Significant**
"I would work with such a robot on such kind of tasks in a working environment."	4.50	3.43	$F(1, 26) = 9.98, p = 0.004$ **Significant**

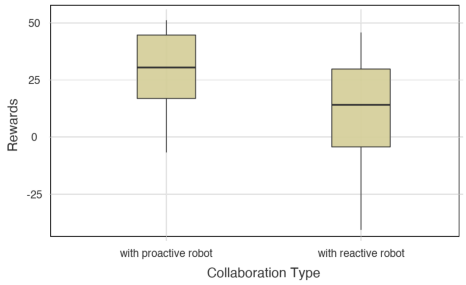
The participants state their agreement over a 5-step Likert scale and the results are averaged over 14 participants. The ANOVA results compare the performance of the proactive and the reactive robot.

Finally, Figure 13(b) shows the rewards gathered by the cobot, which is the combination of task success and the number of warnings. As expected, the proactive robot has received 2.6 times more rewards than the reactive robot ($p < 0.0001$). With that and the analysis above, we conclude that our A-POMDP cobot model (the proactive robot) leads to a more efficient and natural collaboration when compared to the same cobot that does not handle a human's unanticipated behaviors, which supports **Hypothesis 2**.

Hypothesis 3 is evaluated through the subjective statements of the participants in Table 4. Firstly, since the collaboration with both of the cobots achieved very high success rates, the participants rated their trust in both of the cobots very high (the proactive robot with the mean rating of 4.43, the reactive robot with 4.05 out of 5.00 in Table 4(c)). The participants still thought that the proactive robot is significantly more trustworthy than the reactive one (with $p = 0.041$). The participants think that the proactive robot took over the task with significantly more accurate timing, i.e., when they needed assistance, than the reactive robot (with $p = 0.002$). This is also supported by the statements of "The robot acted as I expected" and "The robot was able to adapt to my assistance needs" that are both rated significantly higher for the proactive robot (with $p = 0.012$ and $p = 0.004$, respectively, in Table 4(b)). Similarly, another consistent analysis is made for the negative statement, "I did not want the robot to take over the task", which is rated significantly



(a) Number of warnings received



(b) Rewards gathered by the robots

Measurements	Mean values of proactive robot	Mean values of reactive robot	ANOVA
Number of warnings received	1.88	6.79	$F(1, 82) = 37.68, p < 0.0001$ **Significant**
Total rewards gathered	28.95	10.88	$F(1, 82) = 17.00, p < 0.0001$ **Significant**

(c) Mean values and the ANOVA results

Fig. 13. (a), (b) Box and whisker plots of the number of warnings the proactive robot and the reactive robot have received from a human and the rewards they gather during a task, averaged over 14 participants on 42 tasks for each of the cobots in total. (c) Mean values of the two cobots and the ANOVA results for the comparison of their performance.

higher for the reactive robot that took over more frequently ($p = 0.030$ in Table 4(c)). All in all, these analyses point out a better collaboration experience with the proactive robot, resulting in a higher level of trust (The sum of scores under Table 4(c), i.e., the Likert scale for the trust and collaboration capability of a robot, significantly favors the proactive case).

In general, better anticipation of a person's assistance needs, respecting her preferences, and greater trust suggest a higher acceptance for the proactive robot. In addition, the proactive robot contributed to an increased performance of its partner, and led to more efficient task completion. We thus conclude that the proactive robot has more positive teammate traits than the reactive one. The participants also indirectly support this by stating that they would prefer to work with the proactive robot on this kind of demanding tasks significantly more than the reactive robot, even though both of the cobots are rated high ($\mu_{proactive} = 4.50$ out of 5.00, with $p = 0.004$ as shown in Table 4(c)). More positive teammate traits and a higher trust may already indicate a higher perceived collaboration for the participants. As a supporting statement, the participants think that the reactive robot was significantly more competitive in its behaviors, whereas the ratings for the proactive robot were below average (with $p < 0.001$ in Table 4(c)). However, since the partners share a mutual goal, competition is not encouraged for better team performance. Finally, the participants affirm that they feel more comfortable with the proactive robot ($p = 0.003$) and they are more pleased collaborating with it ($p = 0.011$), all pointing out a higher perceived collaboration for the proactive robot.

Many of the positive traits of the proactive robot mentioned is a result of better human adaptation skills, which is hard to directly observe and evaluate in HRI in general, especially when it requires reasoning about the hidden human states in a static environment. In simulation experiments, we are able to track the accuracy of the belief estimation over human states since the ground-truth information of the human states is available through the simulated human decision models (in Figure 5). In our previous study, we find that the average estimation accuracy decreases proportionally when the frequency of unexpected human behaviors increases [16]. In real-world experiments, the ground-truth information of whether a cobot has accurately

anticipated a human's state and adapted to the human accordingly is not explicit and can only be known to the interacting human. Hence, we ask the participants to broadly evaluate. In Table 4(b), we show that the participants think the proactive robot was able to adapt significantly better to their assistance needs (with $p = 0.004$) whereas the reactive robot behaves more repetitively rather than responding to their changing behaviors ($p < 0.001$). The sum of scores in Table 4(b) (Likert scale on overall adaptation capability) also significantly favors the proactive case. Finally, at the end of the experiments we asked the participants two direct technical comparison questions where the robots are renamed anonymously (see comparison category in Table 3). For the first question, the vast majority of the participants,i.e., 71.4%, picked the reactive robot as the cobot that is following preset rules instead of responding to their changing needs and preferences. For the second one, 78.6% of the subjects picked the proactive robot as the one that was learning and adapting better to the participant's assistance needs. From these statements we conclude that our A-POMDP model adapting to a human's unanticipated behaviors shows better adaptation skills and it has a higher perceived collaboration, trust, and positive teammate traits, than a cobot model that does not handle such behaviors, which supports **Hypothesis 3**.

With this experiment, we show the negative impact of the unanticipated human behaviors on the fluency of a collaboration, which are mostly overlooked in HRC studies. Despite the diverse backgrounds of the participants, their statements show great consistency, suggesting that the unexpected cobot interference occurs mostly due to the wrong anticipation of and adaptation to the current behavior of a person, which then results in a significantly less efficient and natural collaboration. Since our robot decision model is a POMDP, it is not learning from the history of interaction but it was able to reach to complex conclusions about the human states. For instance, referring to Figure 2, when our robot observes *a subtask succeeded* several times, it is very likely to anticipate that the state of *Human is not struggling* for a participant placing the cubes correctly. In that case, if the participant is observed to be idling for a long time (inactivity), the robot is still likely to anticipate that the human will take care of the job, and so it does not interfere. In other words, the robot concludes that "the human may still be assessing the subtask (after a long wait) as she has mostly succeeded so far". If this inactivity of the participant continues or if it is followed by an observation of *a subtask failed*, the likelihood of transitioning to the state of *Human may be tired* significantly increases. After such a belief update, the robot is likely to offer assistance and act on the task. The robot's belief translates into: "A longer wait may indicate tiredness after a decrease in her performance. I better assist her with the task". Such conclusions are reached due to the probabilistic distributions over our state machine design with multiple anticipation stages (see Figure 2). With this experiment and analysis, we validate our cobot's extended short-term adaptation skills on a real setup, which is in line with our simulation experiments in [16].

We are aware that the same intrinsic parameters of our A-POMDP design may not respond reliably to all human types. For that, our first attempt was to incorporate different human types into our A-POMDP design. After we optimize for an online POMDP solver, the model in its current design (in Figure 2) takes approximately *1 second* to respond to each observation in real-time (see in Section 5.2.2). However, incorporating various human types in the same model results in more complex transition and observation functions. The robot's responses have become significantly inaccurate with a response time of approximately *3 seconds* for each action decision. When we deploy the model for early experiments, many of the tasks could not be completed (e.g., due to a participant's frustration) or completed with significantly lower success rates. In the end, the POMDP model has become inapplicable and the experiments have been inconclusive. That is, we could not generate any quantitative results. Therefore, instead of modeling human types as a latent variable in a single model, our intuition is to design a separate Bayesian inference procedure for the human states that change less frequently (i.e., the human types), whereas the A-POMDP handles

only more frequent human dynamics. For that, the adaptive policy selection mechanism provides a faster and further destabilization of the robot traits by selecting (not learning) different policies according to the accumulated knowledge in the long-term. The next section discusses this long-term adaptation mechanism.

6.2 Evaluation of Long-term Adaptation and Complete Framework

In this section, we evaluate the extended long-term human adaptation and the integrated performance of FABRIC, i.e., the full system with the *adaptive policy selection* component in Figure 1. In this experiment, our goal is to support the hypotheses:

Hypothesis 4. *Our collaboration setup and the task induce changes in long-term human characteristics, such as their expertise and collaborativeness.*

Hypothesis 5. *Our framework with A-POMDP models and the ABPS mechanism (the full FABRIC) provides a personalized, fast, and reliable adaptation to both short- and long-term changing human behaviors and characteristics, while it is perceived to have high collaboration skills, positive teammate traits, and trust.*

As mentioned in Section 5.4, the base model used in creating the policy library is the proactive robot model from the experiments in Section 6.1. This has already led to more natural and efficient collaboration in a challenging work environment. However, we have observed dynamic characteristics and preferences in the participants during the experiments. Selecting different policies would provide even better adaptation toward more personalized collaboration. In this experiment, our first goal is to show that the ABPS mechanism is able to provide this adaptation fast and reliably, improving a cobot's collaboration performance. Also, we want to prove the effectiveness and applicability of our complete system in a real-world scenario.

6.2.1 Objective and Subjective Measures. We evaluate the long-term adaptation capability of the system by highlighting the long-term differences in human behaviors and the cobot's ability to detect and respond to such changes. We use the same objective measures as in Section 6.1.1, which are the overall task success rate, the number of warnings received from a participant, a human's success rate, a human's contribution to the overall success and the task efficiency, but we also analyze their changes in time over the course of a collaboration for the long-term effects. In addition, we calculate the regret for a selected policy which denotes the distance of a total discounted reward collected in a task from the maximum utility, i.e., a discounted reward that can be obtained by the best policy for the current human type. The subjective measures are obtained from the questionnaires that are given to the participants (in Table 5). We use some of the statements from the previous experiments with minor additions to analyze the participant statements over time. This gives more insights on the dynamics of the participants' changing stamina, motivation, and perceived difficulty of the tasks, and how they perceive the cobot's collaboration skills over time.

6.2.2 Experiment Protocol. We invited 11 people, who again had no previous interaction with a robot, and are from different backgrounds and demographics (ages between 18 and 35). As mentioned, our purpose is to show the effectiveness and reliability of our complete framework, along with its long-term adaptation capability. We follow mostly the same protocol as in Section 6.1.2). The only difference is that we do not run and compare different cobot decision models as in the previous experiment but only the complete framework. However, the ABPS mechanism may select a different decision model for a task; therefore, the participants are asked to compare the cobot performance between the tasks, where possible, without knowing that its strategy may change. Each participant works on eight tasks in a row and fills out a survey evaluating the performance

Table 5. Subjective Statements Asked during the Experiments

Questionnaire Statements	Category
"The task was challenging for me."	Unanticipated behaviors
"It was exhausting to remember all the placement rules."	
"I lost my attention during the task."	
"I had to warn the robot a lot."	
"The robot was able to anticipate my assistance needs better than in the previous task."	Trust in the robot, the naturalness of the collaboration and the reliability of the robot
"I trust the robot with the task."	
"The robot acted as I expected."	
"The robot took over at the right times when I needed assistance."	
"The robot behaviors seemed more natural (human-like coordination) over time."	Perceived collaboration and teammate traits of the robot and their change over time (**General**)
"The robot showed more positive teammate skills over the course of the experiment."	
"Our coordination with the robot got worse over the course of the experiment."	
"I became more comfortable interacting with the robot over the course of the experiment."	
"The robot was able to adapt to my assistance needs over the course of the experiment."	
"I would work with such a robot on such kind of tasks in a working environment."	
"The robot was learning and adapting to my assistance needs."	True/false questions (**General**)
"The robot showed positive teammate skills and a high-level of coordination."	

of each task right after it is completed. Then, at the end of the experiment they answer another survey for the general statements (in Table 5). We use task type-5, demonstrating the Stroop effect [52] in Figure 7, throughout the experiment (for all eight tasks) to ensure the task conditions remain constant. This task invokes the learning effect and the participants could gain noticeable expertise within an experiment, which normally requires more practice. At the beginning of the experiments the participants did not know the tasks had the Stroop effect, and it took them several tasks to notice and master it. In addition, each experiment takes approximately 1.5 hours to complete; hence, we expect to observe accumulated tiredness or a decrease in motivation.

6.2.3 Results and Discussions. We first analyze the subjective evaluations of the participants on their perceived task difficulty, attention, exhaustion, and collaborativeness. Even though the difficulty and the length of a task always remain the same, the participants have perceived the tasks as being easier and less exhausting over time. This is in line with our measurements from Figure 14(b) visualizing the increasing trend of the human success rates and the successful contribution of the human in a task. We deduce that the participants gained more expertise and got used to the task and the environment, which indicates the practice-effect that also affects their perceived difficulty of and the exhaustion in a task. The collaborativeness of the participants has also changed during the experiments. Figure 14(a) demonstrates that the participants warned the cobot less over time whereas Figure 15(a) shows that their trust in the cobot increased, both indicating that they become more collaborative in time. That said, their expertise, collaborativeness, stamina, and motivation (i.e., through their handling of more subtasks) do change over time and that a cobot should adapt to them. We conclude that our assumption about the change of human characteristics is valid and that we could invoke and observe this during our experiments, which supports Hypothesis 4.

We expect the ABPS mechanism to select a different policy when, for instance, a participant is more motivated to complete a task herself due to her increased expertise. In that case, the cobot should give more space to a human to finalize a task. For example, an A-POMDP policy needs to be selected that less favors a state transition from "human is not struggling" to, e.g., "human may not be capable" (in Figure 2). As another example, some of the participants built up more trust in the cobot resulting in leaving a task more often to the cobot. Such drastic and stochastic behavioral changes are difficult to model in a single decision-making strategy. In Figure 16, we demonstrate how ABPS has responded to such changes. Figure 16(a) gives the total discounted rewards the

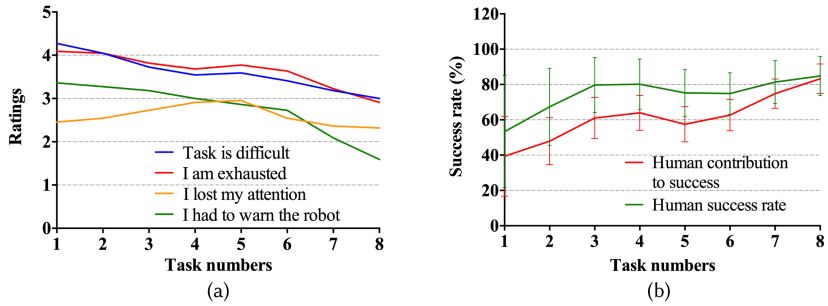


Fig. 14. (a) The subjective ratings of the participants on their cognitive load and trust in the robot over time. (b) The participant's moving averaged success rates and their contribution to success over time.

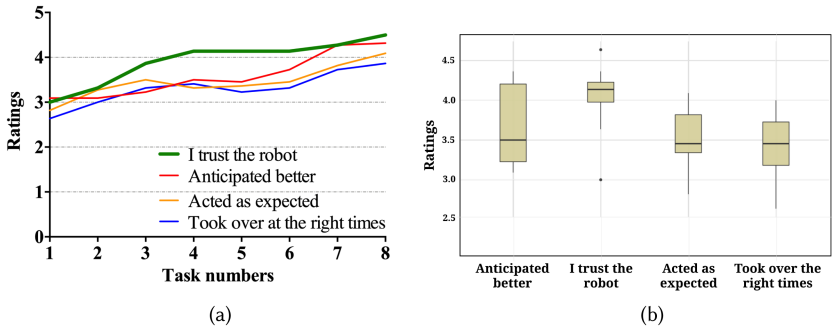


Fig. 15. The subjective evaluations of the participants to the statements of: "The robot was able to anticipate my needs and behaviors better than in the previous task.", "I trust the robot.", "The robot acted as I expected." and "the robot took over at the right times when I needed assistance.". (a) shows the change of the participant ratings over the task assignments, (b) is the box and whisker plots of the overall ratings.

cobot has collected over time. In the same figure, we also show the average rate of ABPS selecting a different policy for the new task to start (i.e., policy change rates as data points). For instance, before the second task starts, ABPS has picked another policy in 45% of the experiments, according to its current estimate of that human type. With that, at the end of the second task the collected rewards on average have been almost doubled compared to the first task. As another example, the number of warnings (in Figure 16(d)) increased at the fourth task with the increasing human contribution (i.e., expertise) as in Figure 14(b). As a reaction, ABPS has picked another strategy at the fifth task in 54% of the experiments, mostly leaving the task to the participants. This has successfully dropped the number of warnings by 20%; however, the participants have reached a less success rate on average (in Figures 14(b) and 16(c)). Then, the cobot has picked another policy at the sixth task, which has balanced the collaboration better and resulted in 14% increase in the task efficiency (in Figure 16(e)). This shows the significant positive effect of the ABPS mechanism.

Considering the plots in Figure 16, by the fourth task, ABPS has found a policy that is nearly optimal for the interacted humans, when averaged over all of the experiments. In general, we can say that a mutual steady-state has almost been reached after the third task where both the human and the cobot have more or less satisfactory coordination (see the stabilizing curves in the plots). This is also stated by the participants through their almost stabilized ratings to, for example, "I trust the robot" in Figure 15(b) and their contribution to success in Figure 14(b) after the third task. We can see the same effect also on the task efficiency in Figure 16(e). We note again the efficiency

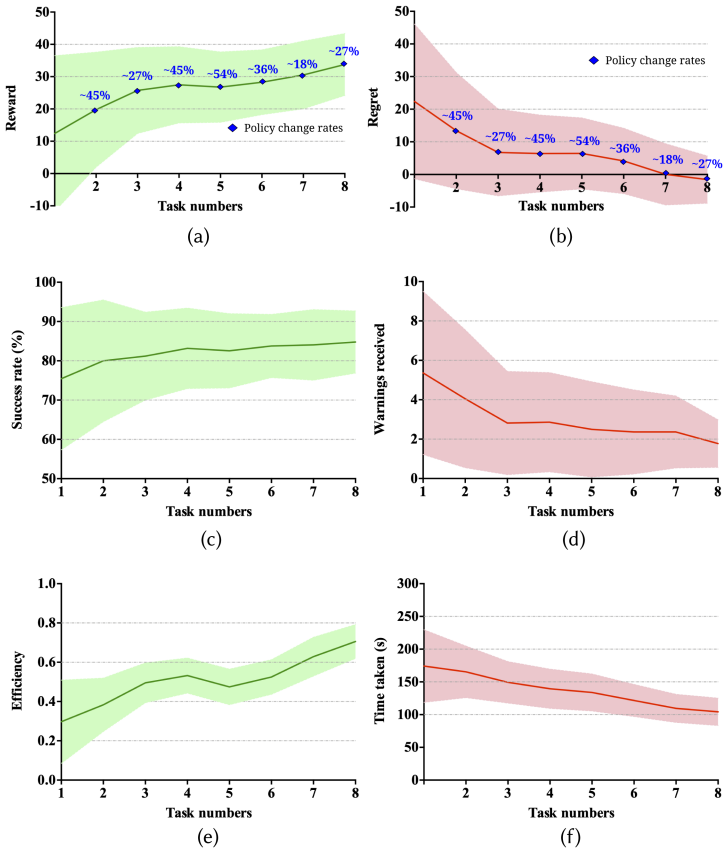


Fig. 16. The first two plots give the performance of ABPS: (a) moving average total discounted reward over task assignments collected by the cobot; (b) moving average regret over task assignments with error bars denoting the standard deviation, collected by the cobot. The policy change rates on these two figures denote the average rate of the cobot selecting a different policy before the task starts, whereas the reward and the regret values on the data points are collected at the end of that task. The others show the overall performance of the cobot averaged over the task assignments: (c) moving average of the task success rate; (d) number of warnings received by the cobot; (e) moving average of the task efficiency as in Section 6.1.1; (f) moving average of the task durations in seconds.

drop at the fifth task; however, this is also compensated at the next task by the cobot. The constant increase in the efficiency after the sixth task can be explained by the increasing human expertise (in Figure 14(b)) and the cobot's successful adaptation to it. This adaptation is mutual, where both the human and the cobot have finally achieved higher levels of coordination. The cobot first selected a strategy (exploration) that has left almost the entire task to the participants at the fifth task. But eventually it could find strategies that mostly leave the tasks to the participants and take over the tasks only if the human failed consistently or idled for longer. The latter was often observed after the participants discovered when the cobot usually takes over and so they were intentionally waiting for the cobot to pick up when they do not remember the rules, which is another example of the mutual adaptation.

Finally, in Figure 16(b), we show the performance of ABPS also through the change of the regret values over time. Since the actual human types are unknown to the cobot and to us, for the regret

Table 6. Average Ratings of the Participant Answers to the General Statements Asked at the End of Each Experiment

	General Statements	Ratings
Mean (likert)	Robot showed positive teammate traits	4.1
	Coordination with robot got worse	1.8
	I became more comfortable	4.3
	Robot was able to adapt	4.3
	I would work with such a robot	4.1
True/False	Robot was learning and adapting	Agreed: 91 %
	Robot showed a high-level coordination	Agreed: 82 %

calculation we approximate it by taking the average of the maximum discounted rewards collected by the policies in the policy library during the training phase run on the simulation environment (see Section 5.4). Hence, we observe negative values of regret at the last task in Figure 16(b) since the rewards collected at this task were more than this assumed maximum utility value. Overall, the moving average regret has reached an almost steady-state after the third task. Then, thanks to the increasing collaborative success with the human expertise and the cobot's adaptation to it with a different policy that received less number of warnings, the regret values have drastically decreased after the fifth task. All in all, such human characteristic changes have been observed between the first and the third tasks and between the fifth and the seventh tasks, which shows that ABPS has gradually adapted to them despite their constant change throughout the experiments. We note that the response of ABPS is fast but it could be faster with more accurate human type estimation. Since the actual type of a participant is unknown to the cobot and to us at the beginning of an experiment and since it dynamically changes throughout the experiments, we do not know the true optimal robot policy or policies for the participant in hindsight. Therefore, it is not possible to objectively measure the convergence rate for ABPS. In our previous study [15], we conducted experiments against simulated human models with known types and the best performer robot policy in hindsight. The convergence to the optimal policy has been found to be around the 6th iteration (i.e., the 6th task); nonetheless, the value depends on the duration of a task (i.e., the collaboration scenario) and the number of human observations collected. Most importantly, we show that the system is reliable as it has continuously contributed positively to the overall task efficiency and human satisfaction, i.e., naturalness, as shown in Figure 15.

In addition to the performance of ABPS, we also evaluate the collaboration skills and teammate traits of our cobot perceived by the participants, and their trust in the cobot. In general, the better anticipation of a person's assistance needs, the correct timing in assisting, the expected and reliable cobot responses, and an increasing trust suggest a higher acceptance for the cobot over the course of the experiments. We see in Figure 15(a) that, on average, the participants were satisfied with the performance of the cobot throughout the experiment with the constantly increasing ratings of the given statements, reaching up to the levels of 75% of the maximum rating. In particular, the trust level has reached over 90% of the maximum rating by the end of the experiments. Our cobot could provide positive teammate traits through anticipating and respecting its partner's needs and preferences, contributing to the increased performance of its partner and leading to more efficient task completion. Supporting this, the participants evaluated that the cobot showed positive teammate traits, it was able to adapt to the changing human needs and preferences, they became more comfortable with the cobot, and they would work with such a cobot with the mean ratings over 4 out of 5 as given in Table 6. In addition, 91% of the participants agreed that the cobot was learning and adapting, and 82% of them confirmed that the cobot has high coordination skills.

All these objective measures and the positive subjective statements of the participants support **Hypothesis 5**.

7 CONCLUSION

In this article, we focus on an extended human adaptation of collaborative robots (cobots). We propose solutions for the following research challenges: How to anticipate and adapt to unanticipated human behaviors in short-term; how to handle a diversity of long-term human characteristics for a personalized collaboration; how to develop accurate human models that simulate a great diversity of human behaviors in work environments; and how to design a real collaboration experiment that does not assume turn-taking, does not constrain human intention space, and invokes unanticipated human behaviors to properly evaluate our robot's adaptation goals. With the purpose of obtaining an integrated system, we devise our novel lightweight autonomous framework, called FABRIC, that hierarchically integrates our approaches to the challenges above.

To design and evaluate our framework, we propose a pipeline that trains and runs rigorous tests in simulation, then improves and deploys the solution in the real-world for user studies. We first design a novel simulated human model and a 3D factory environment that samples such dynamic human behaviors. We are aware of the possible biases in the experiments which could be introduced by the simulated humans due to the limited size of the action space of these humans and the fact that they are hand-coded. We stress that these are not necessarily accurate models; however, the abstracted states in our design are shown to be observed in real humans through the calibration experiments. On the other hand, we also show that the simulation provides even more diversity than the user studies thanks to its scalable behavior sampling. This and the calibration experiments have proven the necessity of our human simulations in training and for rigorous tests. We then transfer the simulation results into the real-world through our novel experiment setup and our collaboration task that together induce cognitive load on the human participants. We show that we could observe a variety of human responses and preferences, including human behaviors that lead to mistakes and changing human expertise.

For the short-term adaptation of FABRIC, our approach is our novel A-POMDP model design adapting to a human's changing *intent, attention, tiredness, and capability*, to better estimate if the human needs help and whether the cobot should intervene. Our first round of user studies have shown that handling such human variability increases the overall efficiency and the naturalness of an HRC, also leading to more positive teammate traits and higher trust as perceived by the participants. For the long-term adaptation, we introduce our novel adaptive Bayesian policy selection, ABPS, that runs on top of several A-POMDPs with distinct intrinsic parameters. Toward a personalized collaboration, ABPS selects a model according to an estimate of a human's workplace characteristics, we call types, such as her levels of expertise, stamina, attention, and collaboration preferences. We conduct another user study that deploys our complete framework, FABRIC. At first, we have shown that ABPS provides fast and reliable policy selection in adapting to unknown and changing human types. Then, the objective and subjective results have demonstrated that FABRIC is able to reliably operate in our dynamic environment that does not follow a turn-taking collaboration by fluently coordinating with humans thanks to its extended human adaptation. FABRIC has provided significant improvements in a cobot's human adaptation toward a more natural and efficient collaboration with high perceived teammate skills and trust. We believe that such an adaptation will positively contribute to the acceptance and the long-term use of cobots.

7.1 Limitations and Future Works

Our human type estimation is not accurate enough for some cases, such as, when estimating whether a human is tired or a beginner after she idles longer. We are aware that it is a nontrivial

task to differentiate such hidden human states. Nonetheless, we believe that our assumption of a limited human action space is also a limiting factor for this. Our simulated humans are also modeled with the same assumption; hence, some simulation runs have generated quite similar behavior patterns for certain human types in the long-term. In this study, since our goal was to highlight the importance of handling the unanticipated human behaviors, the cobot has responded similarly to such cases by offering more assistance. Hence, we have not observed a significant impact of this inaccuracy. However, during the user studies we found that an accurate estimation and more tailored handling of such behaviors would result in even more efficient collaboration. In the future, we believe that more data should be collected from larger scale long-term user studies to model the long-term traits of humans and to generate more accurate human simulations.

During our long-term experiments, some participants have gained enough experience over time and perfectly achieved the tasks alone. In such cases, they started to question the necessity of the cobot. To better analyze our cobot's contribution when the human characteristics change, we need to repeat our experiments that compare the human-alone performance with the cobot collaboration, also for the long-term. Two participants also added that calling for cobot assistance only when they want to would yield more efficient collaboration. For that, we may experiment on a command and control system to compare it with FABRIC's performance in the long-term. This can also help us analyze the additional cognitive load such a command system may put on the human operators. Additionally, further validation of the contribution of ABPS to the collaboration is needed through a more comprehensive comparison of it with the best policy in hindsight running alone. Finally, we can further improve our system's reliable adaptation and response time by developing a policy change mechanism even during a task.

To assist with our system successfully scaling with the increasing complexity of tasks, the states and actions specific to a task are not included in our decision models. For broader applicability of FABRIC, we keep the human adaptation mechanisms agnostic to the task details themselves. The *decisional level* of FABRIC only monitors the progress of a task and the human states, and plans for adaptive assistance and task allocation. A decision model only needs to know the progress of a task (i.e., success or failure), which are abstracted in the *functional level* of the framework. Similarly, task related motions and control are tailored under the *functional level* for a specific application. For more complex scenarios, e.g., non-concurrent tasks, we believe our system is applicable through running parallel decision models (e.g., for three simultaneous tasks, three instances of the same selected policy may run in parallel to track and regulate each task separately). That said, one direction of future work is to deploy the system on a larger scale industrial setup with more complex tasks to prove its applicability for a broader impact.

APPENDICES

A DEVELOPING, TRAINING, AND TESTING OUR FRAMEWORK

Here we provide an overview of the development, training and testing phases of our framework by following the pipeline steps. Figure 17 summarizes all activities with their chronological order and also indexes where these steps are described in this article.

B VALIDATING THE SIMULATED HUMAN MODELS

In this section, we validate the reliability of the simulated human models by analyzing and comparing their generated human observations with the real observed behaviors from the humans. Our goal is to show that our novel simulated MDP human models generate reliable human behaviors. Nevertheless, we note that not all of the behaviors generated from this modeling scheme are necessarily accurate representations of the real human behaviors since they are sampled with a

Pipeline Steps	Completed activities	Index	Details
Step (1)	Anticipatory Decision-Making component of FABRIC	Section 3.2.1	A proactive model, i.e., A-POMDP base model, and a reactive model to compare
	Sensing and Actuator components of FABRIC for simulation	Section 5.1	Simple Human Activity Recognition (HAR) module and robot actions in simulation
Step (2)	Human decision models and a 3D simulated factory env.	Section 5.1.1	To simulate human behaviors for a pick & place task on an assembly line
Step (3)	Training robot decision-making models	Görür et al. (2018)	both proactive and reactive models are run against many different and randomly generated simulated human behaviors
Step (4)	Simulation experiments: HRC with 1) proactive robot, 2) reactive robot		
	Our extended short-term human adaptation validated on simulation		proactive robot yields to more efficient and natural collaboration
Step (5)	Improved robot decision-making models with the feedback from Step (4)	Section 3.2.1	**FABRIC decisional-level beta version**
Step (1)	Decision Model Library component of FABRIC	Section 3.2.2	i.e., several A-POMDP models, each of which handling a unique human type
	Policy Selection component of FABRIC		ABPS deployed over the Decision Model Library
Step (2)	Improved human decision models (v2) and extended task for more work load	Section 5.1.1	Various human types and more responsive human models. Multiple pick & place task that spans a longer time
Step (3)	Training the policy selection component (ABPS)	Görür et al. (2019)	Against various simulated human behaviors and types.
Step (4)	Simulation experiments: HRC using FABRIC with and without Policy Selection		Against sampled various long- and short-term human behaviors
	Our extended long-term human adaptation validated on simulation		FABRIC with the policy selection yields to more efficient and natural collaboration than FABRIC with decisional-level beta version (only with single A-POMDP)
Step (5)	Improved policy selection component with the feedback from Step (4)	Section 3.2.2	**Complete FABRIC beta version**
Step (6)	An experiment setup also invoking unanticipated human behaviors	Section 5.2.1	E.g., distracting human attention with a distributed setup
	A collaboration task for the real setup to induce cognitive load on humans		Pick & place different colored objects with complex rules to induce cognitive load
Step (7)	Sensing component of FABRIC for the real world experiments	Section 5.2.2	Real-time HAR based on tracked human hand and head gestures, synchronize robot decisions with human responses
	Manual adjustments on the Decision-Making model (A-POMDP)		Toward the new task design and for a real life interaction capability (synchronizing with real humans)
	World model for FABRIC		For a safe and fluent physical human interaction with human proxemics considered
Step (8)	Calibration experiments with 8 people against FABRIC	Section 5.3, Appendix 8.3	Validating the setup and the task design toward invoking unanticipated human behaviors. Validating the robot's expressiveness. Testing the A-POMDP model and the Sensing component against real humans in real-time.
	Our novel experiment setup validated		
	Improved Sensing component and the World Model of FABRIC for the real world (v2)		Synchronization of human actions and robot responses: trigger mechanism for robot decision update
Step (9)	Training robot decision-making models (v2) and the Sensing Component	Section 5.4, Appendix 8.3.2	proactive (A-POMDP) and reactive models are trained with the real subject data from the calibration experiments and the simulated human data. Sensing component is improved for real-time HAR.
Step (10)	Evaluation metrics for analyzing the short-term adaptation	Section 6.1.1	Objective and subjective analyses toward efficient, natural, trusted, and fluent collaboration
Step (11)	Introducing the participants to the experiment and practicing the collaboration	Section 6.1.2	Getting familiar with the task, the setup, and the robot and its actions
Step (12)	User Studies-1: Extended short-term human adaptation (A-POMDP) validated	Section 6.1.3	Within subject experiments with the proactive and reactive robots, running FABRIC without the policy selection. Comparative analyses prove our extended short-term adaptation. **(FABRIC decisional-level alpha version)**
Step (8)	Decision Model Library component of FABRIC (v2)	Section 5.4	Several validated A-POMDP models generated
Step (9)	Training Policy Selection component of FABRIC (v2)		ABPS is trained using the collected human data from the previous experiments and the simulated human data for diversity
Step (10)	Evaluation metrics for analyzing long-term adaptation and the complete FABRIC	Section 6.2.1	Objective and subjective analyses toward the efficiency, naturalness, trust, adaptation, and the fluency of collaboration.
Step (11)	Introducing the new participants to the setup and practicing the collaboration	Section 6.2.2	Getting familiar with the task, the setup, and the robot and its actions
Step (12)	User Studies-2: Extended long-term human adaptation and the complete FABRIC are validated	Section 6.2.3	Proving the necessity of the policy selection and the applicability of the framework. FABRIC is perceived with high collaboration skills, positive teammate traits and leads to a fluent collaboration. **(Complete FABRIC alpha version)**

Fig. 17. Our pipeline for integration, deployment and the evaluation of an anticipatory collaborative robot with extended human adaptation (FABRIC). The layers reflect how the components in our FABRIC framework (in Figure 1) are integrated and deployed first in simulation then on a real setup, following our pipeline in Figure 3.

random factor to increase the variety of the behaviors. In fact, we believe that such a single human decision model that reflects the greater diversity of human behaviors may not even be possible. In the literature, there exist some abstracted categorizations of such behaviors that reflect some of the possible human characteristics. This is inherited also in our model design, where we abstract the human states in a work environment into the certain intention and behavior sets, which are inspired by the literature. In particular, we are indifferent to the real motivations behind the states the humans are in, which can have infinitely many possibilities (see Figure 5). Hence, the generated human behaviors from our simulated models are nothing but the sampled reflections of these abstracted states, and the diversity is reached through a random walk between them.

Our approach to validate the human models is to compare the generated observations from the simulated models and the real human observations collected during the evaluation of short-term adaptation in Section 6.1 (i.e., from 14 participants). As we discussed before, we expect the simulated models to generate much more diversity of human behaviors and characteristics than the ones we have observed from only 14 people during the user studies. It is also possible that some of the observations from the generated models do not reflect a real scenario; however, the

Observations generated from	Likelihood of the obervations generated from: Model-1	Model-2	Model-3	Model-4	Model-5	Model-6	Model-7	Model-8
Model-1	0.227	0.226	0.178	0.126	0.087	0.096	0.129	0.078
Model-2	0.205	0.223	0.155	0.113	0.072	0.085	0.105	0.063
Model-3	0.187	0.179	0.215	0.091	0.047	0.039	0.060	0.044
Model-4	0.176	0.171	0.108	0.212	0.042	0.036	0.048	0.039
Model-5	0.128	0.129	0.088	0.078	0.185	0.102	0.044	0.058
Model-6	0.132	0.133	0.090	0.079	0.066	0.193	0.043	0.058
Model-7	0.176	0.190	0.186	0.144	0.083	0.052	0.228	0.124
Model-8	0.107	0.101	0.083	0.068	0.041	0.033	0.049	0.172

(a)

Observations collected from	Likelihood of the obervations generated from: Model-1	Model-2	Model-3	Model-4	Model-5	Model-6	Model-7	Model-8
Participant-1	0.208	0.146	0.087	0.082	0.081	0.083	0.110	0.061
Participant-2	0.141	0.132	0.082	0.075	0.071	0.072	0.122	0.058
Participant-3	0.226	0.189	0.094	0.094	0.100	0.102	0.062	0.062
Participant-4	0.165	0.153	0.098	0.087	0.082	0.083	0.174	0.067
Participant-5	0.164	0.154	0.080	0.080	0.085	0.086	0.063	0.055
Participant-6	0.143	0.132	0.077	0.073	0.069	0.071	0.095	0.051
Participant-7	0.133	0.126	0.208	0.080	0.072	0.073	0.197	0.072
Participant-8	0.128	0.119	0.066	0.065	0.064	0.065	0.068	0.045
Participant-9	0.138	0.132	0.078	0.073	0.076	0.077	0.097	0.057
Participant-10	0.133	0.127	0.086	0.077	0.072	0.073	0.137	0.063
Participant-11	0.153	0.143	0.116	0.085	0.079	0.080	0.166	0.069
Participant-12	0.127	0.122	0.071	0.068	0.071	0.073	0.083	0.052
Participant-13	0.124	0.118	0.076	0.069	0.068	0.069	0.115	0.056
Participant-14	0.175	0.201	0.088	0.084	0.083	0.162	0.090	0.057

(b)

Fig. 18. The likelihoods of: (a) the simulated observations generated from the same simulated models; (b) the participant observations generated from the simulated models. The darker blue the cells get, the higher the probability is.

models we have used throughout the thesis are run and tested on our 3D simulated environment several times to make sure that they do not always exert unreliable human behaviors, e.g., a human is tired at the very beginning, a human constantly fails and does not let the robot take over. This is configured by manually tuning the decision models. To prove that the models are able to reflect real-life scenarios, we calculate the likelihood of an observation set being generated from our simulated human models. Each task starts with a task assignment and ends with either a global success or a global fail for both the simulated human models and real humans. Also, the action sets are confined within the work environment; hence, both of the real and the simulated humans generate an observation of interest (abstracted observations for the robot) from the same observation space (in Section 5.2.2). Because of that, the generated observation sets differ from each other in terms of the sequence of the executed human actions.

We first take each of the human actions observed in a task one by one sequentially (that is, with the frequency of observation update as in Section 5.2.2), and calculate the probability of the action generated by the human states in our model design. This gives us a belief distribution on the current human state. For simplicity, we call it “action belief”. Afterward, we calculate

the current belief starting from the initial belief distribution of a simulated human model, and using the action taken and the state transition probabilities of the human model. This gives us an estimate of the actual belief state that the human model would be in if the generated action was taken from it, we call it "current belief". The multiplication of the action belief and the current belief gives us the likelihood of that action generated by the model. We keep multiplying the likelihood values for each observed action in an observation set until a task ends (i.e., until the end of the set). Then, we average the likelihood values obtained from each task to obtain one value for comparison. We repeat this for all of the eight simulated human models we have created and used in our previous experiments, and for all of the real observations obtained from 14 participants. We give the resulting likelihood values of the simulated observations being generated from the same simulation models in Figure 18(a), whereas in Figure 18(b), we show the resulting likelihood values of the real observations emitted by the participants being generated from the simulated human models. The main reason behind the calculation of the likelihoods of the simulated observations is to have a ground truth, i.e., to know what is the best likelihood value that can actually be obtained given the randomness and all possible state transitions in the simulated humans.

Figure 18(a) shows the distribution of these values over each model. The largest likelihoods are all the diagonal values, which states that the generated observations from each of the models resemble the best likelihoods as expected. That means despite the random walk, the generated actions still reflect the intrinsic characteristics of their models. Yet the values are very small, ranging between 0.17 and 0.23. This shows the randomness of our MDP state sampling. Our belief is that this would give enough diversity even within the same model itself, which reflects the real human behaviors better and which is also our goal to observe a broader range of human dynamics. In Figure 18(b), we show the same calculation, this time for the participant observations. For 14 participants, we observe that the maximum likelihood values are also in the range of 0.14 to 0.23, where each participant shows a high resemblance to at least one of the models. This indicates that the probability of a real human observation being generated from one of our simulated models is as likely as they are actually generated from that model. Hence, the observations show a great similarity, supporting the reliability of our models generating realistic human behaviors. It is not possible to run variance analysis as the number of the human observations is very small compared to the simulated ones. Finally, as it is visualized in Figure 18(b), the participants mostly show a resemblance only with three or four out of eight simulated models. This supports our idea that the user studies in lab environments are less likely to provide enough diversity of human behaviors for training and testing collaborative robots. Our simulated models, specifically, Model-4,-5, and -8 in our case, contribute greatly to that diversity.

C MORE ON THE CALIBRATION EXPERIMENTS

C.1 Evaluating Cognitive Load of the Tasks

Our collaboration setup aims at placing a cognitive load on the humans and the degree of cognitive load may differ between the task types. As a result of the cognitive load and long working hours, unanticipated human behaviors should be invoked and observed during a task. To achieve this, we ran the calibration experiments as part of a master's thesis study at Technische Universität Berlin [28]. Here, we summarize the experiments and the results. We let eight participants and our cobot collaborate over an extended period (approximately 1.5 hours each) executing our tasks designed for the experiments (see in Figure 7). Our goal is to compare the types and choose suitable ones for the final experiments. Our criteria are: (1) The task is cognitively demanding enough to observe unanticipated human behaviors and (2) the task is easy and motivating enough to keep the human engaged in the collaboration. Hence, we measure the rewards the cobot receives, the

number of cobot interferences to count how many times the cobot has taken over the task, and the number of warning gestures a participant made during a task.

Table 7. Subjective Statements Asked to Evaluate the Cognitive Load

Questionnaire Statements	Category
"The tasks were tedious (boring and monotonous)."	Overall cognitive load
"The task was challenging for me."	
"Over the course of the experiment, the tasks felt more and more tedious." **(General)**	
"A robot collaboration is beneficial/needed in such tasks." **(General)**	
"The tasks became more challenging over the course of the experiments." **(General)**	
"It was exhausting to remember all the placement rules."	
"It became harder to remember the rules over the course of the task."	
"I lost my attention during the task."	
"The robot helped me remember the rules"	
"I became more and more exhausted during the experiment." **(General)**	
"It became harder to focus over the course of the experiment." **(General)**	
"How mentally demanding was the task?"	Task load - NASA-TLX
"How physically demanding was the task?"	
"How hurried or rushed was the pace of the task?"	
"How successful were you in accomplishing what you were asked to do?"	
"How hard did you have to work to accomplish your level of performance?"	
"How insecure, discouraged, irritated, stressed, and annoyed were you?"	

The subjective measures are collected by means of questionnaire responses that the participants complete either after a task is completed or at the end of the experiment. We use a 5-step Likert scale to evaluate each of the statements in the questionnaire (in Table 7). Additionally, the NASA-TLX measures are used to rate the task load induced over the participants on a 20 point scale [18]. It measures the load in six dimensions, namely, *mental demand*, *physical demand*, *temporal demand*, *performance*, *effort*, and *frustration*. In Table 7, the statements marked with the type **General** target the experiment in general, and so they are asked only once after the whole experiment is completed. The other statements are task-specific and they are repeatedly asked after each task completion. Also, the experiments are conducted as a within-subject design to better compare the effect of each task type.

During the experiments, we noticed that the task type-3 and type-5 (in Figure 7(c) and (e), respectively) are particularly difficult for humans, which has led the participants to always leave the tasks to the cobot. Therefore, we drop them from this experiment that cover shorter periods of collaboration on a same task. Each participant collaborates with the cobot on each of the three task types we would like to examine (type-1, type-2, and type-4 as given in Figure 7) three times in a changing order. This ensures that the inspected effects are caused by the difference of the cognitive loads induced by the task types rather than by a possibly heterogeneous participant group composition or the practice-effect the participants may gain throughout the experiment.

We first use a training round to introduce our industrial scenario of a human and the cobot collaborating on an assembly line. We also add that the tasks are initially assigned to the participants, leading to higher efficiency and a score if they are done by them, and the cobot's job is to assist whenever needed. We demonstrate to the participants how to grasp and place the objects and how to interact with the cobot to avoid practice-effect. We also remind them that whenever an object is detected in one of the containers, the participants should wait for the audio feedback indicating that the placement is processed. In total, eight participants interacted with the cobot using the same A-POMDP decision model (i.e., the base model described in Section 5.4). Starting with NASA-TLX rates, since in our experiments the tasks have been designed to be cognitively demanding, our focus is on the mental demand dimension that is depicted in Figure 19. We discard

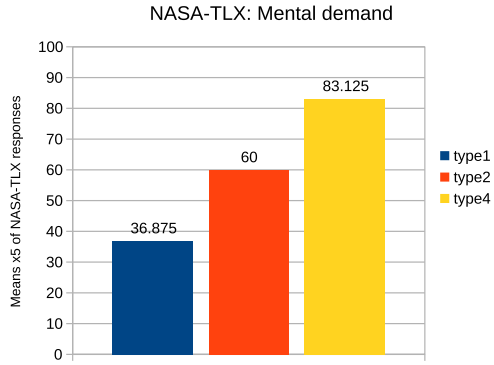
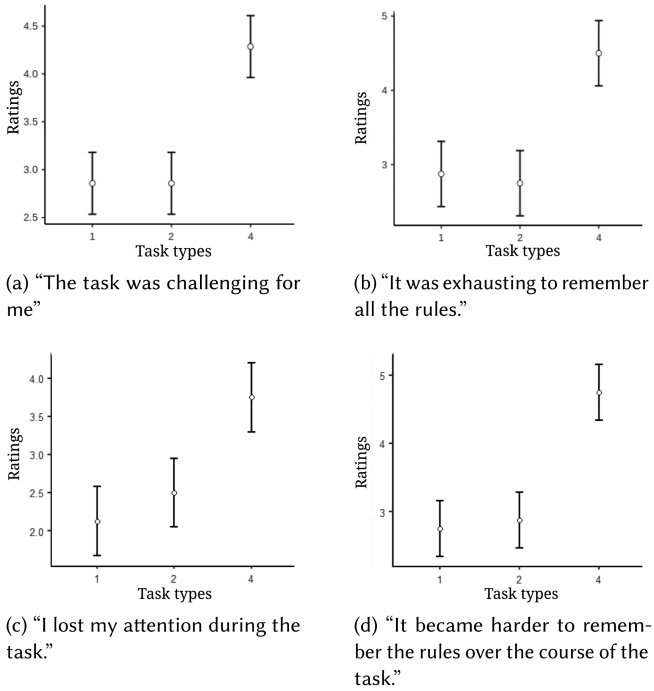


Fig. 19. NASA-TLX scores.



(a) "The task was challenging for me"

(b) "It was exhausting to remember all the rules."

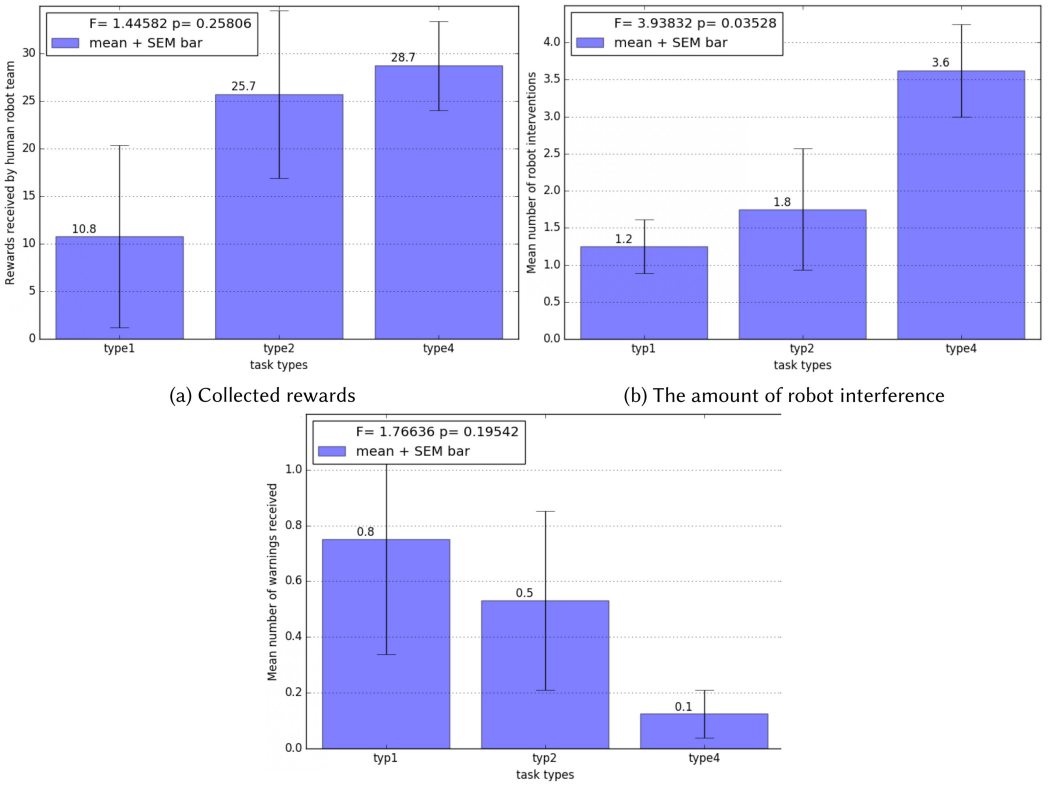
(c) "I lost my attention during the task."

(d) "It became harder to remember the rules over the course of the task."

Fig. 20. Mean plot of the participant ratings in 5-step Likert scale to the statements for the perceived task difficulty and the cognitive load for different task types.

the physical load in our analysis as the participants already stated that they did not feel this. The task type-4 has been reported to require the highest mental workload with a mean score of 83.125 out of 100, while the task type-2 had a score of 60, and for the type-1, it is 36.875.

For the survey responses on the task difficulty, one-way ANOVA results show that the three task types significantly differ from each other concerning how challenging they were perceived by the participants ($p = 0.007$, in Figure 20(a)). Additionally, an η^2 of 0.42 indicates a large effect size. The post-hoc Tukey-HSD reveals that the mean difficulty for task type-4, 4.29 out of 5, is significantly larger than the others ($P_{tukey} = 0.015$), which is in line with the NASA-TLX test. Similarly, task type-4 was significantly more exhaustive and caused more distraction, as



(a) Collected rewards

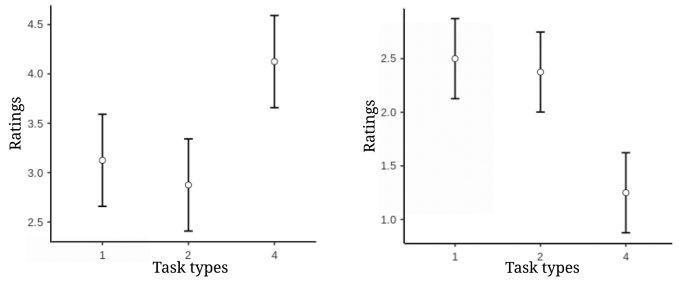
(b) The amount of robot interference

(c) The amount of human warnings

Fig. 21. Quantitative measures averaged over the participants for each task type.

perceived by the participants (in Figure 20). As the task rules are perceived to be difficult, the participants needed to look at the task monitor (in Figure 6) several times during a task to track its current state. This is also recognized by the cobot as distracted since the attention is removed from the work environment. Finally, the participants agree that they became increasingly tired during the experiment for all task types in general ($mean = 3.875$). This indicates that even though a task type is kept the same, it is perceived to be more demanding over time. This perceived difficulty was particularly strong with type-4 (mean = 4.75). From the analysis above, we deduce that type-4 is significantly challenging and induces a cognitive load. In addition, it is well suited for the cases in which the perceived cognitive load on the participants increases over time.

Since we require collaboration, it should not be too overwhelming for the participants, or demotivating to leave the task completely to the cobot. For this purpose, we check the success rates, cobot's interference, and the participants' opinions about the cobot and the reliability of its reactions in a task. First of all, Figure 21(a) compares the received rewards ($mean_{type1} = 10.8, mean_{type2} = 25.7, mean_{type3} = 28.7$) with no significant difference between the rewards the cobot received after each task type ($p = 0.258$). This indicates that, even though type-4 is stated to be the most difficult, the success rates are still higher. This is mainly due to significantly large contributions of the cobot in this task as shown in Figure 21(b). A cobot interference describes a successful take over of the cobot, which is a successful placement of an object without receiving a warning from a participant. In other words, the participants also approve the cobot's assistance



(a) "The robot helped me to remember the rules"

(b) "I would have scored better without the robot"

Fig. 22. Mean plot of the participant ratings in 5-step Likert scale to the statements related to how they perceive the robot's collaboration during different task types.

offer. Most cobot interferences occurred during task type-4 ($mean = 3.6$), which is significantly higher than task type-2 ($mean = 1.8$) and type-1 ($mean = 1.2$) with $p = 0.035$. However, the warning levels were low as given in Figure 21(c). Finally, the participants also agree that the cobot, in general, helped them to remember the rules in task type-4 ($mean = 4.13$), and in type-4 they would have scored significantly worse without the cobot's help than in tasks type-1, $p = 0.027$ and type-2, $p = 0.045$ as shown in Figure 22(b). As a result, it can be stated that task type-4 has led to a better coordination where the cobot could effectively support its human collaborator.

In summary, we conclude that the setup is effective in invoking unanticipated human behaviors. This is clear in task type-4 through the participant responses on not remembering the rules, their lost attention, increasing exhaustion over the course of the experiment, and from very high NASA-TLX scores. In addition, the cobot has also successfully estimated these behaviors by correctly offering its assistance. These analyses show that the experiment setup is able to place a cognitive load on the humans, and the higher the cognitive load is the more unanticipated human behaviors, followed by more human errors. As mentioned, we choose task type-4 as the main collaboration task to evaluate our cobot's short-term adaptation capabilities. For the evaluation of long-term adaptation, we select to work with task type-5, which has the same configuration as type-4 along with an additional Stroop effect (in Figure 7). Even though it was very challenging, type-5 was very suitable for a longer collaboration as it leads to a noticeable change in human characteristics, like expertise through the learning effect (i.e., learning the Stroop effect) as discussed in Section 6.2.3 of the main article.

C.2 Durations of Human Actions

The human action update frequency in Figure 9 depends on the collaboration setup and the task. Hence, we analyze the collected observations from the calibration experiments for the average duration of human actions. As Table 8 indicates, it takes between 3–5 seconds for a human to grasp an object and place it in one of the containers and another 1–2 seconds to return from the container to the conveyor belt, which is the longest human action on average. We, therefore, set the same action timeout to be 3 seconds (see Figure 9), which is the average time needed for an action to be completed. This also makes sure that during the longest action, the observation update informs the decision-making block at least twice, stating that the human is progressing with the task. As soon as the human action changes, it is processed by triggering a response on the cobot decision-making. This timeout has been extensively tested while interacting with the environment and has shown to deliver a good balance between a timely reaction to new observations and reliable responses during long-lasting and continuous interaction.

Table 8. Average Durations of Human Actions while Interacting with the Cobot

Human Action	Duration of the action [seconds]
grasping and placing	3-5
idle times between grasp attempts	1-3
warning the robot	3-4

ACKNOWLEDGMENTS

We would like to thank Guy Hoffman for his valuable feedback, Elia Kargruber, Minh Nghiem Vi, and Güner Dilsad Er for their contributions.

REFERENCES

[1] Sharath Chandra Akkaladevi, Matthias Plasch, Christian Eitzinger, Sriniwas Chowdhary Maddukuri, and Bernhard Rinner. 2017. Towards learning to handle deviations using user preferences in a human robot collaboration scenario. In *Proceedings of the Intelligent Human Computer Interaction*. Anupam Basu, Sukhendu Das, Patrick Horain, and Samit Bhattacharya (Eds.), Springer International Publishing, Cham, 3–14.

[2] R. Alami, R. Chatila, S. Fleury, M. Ghallab, and F. Ingrand. 1998. An architecture for autonomy. *The International Journal of Robotics Research* 17, 4 (1998), 315–337.

[3] Stefano V. Albrecht and Peter Stone. 2018. Autonomous agents modelling other agents: A comprehensive survey and open problems. *Artificial Intelligence* 258 (2018), 66–95.

[4] Antonio Andriella, Carme Torras, and Guillem Alenyà. 2019. Short-term human–robot interaction adaptability in real-world environments. *International Journal of Social Robotics* 12, 3 (2019), 639–657 .

[5] Chris L. Baker and Joshua B. Tenenbaum. 2014. Modeling human plan recognition using bayesian theory of mind. In *Proceedings of the Plan, Activity, and Intent Recognition: Theory and Practice*. 177–204.

[6] Tirthankar Bandyopadhyay, Kok Sung Won, Emilio Frazzoli, David Hsu, Wee Sun Lee, and Daniela Rus. 2013. Intention-aware motion planning. In *Proceedings of the Algorithmic Foundations of Robotics X*. Emilio Frazzoli, Tomas Lozano-Perez, Nicholas Roy, and Daniela Rus (Eds.), Springer, Berlin 475–491.

[7] Andrea Bauer, Dirk Wollherr, and Martin Buss. 2008. Human-robot collaboration: A survey. *International Journal of Humanoid Robotics* 05, 01 (2008), 47–66.

[8] Frank Broz, Illah Nourbakhsh, and Reid Simmons. 2013. Planning for human-robot interaction in socially situated tasks: The impact of representing time and intention. *International Journal of Social Robotics* 5, 2 (2013), 193–214.

[9] E. Calisgan, A. Haddadi, H. F. M. Van der Loos, J. A. Alcazar, and E. A. Croft. 2012. Identifying nonverbal cues for automated human-robot turn-taking. In *Proceedings of the 2012 IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication*. IEEE, Paris, France, 418–423.

[10] Crystal Chao and Andrea L. Thomaz. 2012. Timing in multimodal turn-taking interactions: Control and analysis using timed petri nets. *Journal of Human-Robot Interaction* 1, 1 (2012), 4–25.

[11] Min Chen, Stefanos Nikolaidis, Harold Soh, David Hsu, and Siddhartha Srinivasa. 2018. Planning with trust for human-robot collaboration. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction (HRI'18)*. ACM, 307–315.

[12] Sandra Devin and Rachid Alami. 2016. An implemented theory of mind to improve human-robot shared plans execution. In *Proceedings of the 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI'16)*. 319–326.

[13] Matthew Gombolay, Anna Bair, Cindy Huang, and Julie Shah. 2017. Computational design of mixed-initiative human-robot teaming that considers human factors. *The International Journal of Robotics Research* 36, 5-7 (2017), 597–617.

[14] O. Can Görür and Şahin Albayrak. 2016. A cognitive architecture incorporating theory of mind in social robots towards their personal assistance at home. In *Proceedings of the Workshop on Bio-inspired Social Robot Learning in Home Scenarios at IEEE/RSJ International Conference on Intelligent Robots and Systems 2016 (IROS'16)*.

[15] O. Can Görür, Benjamin Rosman, and Sahin Albayrak. 2019. Anticipatory Bayesian policy selection for online adaptation of collaborative robots to unknown human types. In *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems (AAMAS'19)*. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 77–85.

[16] O. Can Görür, Benjamin Rosman, Fikret Sivrikaya, and Sahin Albayrak. 2018. Social cobots: Anticipatory decision-making for collaborative robots incorporating unexpected human behaviors. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction (HRI'18)*. 398–406.

[17] Elena Corina Grigore, Alessandro Roncone, Olivier Mangin, and Brian Scassellati. 2018. Preference-based assistance prediction for human-robot collaboration tasks. In *Proceedings of the 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. 4441–4448.

[18] Sandra G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (task load index): Results of empirical and theoretical research. In *Proceedings of the Advances in Psychology*. Elsevier, 139–183.
[19] Laura M. Hiatt, Anthony M. Harrison, and J. Gregory Trafton. 2011. Accommodating human variability in human-robot teams through theory of mind. In *Proceedings of the 22nd International Joint Conference on Artificial Intelligence (IJCAI'11)*. 2066–2071.
[20] Laura M. Hiatt, Cody Narber, Esube Bekele, Sangeet S. Khemlani, and J. Gregory Trafton. 2017. Human modeling for human-robot collaboration. *The International Journal of Robotics Research* 36, 5-7 (2017), 580–596.
[21] Guy Hoffman. 2019. *Anki, Jibo, and Kuri: What We Can Learn from Social Robots That Didn't Make It*. Technical Report. IEEE Spectrum.
[22] Guy Hoffman. 2019. Evaluating fluency in human-robot collaboration. *IEEE Transactions on Human-Machine Systems* 49, 3 (2019), 209–218.
[23] Guy Hoffman and Cynthia Breazeal. 2007. Effects of anticipatory action on human-robot teamwork efficiency, fluency, and perception of team. In *Proceedings of the ACM/IEEE International Conference on Human-robot Interaction (HRI'07)*. 1–8.
[24] Steven Holtzen, Yibiao Zhao, Tao Gao, and Song-chun Zhu. 2016. Inferring human intent from video by sampling hierarchical plans. In *Proceedings of the 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'16)*. 1489–1496.
[25] Chien Ming Huang and Bilge Mutlu. 2016. Anticipatory robot control for efficient human-robot collaboration. In *Proceedings of the 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI'16)*. 83–90.
[26] B. Irfan, A. Ramachandran, S. Spaulding, D. F. Glas, I. Leite, and K. L. Koay. 2019. Personalization in long-term human-robot interaction. In *Proceedings of the 2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 685–686.
[27] Qiang Ji, Peilin Lan, and Carl Looney. 2006. A probabilistic framework for modeling and real-time monitoring human fatigue. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans* 36, 5 (2006), 862–875.
[28] Elia Jona Kargruber. 2019. *Toward User Studies in Real-World Settings: Application of an Autonomous Anticipatory Robotic System for Evaluating Human-Robot Collaboration on an Assembly Line*. Master's thesis. Technische Universität Berlin, Berlin, Germany.
[29] Markus Koppenborg, Peter Nickel, Birgit Naber, Andy Lungfiel, and Michael Huelke. 2017. Effects of movement speed and predictability in human-robot collaboration. *Human Factors and Ergonomics in Manufacturing and Service Industries* 27, 4 (2017), 197–209.
[30] Hema S. Koppula, Ashesh Jain, and Ashutosh Saxena. 2016. *Anticipatory Planning for Human-Robot Teams*. Springer International Publishing, Cham, 453–470.
[31] KUKA. 2016. *Hello Industrie 4.0 Glossary*. Technical Report. Kuka AG. Retrieved from https://www.kuka.com/-/media/kuka-corporate/documents/press/industry-4-0-glossary.pdf.
[32] Przemyslaw A. Lasota, Terrence Fong, and Julie A. Shah. 2017. A survey of methods for safe human-robot interaction. *Foundations and Trends in Robotics* 5, 4 (2017), 261–349.
[33] Iolanda Leite, Ginevra Castellano, André Pereira, Carlos Martinho, and Ana Paiva. 2014. Empathic robots for long-term interaction. *International Journal of Social Robotics* 6, 3 (2014), 329–341.
[34] Iolanda Leite, Carlos Martinho, and Ana Paiva. 2013. Social robots for long-term interaction: A survey. *International Journal of Social Robotics* 5, 2 (2013), 291–308.
[35] Huao Li, Tianwei Ni, Siddharth Agrawal, Fan Jia, Suhas Raja, Yikang Gui, Dana Hughes, Michael Lewis, and Katia Sycara. 2021. Individualized mutual adaptation in human-agent teams. *IEEE Transactions on Human-Machine Systems* 51, 6 (2021), 706–714.
[36] J. Li, T. Lei, and F. Zhang. 2018. An gaussian-mixture hidden markov models for action recognition based on key frame. In *Proceedings of the 2018 11th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*. 1–5.
[37] S. McGuire, P. M. Furlong, C. Heckman, S. Julier, D. Szafir, and N. Ahmed. 2018. Failure is not an option: Policy learning for adaptive recovery in space operations. *IEEE Robotics and Automation Letters* 3, 3 (2018), 1639–1646.
[38] Grégoire Milliez, Raphaël Lallement, Michelangelo Fiore, and Rachid Alami. 2016. Using human knowledge awareness to adapt collaborative plan generation, explanation and monitoring. In *Proceedings of the 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI'16)*. 43–50.
[39] Thibaut Munzer, Marc Toussaint, and Manuel Lopes. 2018. Efficient behavior learning in human—robot collaboration. *Autonomous Robots* 42, 5 (2018), 1103–1115.
[40] Heramb Nemlekar, Jignesh Modi, Satyandra K. Gupta, and Stefanos Nikolaidis. 2021. Two-stage clustering of human preferences for action prediction in assembly tasks. In *Proceedings of the 2021 IEEE International Conference on Robotics and Automation (ICRA)*. 3487–3494.

[41] Stefanos Nikolaidis, David Hsu, and Siddhartha Srinivasa. 2017. Human-robot mutual adaptation in collaborative tasks: Models and experiments. *The International Journal of Robotics Research* 36, 5-7 (2017), 618–634.

[42] Stefanos Nikolaidis, Przemyslaw Lasota, Ramya Ramakrishnan, and Julie Shah. 2015. Improved human–robot team performance through cross-training, an approach inspired by human team training practices. *The International Journal of Robotics Research* 34, 14 (2015), 1711–1730.

[43] Stefanos Nikolaidis, Ramya Ramakrishnan, Keren Gu, and Julie Shah. 2015. Efficient model learning from joint-action demonstrations for human-robot collaborative tasks. In *Proceedings of the 10th ACM/IEEE International Conference on Human-Robot Interaction (HRI'15)*. 189–196.

[44] Stefanos Nikolaidis, Yu Xiang Zhu, David Hsu, and Siddhartha Srinivasa. 2017. Human-robot mutual adaptation in shared autonomy. In *Proceedings of the 12th ACM/IEEE International Conference on Human-Robot Interaction (HRI'17)*. 294–302.

[45] Svetlin Penkov, Alejandro Bordallo, and Subramanian Ramamoorthy. 2016. Inverse eye tracking for intention inference and symbol grounding in human-robot collaboration. In *Proceedings of the Robotics: Science and Systems (RSS), Workshop on Planning for Human-Robot Interaction*.

[46] Vidyasagar Rajendran, Pamela Carreno-Medrano, Wesley Fisher, Alexander Werner, and Dana Kulic. 2020. A framework for human-robot interaction user studies. In *Proceedings of the 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'20)*. 6215–6222.

[47] Ramya Ramakrishnan, Chongjie Zhang, and Julie Shah. 2017. Perturbation training for human-robot teams. *Journal of Artificial Intelligence Research* 59, 1 (2017), 495–541.

[48] Alina Roitberg, Alexander Perzylo, Nikhil Somani, Manuel Giuliani, Markus Rickert, and Alois Knoll. 2014. Human activity recognition in the context of industrial human-robot interaction. In *Proceedings of the Signal and Information Processing Association Annual Summit and Conference (APSIPA), 2014 Asia-Pacific*. 1–10.

[49] Benjamin Rosman, Majd Hawasly, and Subramanian Ramamoorthy. 2016. Bayesian policy reuse. *Machine Learning* 104, 1 (2016), 99–127.

[50] Valerio Sanelli, Michael Cashmore, Daniele Magazzeni, and Luca Iocchi. 2017. Short-term human-robot interaction through conditional planning and execution. In *Proceedings of the International Conference on Automated Planning and Scheduling*.

[51] Mario Selvaggio, Marco Cognetti, Stefanos Nikolaidis, Serena Ivaldi, and Bruno Siciliano. 2021. Autonomy in physical human-robot interaction: A brief survey. *IEEE Robotics and Automation Letters* 6, 4 (2021), 7989–7996.

[52] J. Ridley Stroop. 1935. Studies of interference in serial verbal reactions. *Journal of Experimental Psychology* 18, 6 (1935), 643.

[53] Ron Sun, Xi Zhang, and Robert Mathews. 2006. Modeling meta-cognition in a cognitive architecture. *Cognitive Systems Research* 7, 4 (2006), 327–338.

[54] Adriana Tapus, Maja J. Mataric, and Brian Scassellati. 2007. The grand challenges in socially assistive robotics. *IEEE Robotics and Automation Magazine* 14, 1 (2007), 35–42.

[55] Silvia Tulli, Diego Agustín Ambrossio, Amro Najjar, and Francisco J. Rodríguez Lera. 2019. Great expectations and aborted business initiatives: The paradox of social robot between research and industry. In *Proceedings of the BNAIC/BENELEARN*.

[56] Valeria Villani, Fabio Pini, Francesco Leali, and Cristian Secchi. 2018. Survey on human–robot collaboration in industrial settings: Safety, intuitive interfaces and applications. *Mechatronics* 55 (2018), 248–266.

[57] Matt Webster, David Western, Dejanira Araiza-Illan, Clare Dixon, Kerstin Eder, Michael Fisher, and Anthony G. Pipe. 2020. A corroborative approach to verification and validation of human–robot teams. *The International Journal of Robotics Research* 39, 1 (2020), 73–99.

[58] Nan Ye, Adhiraj Somani, David Hsu, and Wee Sun Lee. 2017. DESPOT: Online POMDP planning with regularization. *Journal of Artificial Intelligence Research* 58, 1 (2017), 231–266.

Received 25 January 2022; revised 2 January 2023; accepted 26 January 2023