Social Cobots: Anticipatory Decision-Making for Collaborative Robots with Extended Human Adaptation

vorgelegt von
M.Sc. Electrical and Electronics Engineering
Orhan Can Görür
ORCID: 0000-0002-5140-5580

an der Fakultät IV – Elektrotechnik und Informatik
der Technischen Universität Berlin
zur Erlangung des akademischen Grades

Doktor der Ingenieurwissenschaften
– Dr.-Ing. –

genehmigte Dissertation

Promotionsausschuss:

Vorsitzender: Prof. Dr. Oliver Brock (TU Berlin)
Gutachter: Prof. Dr.-Ing. Sahin Albayrak (TU Berlin)
Gutachter: Prof. Dr.-Ing. Sami Haddadin (TU München)
Gutachter: Prof. Dr. Guy Hoffman (Cornell University, USA)

Tag der wissenschaftlichen Aussprache: 11.11.2020

Berlin 2020

Abstract

One of the biggest concerns about collaborative robots (cobots) is their ability to adapt during their interactions with humans, which typically exhibit an immense diversity. We present an autonomous framework as a robot's real-time decision-making mechanism to anticipate a great deal of human's characteristics and behaviors, so as to adapt accordingly toward a personalized human-robot collaboration. As a key component of cobots working with humans, existing decision-making approaches try to model the uncertainty in human behaviors as latent variables. However, as more possible contingencies are covered by such intention-aware models, they face slow convergence times and less accurate responses. Hence, they naturally limit the human intention space, leaving many human behaviors rather undiscovered. Our premise in this thesis is to anticipate and adapt to a greater extent of human characteristics and behaviors, including human errors, toward fluent coordination between a human and the robot. We call the collection of such rather overlooked human contingencies "unanticipated" human behaviors. To develop an extended adaptation capability, our novel anticipatory architecture handles such behaviors in two levels: short-term context-dependent human behaviors, e.g., lost attention, are anticipated and adapted in the cognitive level; long-term changing human characteristics, e.g., expertise, are handled in meta-cognitive level decision-making.

In the cognitive level of the architecture, we focus on the robot's short-term adaptation. A novel stochastic decision-making mechanism is implemented using a partially observable Markov decision process (POMDP), anticipating a human's state of mind in two-stages. We name this model design Anticipatory POMDP (A-POMDP). In the first stage, it anticipates the human's task-related availability, intent (motivation), and capability during the collaboration. In the second, it further reasons about these states to anticipate the human's true need for help. Our contribution at this level lies in the ability of our model to handle the following unanticipated human conditions: 1) when the human's intention is estimated to be irrelevant to the assigned task and maybe unknown to the robot, e.g., motivation is lost, the onset of tiredness; 2) when the human's intention is relevant but the human does not want the robot's assistance in the given context, e.g., because of the human's task-relevant distrust for the robot.

At the meta-cognitive level, we handle the robot's long-term adaptation. We present a novel anticipatory policy selection mechanism built on existing intention-aware models, where a robot is required to choose from an existing set of policies based on an estimate of the human. Such policies are generated from different decision-making strategies we develop using our intention-aware POMDP design, i.e., A-POMDP, at the cognitive-level. Our contribution is the Anticipatory Bayesian Policy Selection (ABPS) mechanism which selects from a library of different response policies that are generated from such models and converges to a reliable policy after as few interactions as possible when faced with unknown humans and their dynamic behaviors. The selection is based on the estimation of the human in terms of long-term workplace characteristics that we call *types*, such as level of expertise, stamina, attention, and collaborativeness.

The first round of evaluations is conducted on our simulated factory environment consisting of crafted human models collaborating at a conveyor belt on a pick and place task with our robot. Our contribution here is our novel human models that simulate human behaviors with several contingencies in their characteristics and behaviors that are responsive to the context of the collaborated task, the environment, and the robot behaviors. This allows rigorous tests for our robot model designs when faced with greater uncertainty, including unanticipated human behaviors. Our first results on simulation show that integrating our intention-aware model, i.e., A-POMDP design, into a robot's decision-making process increases the efficiency and naturalness of the collaboration when compared to a model that does not handle such unanticipated human behaviors. Our second results show that incorporating our policy selection mechanism, i.e., ABPS, on top of the A-POMDP models further contributes positively to the efficiency and naturalness when compared to the best intention-aware model in hindsight running alone.

Lastly, we integrate our solutions as an autonomous robotic system with a real-time human-in-the-loop interaction capability. We validate the results obtained from the simulation environment in a real-world setup of a small scale conveyor belt and a robot arm. Our novel setup features such unanticipated human behaviors by presenting a cognitively challenging task of picking, sorting, and placing color-coded products on a conveyor belt within a distractive environment. We conduct user studies, and the objective and subjective analyses validate our earlier simulation findings. The results show that our applied framework with extended human adaptation, covering unanticipated human behaviors and their long-term changing characteristics, increases the efficiency and naturalness of the collaboration with a higher perceived collaboration, positive teammate traits, and human trust when compared to existing anticipatory solutions that do not handle such behaviors.

Zusammenfassung

Eines der größten Probleme über kollaborative Roboter (Cobots) betrifft ihre Anpassungsfähigkeit während ihrer Interaktionen mit Menschen, die typischerweise eine immense Vielfalt aufweisen. Wir stellen einen autonomen Framework als Echtzeit-Entscheidungsmechanismus eines Roboters vor, der einen Großteil der Eigenschaften und Verhaltensweisen des Menschen antizipiert, um sich entsprechend an eine personalisierte Mensch-Roboter-Kollaboration anzupassen. Als eine Schlüsselkomponente von Cobots, die mit Menschen arbeiten, versuchen bestehende Entscheidungsansätze, die Unsicherheit in menschlichen Verhaltensweisen als latente Variablen zu modellieren. Da jedoch mehr mögliche Eventualitäten von solchen Intentions-bewusste Modellen abgedeckt werden, sehen sie sich mit langsamen Konvergenzzeiten und weniger genauen Reaktionen konfrontiert. Daher schränken sie den Raum menschlicher Intentionen natürlich ein und lassen viele menschliche Verhaltensweisen eher unentdeckt. Unsere Prämisse in dieser Dissertation ist die Antizipation und Anpassung an ein größeres Maß an menschlichen Eigenschaften und Verhaltensweisen, einschließlich menschlicher Fehler, in Richtung einer flüssigen Koordination zwischen Mensch und Roboter. Die Sammlung solcher eher übersehener menschlicher Eventualitäten nennen wir "unvorhergesehene" menschliche Verhaltensweisen. Um eine erweiterte Anpassungsfähigkeit zu entwickeln, behandelt unsere neuartige antizipierende Architektur solche Verhaltensweisen auf zwei Ebenen: kurzfristige kontextabhängige menschliche Verhaltensweisen, z.B. verlorene Aufmerksamkeit, werden auf der kognitiven Ebene antizipiert und angepasst; langfristig sich verändernde menschliche Charakteristiken, z.B. das Expertenwissen, werden in der Entscheidungsfindung auf der meta-kognitiven Ebene behandelt.

Auf der kognitiven Ebene der Architektur konzentrieren wir uns auf die kurzfristige Anpassung des Roboters. Ein neuartiger stochastischer Entscheidungsmechanismus wird unter Verwendung eines teilweise beobachtbaren Markov-Entscheidungsprozesses (partially observable Markov decision process, POMDP) implementiert, der den Geisteszustand eines Menschen in zwei Stufen antizipiert. Wir nennen diesen Modellentwurf Antizipierender POMDP (A-POMDP). In der ersten Stufe antizipiert es die aufgabenbezogene Verfügbarkeit, Intention (Motivation) und Fähigkeit des Menschen während der Kol-

laboration. In der zweiten Stufe werden diese Zustände weiter begründet, um das tatsächliche Bedürfnis des Menschen nach Hilfe zu antizipieren. Unser Beitrag auf dieser Ebene liegt in der Fähigkeit unseres Modells, mit unvorhergesehenen menschlichen Zuständen umzugehen: 1) wenn die Intention des Menschen als irrelevant für die zugewiesene Aufgabe eingeschätzt wird und dem Roboter möglicherweise unbekannt ist, z.B. wenn die Motivation verloren geht, die Müdigkeit einsetzt; 2) wenn die Intention des Menschen relevant ist, der Mensch aber die Hilfe des Roboters im gegebenen Kontext nicht möchte, z.B. wegen des aufgabenrelevanten Misstrauens des Menschen gegenüber dem Roboter.

Auf der meta-kognitiven Ebene kümmern wir uns um die langfristige Anpassung des Roboters. Wir stellen einen neuartigen antizipatorischen Policy-Auswahlmechanismus vor, der auf bestehenden Intentions-bewusste Modellen aufbaut, bei denen ein Roboter auf der Grundlage einer Einschätzung des Menschen aus einem bestehenden Satz von Strategien (Policies) auswählen muss. Solche Richtlinien werden aus verschiedenen Entscheidungsstrategien generiert, die wir mit Hilfe unseres Intentions-bewusste POMDP-Designs, d.h. A-POMDP, auf der kognitiven Ebene entwickeln. Unser Beitrag ist der Mechanismus der Anticipatory Bayesian Policy Selection (ABPS), der aus einer Bibliothek verschiedener Reaktionsstrategien, die aus solchen Modellen generiert werden, eine Auswahl trifft und nach so wenigen Interaktionen wie möglich zu einer zuverlässigen Strategie (Policy) konvergiert, wenn er mit unbekannten Menschen und ihren dynamischen Verhaltensweisen konfrontiert wird. Die Auswahl basiert auf der Einschätzung des Menschen in Bezug auf langfristige Arbeitsplatzmerkmale, die wir als *Typen* bezeichnen, wie z.B. Grad der Fachkenntnis, Ausdauer, Aufmerksamkeit und Kooperationsfähigkeit.

Die erste Bewertungsrunde wird in unserer simulierten Fabrikumgebung durchgeführt, die aus handwerklich gefertigten Menschenmodellen besteht, die mit unserem Roboter an einem Förderband bei einer Pick&Place-Aufgabe zusammenarbeiten. Unser Beitrag dazu sind unsere neuartigen menschlichen Modelle, die menschliche Verhaltensweisen mit verschiedenen Kontingenzen in ihren Charaktereigenschaften und Verhaltensweisen simulieren, die auf den Kontext der kollaborierten Aufgabe, die Umgebung und das Verhalten des Roboters reagieren. Dies ermöglicht strenge Tests für unsere Robotermodellentwürfe, wenn sie mit größerer Unsicherheit konfrontiert sind, einschließlich unvorhergesehener menschlicher Verhaltensweisen. Unsere ersten Simulationsergebnisse zeigen, dass die Integration unseres A-POMDP-Designs in den Entscheidungsprozess eines Roboters die Effizienz und Natürlichkeit der Zusammenarbeit im Vergleich zu einem Modell erhöht, das solche unvorhergesehenen menschlichen Verhaltensweisen nicht berücksichtigt. Unsere zweiten Ergebnisse zeigen, dass die Einbeziehung unseres Policy-Selection-Mechanismus, d.h. ABPS, zusätzlich zu den A-POMDP-Modellen einen weit-

eren positiven Beitrag zur Effizienz und Natürlichkeit leistet, wenn man es mit dem besten Intentions-bewusste-Modell im Rückblick allein vergleicht.

Schließlich integrieren wir unsere Lösungen als autonomes Robotersystem mit einer Echtzeit-Mensch-in-the-Loop-Interaktionsfähigkeit. Wir validieren die aus der Simulationsumgebung gewonnenen Ergebnisse in einem realen Experiment-Aufbau bestehend aus einem kleinen Förderband und einem Roboterarm. Unser neuartiger Experiment-Aufbau zeichnet sich durch solch unvorhergesehener menschliche Verhaltensweisen aus, indem er eine kognitiv herausfordernde Aufgabe präsentiert, die darin besteht, farbcodierte Produkte auf einem Förderband in einer ablenkenden Umgebung aufzunehmen, zu sortieren und zu platzieren. Wir führen Benutzerstudien durch, und die qualitativen und quantitativen Analysen validieren unsere früheren Simulationsergebnisse. Die Ergebnisse zeigen, dass unser angewandtes Rahmenwerk mit erweiterter menschlicher Anpassung, das unvorhergesehene menschliche Verhaltensweisen und ihre sich langfristig verändernden Charakteristika abdeckt, die Effizienz und Natürlichkeit der Zusammenarbeit mit einer höheren wahrgenommenen Fähigkeit zur Zusammenarbeit, positiven Teamkollegeneigenschaften und menschlichem Vertrauen im Vergleich zu bestehenden vorausschauenden Lösungen, die mit solchen Verhaltensweisen nicht umgehen, erhöht.

Acknowledgements

First of all, I would like to acknowledge my advisor, Prof. Dr. Dr. h. c. Şahin Albayrak, for providing the opportunity and the infrastructure to pursue my Ph.D. research at DAI-Labor, Technische Universität Berlin. I also gladly express my gratitude to Prof. Dr. Oliver Brock, Prof. Dr.-Ing. Sami Haddadin, and Prof. Dr. Guy Hoffman for their valuable feedback and for taking part in my thesis committee. In particular, Guy Hoffman closely reviewed the study and supported it over years. For that, I express my sincere gratitude. Most importantly, I cannot thank enough my collaborator, Dr. Benjamin Rosman, who has supervised me throughout this research. It has always been very exciting and inspiring to brainstorm on ideas for more than three years almost every week. All of these valuable conversations are now presented as 204 pages of knowledge in this thesis. Without his guidance, this thesis would never reach a successful end. I cannot wait for further opportunities to collaborate with Benji, who also happens to be the coolest supervisor in the robotics world. Furthermore, my special thanks go to Dr. Fikret Sivrikaya whose contribution to this thesis work is also enormous with his great insights. I appreciate all of his support and technical feedback for my research, as well as his valuable advice in general. Also, Dr. Nuri Kayaoglu has always provided great help not only with the difficult Ph.D. life but also for all the aspects of living abroad in Germany. Finally, I will always be thankful to my Master's thesis supervisor, Prof. Dr. Aydan Erkmen, for guiding me into this amazing topic of research.

I am also thankful for all the colleagues that collaborated with me in various research activities and shared their knowledge with me during our time in DAI-Labor and GT-ARC. My dearest colleagues, Doruk Şahinel, Cem Akpolat, and Dr. Elif Eryılmaz-Sigwarth have always been there for me. I cannot forget the friendships we developed over the years and our endless support for each other. They are three of the colleagues that I consider lifelong friends. The other one, Dr. Christopher-Eyk Hrabia, was there with his valuable feedback and endless support. It was a great pleasure to work on several robotics activities with him. Besides, sharing our fears and doubts during our Ph.D. career was very helpful. In general, I find myself very lucky to have the most amazing colleagues. Denis Pozo has became also a valuable friend. Nils Masuch and

Dr. Tobias Küster also reviewed this thesis and provided valuable insights, for which I express my deepest gratitude. And all the colleagues I worked with, namely, Dr. Axel Hessler, Dr. Yuan Xu, Martin Berger, Mathias Wilhelm (also supervised me during my internship in 2009), Michael Burkhardt, Marco Lützenberger, Tuguldur Erdene-Ochir, Cem Bila, Nizar Ben-Sassi, Tobias Dörsch, and Frederic Wilhelm, I thank you all. Not to forget, I am particularly grateful for all the contributions of the students who did their internships and Bachelor's and Master's theses under my supervision. In particular, I would like to thank Elia Kargruber for his enormous effort and great ideas in building our experiment setup for the user studies. It was also a pleasure to mentor Güner Dilşad Er, Minh Nghiem Vi, Elif Damla Gültekin, Oğuz Özdemir, Xin Yu, Jean Piere Demir, Alperen Akgül, Mustafa Kılınç, and Robin Scholtz who also contributed greatly to this study.

Ph.D. is a long journey that also requires strong nerves. My dearest friends, who are also my Berlin family, helped me greatly with that. I would like to expose the names of these amazing people, Arda Yıldırım, Andrea Cerri, Meriç Benderlioğlu, Vincent Sigwarth, Kamila Kutz, Selman Bulut, Ana Fernandez, Carlos Vieira, Arthur Melo, Joao Fernandez, and Thiane Carneiro Melo. I cannot describe how much effort they put to cheer me up and to motivate me toward this degree. They all helped me love this city, which was actually not that difficult, and live in it with the tough Ph.D. path, which was very difficult. I am also greateful to all the friends living around the world but still sending their endless support, Barış Çiftçi, Göksu Çınar, Uğur Doğan Gül (who is also the first supervisor of my professional life), Mustafa Yavaş, Cihan Göksu, and the Ankara crew.

My family deserves a dedicated book with the title "I Can't Thank You Enough!". The book would cover all my life including this dissertation that I see also as their success. Hence, I would like to dedicate this dissertation to them. My parents, Gürin Karacaer-Görür and Rıfat Görür, and my brother and friend, Caner Görür (teali), have always provided unconditional trust, love, and support also throughout my Ph.D. life thousands of kilometers away from them. They made me who I am today and supported me with all of my decisions, unconditionally. I am also thankful to all my other families, namely, the Karacaers, the Tohumoğlus, and the Yüzçeliks, for teaching me everything about life and giving me their endless love. Finally, I would like to thank my lovely wife, Neşe Çakmak-Görür, for being the perfect reason to start my Ph.D. career and for holding my hand from the first year of my Bachelor's degree until today. Her endless love is always my motivation and strength to move forward.

Contents

Acronyms

A-POMDP Anticipatory, Partially Observable Markov Decision Proccess.

ABPS Anticipatory Bayesian Policy Selection.

ANOVA Analysis of variance.

BPR Bayesian Policy Reuse.

CMAB Contextual Multi-Arm Bandit.

cobot collaborative robot.

DOF Degree of Freedom.

FABRIC Framework for Anticipatory Behaviors in Robots toward Interactive human Collaborations.

GMM Gaussian Mixture Model.

GMMHMM Gaussian Mixture Model Hidden Markov Model.

HAR Human Activity Recognition.

HMM Hidden Markov Model.

HRC Human-Robot Collaboration.

HRI Human-Robot Interaction.

MAS Multi-Agent System.

MCMC Markov Chain Monte Carlo.

MDP Markov Decision Process.

MOMDP Mixed Observability Markov Decision Process.

NASA-TLX NASA Task Load Index.

POMDP Partially-Observable Markov Decision Process.

ROS Robot Operating System.

ToM Theory of Mind.

Tukey-HSD Tukey Honest Significant Difference.

WAI Working Alliance Inventory.

Publications

Papers

- 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)*, May 13-17, 2019, Montreal QC, Canada. IFAAMAS, Richland, SC, 77–85.

- 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)*, March 5-8, 2018, Chicago, IL, USA. ACM, New York, NY, USA, 398–406.

- O. Can Görür, Benjamin Rosman, Guy Hoffman, and Sahin Albayrak. 2017. Toward Integrating Theory of Mind into Adaptive Decision-Making of Social Robots to Understand Human Intention. *In Workshop on Intentions in HRI at ACM/IEEE International Conference on Human-Robot Interaction (HRI'17)*, March 6-9, 2017, Vienna, Austria.

- O. Can Görür, Sahin Albayrak. 2016. A Cognitive Architecture Incorporating Theory of Mind in Social Robots towards Their Personal Assistance at Home. *In Workshop on Bio-Inspired Social Robot Learning in Home Scenarios at IROS'16 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'16)*, October 9-14, 2016, Daejeon, Korea.

- O. Can Görür and Aydan M. Erkmen. 2015. *Intention and Body-Mood Engineering via Proactive Robot Moves in HRI. In Handbook of Research on Synthesizing Human Emotion in Intelligent Systems and Robotics, Jordi Vallverdú (Ed.).* IGI Global, 256–284.

1 Introduction

1.1 Motivation and Problem Definition

In the fourth industrial revolution (I4.0), mass customization takes an important role for the future of production. A greater individualization is expected for the consumer products leading to higher variability in produced parts (OECD, 2017). This requires more flexibility and automation in production lines for an efficient operation (Satoglu et al., 2018). For this purpose, I4.0 envisions the spatial and temporal overlapping of human and robot workstations in the production process (Koppenborg et al., 2017). Future factories should efficiently cooperate the productivity, accuracy, strength and repeatability of robots with the flexibility, creativity and cognitive capabilities of humans in mass customization (Pfeiffer, 2016; Rojko, 2017; Villani et al., 2018). A resulting challenge is the need for robots to handle a wider range of production processes and more dynamic human collaborations (Mavridis, 2015a; Rojko, 2017).

Recent advancements in robotics are enabling more human-robot teams to work together for increased productivity. For this purpose, research into human-robot collaboration (HRC) has been mainly inspired by human-human teamwork, the core of which lies in an ability to adapt one's behaviors to the other collaborators from observed behaviors. By doing so, humans select appropriate behavioral responses to maintain reliable and efficient collaboration (Hoffman and Breazeal, 2004). With that in mind, for an efficient HRC, robots should not only consider the environment and monitor the human partner's actions, but also process those actions to anticipate the human's plans and goals in the collaborative task (Hoffman and Breazeal, 2004, 2007; Kuka, 2016; Lasota, Fong, and Shah, 2017). 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, resulting in a more efficient and seamless collaboration (Hiatt, Harrison, and Trafton, 2011; Huang and Mutlu, 2016; Milliez et al., 2016; Nikolaidis et al., 2015). Our motivation is to ensure such an anticipatory autonomous adaptation of robots. We call such robots *social cobots*.

Such an adaptation comes naturally to humans. The effectiveness of our coordination is mostly through our ability to understand the collaborators and alter our plans and actions dynamically, often without verbal communication. Such high-level coordination and adaptation is called fluent collaboration (Hoffman and Breazeal, 2010). The majority of HRC is following command and response patterns, and not giving importance to fluent coordination, which is stated to evoke appreciation and confidence. A robot's acceptance in a work environment collaborating with humans depend heavily on their fluency, that is, well-synchronized coordination of their behaviors with their human partners through mutual adaptation (Hoffman, 2019b). That said, we ask a general question: *What level of coordination, i.e., the extent of human adaptation can/should a robot reach in HRC?*

It is stated that for natural interaction, a robot needs to create adaptive cognitive representations of their cared-for humans (Bodenhausen and Hugenberg, 2011). To ensure the long-term usability of robots and their fluent collaboration, 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 (Tapus, Mataric, and Scassellati, 2007). Moving from this, in this thesis, we categorize a robot's adaptation in HRC as short-term and long-term. Short-term adaptation is defined as the quick adaptation of robots to a dynamic environment without the need for much prior information (e.g., in a single interaction). In contrast, long-term adaptation focuses on behavioral changes on the robot over long periods of time based on the characteristics of the environment; therefore, it requires a memory of previous interactions (e.g., in repeated interactions) (Andriella, Torras, and Alenyà, 2019). That said, we examine the general question above in three main topics, namely, *short-term adaptation*, *long-term adaptation* and *evaluation of adaptation*. The last topic is a step to evaluate

and analyze the extent of human adaptation needed that a robot can reach while still maintaining fluent coordination. These topics are the building blocks of our study, and we organize the entire thesis around them in the core Chapters 4, 5 and 6.

1.1.1 Short-Term Adaptation

For a high-level coordination and fluent HRC, short-term adaptation should cover the following possible collaboration aspects: *i)* when the tasks last a short time and/or come with changing requirements, e.g., dynamic plans for mass customization; *ii)* when a collaborating person exerts dynamic behaviors due to, e.g., their changing mental states like intentions, their quick adaption to changing plans as stated in (Andriella, Torras, and Alenyà, 2019; Hoffman, 2019b; Sanelli et al., 2017). Our belief is that a robot should be able to adapt to its human partners and their dynamic behaviors under such conditions.

There is a large body of research focusing on adopting the ability of humans to interpret and predict others' mental states into robotics in order to anticipate human intentions and plans from observed actions, e.g., (Baker and Tenenbaum, 2014; Holtzen et al., 2016; Scassellati, 2002). A handful of recent studies have extended them to explicitly use these human states in adaptive plan generation and execution, e.g., (Devin and Alami, 2016; Huang and Mutlu, 2016). These studies are built on the following requirements of collaboration: the agents have a common intention, i.e., commitment to a shared goal along with a common belief about the goal state (Cohen and Levesque, 1991), as well as mutual awareness and mutual support, i.e., the willingness to accept support (Hoffman and Breazeal, 2004). In general, the aim of the field has so far been to prove that anticipatory planning increases efficiency in HRC scenarios, but this assumes that these basic requirements hold during the collaboration, and as a result, implicitly makes two common *assumptions*:

i) All of the actions a human executes are relevant to a goal or an intention that is known to the robot (Baker and Tenenbaum, 2014; Devin and Alami, 2016; Görür and Erkmen, 2015; Holtzen et al., 2016; Huang and Mutlu, 2016; Milliez et al., 2016),

ii) Humans always accept the robot's assistance when offered (Devin and Alami, 2016; Huang and Mutlu, 2016; Penkov, Bordallo, and Ramamoorthy, 2016).

In reality, a human's dynamic desires and emotional states could result in stochastic intentions, behaviors and expectations over the course of repeated interactions. This is particularly valid under the aforementioned collaboration aspects. Therefore, a robot making these assumptions might misinterpret human actions, which may result in unreasonable and intrusive robot behaviors, limiting its reliability and applicability in real-life scenarios. As an example, a robot may infer that a human needs an object if it detects the human's gaze on it. However, the human's mental state behind this gaze could be any other task-irrelevant intention (in contrast to *assumption i*). Moreover, even if the human wants the object, they may not want the robot to pick it up, for any of several reasons such as distrust in the robot and the belief that it could damage the object or a desire to remain autonomous (in contrast to *assumption ii*).

We refer to such erroneous human behaviors that may result in a mistake, a task failure or dropped efficiency in a collaboration task as "unanticipated" human behaviors in the context of the given *assumptions* as they are unreasonably uncommon to observe compared to an expected performance from a human. It is stated that although it is crucial, robots have not yet reached a level of design that allows effective management of such human errors or their unexpected behaviors (Honig and Oron-Gilad, 2018). The premise of this thesis regarding a robot's short-term adaptation during a human collaboration is, for the first time to the best of our knowledge, to devise a robot decision-making mechanism that aims to remove these assumptions and model the HRC over the course of repeated interactions to be able to reason about such "unanticipated cases" that could affect the performance of the collaboration. For this purpose, we ask the following research question to be able to extend the short-term adaptation of a cobot:

Research Question 1. *"How can a robot anticipate and adapt to (i.e., model) unanticipated human behaviors, including a human's short-term changing intent (motivation), availability, capability and willingness to collaborate, which could lead to erroneous human behaviors?"*

1.1.2 Long-Term Adaptation

In HRC, long-term adaptation is needed to cover the following possible collaboration aspects: *i)* when collaboration is required over a repetitive task; *ii)* when long-term collaborations with different people are expected (Leite et al., 2014; Leite, Martinho, and Paiva, 2013; Sanelli et al., 2017). Unlike the short-term adaptation that focuses on dynamic conditions like changing human intentions or their changing plans, long-term adaptation handles a robot's ability to distinguish and adapt to various characteristics of collaborating humans that eventually affect their short-term behaviors, e.g., a human with a lower stamina is more likely to need/accept a robot's assistance in a task with physical load after a short time. Such information is not always available to a robot and it may take a long-term interaction to learn such characteristics of individuals. This type of adaptation is especially crucial when a robot is expected to collaborate with multiple people, which is the case in a factory environment with many work shifts.

Towards building robots with broader adaption to humans, many approaches have been proposed, most of which model human intentions and behaviors as a latent variable in robot planning (Bandyopadhyay et al., 2013; Broz, Nourbakhsh, and Simmons, 2013; Chen et al., 2018; Devin and Alami, 2016). An important open problem of such intention-aware models, for their usability in real-life scenarios, is the degree to which they allow for interacting with different human characteristics leading to various intentions (goals) (Albrecht and Stone, 2018). A limitation for such models is that they become computationally expensive and less accurate as a wider variety of human behaviors are modeled (Hiatt et al., 2017; Pöppel and Kopp, 2018). When we consider long-term interactions involving multiple people, their performance is expected to get even worse. In reality, a human's dynamic long-term characteristics, such as level of expertise, stamina (or fatigue), attention, etc. introduce greater diversity in human behaviors throughout repeated interactions (Gombolay et al., 2017). Succeeding to learn and adapt would contribute positively to the fluency of the collaboration also leading to trust in robots and higher acceptance by the collaborating human (Hoffman and Breazeal, 2007).

We believe that it is very difficult to design and/or learn a single model for a person a robot is collaborating with, let alone for different human characteristics. A human's long-term behaviors may change as a reaction to robot responses. For example, a human may become less collaborative when a robot frustrates them by interfering with a task that they would not trust the robot with. A human may also gain expertise in a task. Such conditions might directly affect the human's collaboration preferences, plans and assistance needs from the robot. To ensure the long-term usability of robots, a robot should adapt to both short-term changes in a human collaborator's mental state and long-term characteristics, personal habits, preferences, and trust. We call each different combination of such long-term characteristics a unique *human type*. 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. We ask the following research question to be able to model such a long-term human adaptation in a cobot:

Research Question 2. *"How can a robot anticipate and adapt to a diversity of long-term human characteristics, trust, personal habits and preferences in its decision-making, which may eventually affect a human's short-term behaviors?"*

1.1.3 Evaluation of Adaptation

Human-robot interaction scenarios mostly evaluate their solutions with human involvement and real experience feedback from human subjects. This is crucial in the case of anticipatory robots as the interacted people are the only source of information for the ground truth data on their true mental states. As a result, evaluating the adaptation capability of an anticipatory robot relies heavily on the data collected from the people participating in a user study. As we mention, this introduces a limitation on the diversity of human behaviors encountered during the lab studies, leading to a limited amount of human intentions covered by the existing robotic solutions. However, due to these limitations and the assumptions made, human-aware robots, e.g., assistive robotics and collaborative robots, face a greater diversity of previously unanticipated humans behav-

iors when they are deployed in the wild, limiting their adaptation capabilities (Hoffman, 2019a; Tulli et al., 2019).

In general, benchmarking interactive robots is very difficult due to the lack of availability of the whole dynamics of human behaviors during their evaluations. To decrease the uncertainty, researchers usually need to run long-term and large scale user studies for their human-aware systems to observe a variety of human characteristics and their changing behaviors, e.g., human trust, motivation, preferences, expertise (Leite, Martinho, and Paiva, 2013). Nevertheless, such experiments are very costly and non-practical. Hence, autonomous systems that involve human interactions are not making the same advancements as the other robotics fields that have access to a great range of dynamics in the environments they operate, most of which is generated through simulations. That said, we believe that there is a lack of human simulations that the anticipatory robots can use as a ground truth to train on human diversity and to rigorously test the solutions. Our intuition is that we also need to rigorously run our anticipatory robots on simulation first, which samples a great diversity of human behaviors with short- and long-term dynamics using accurate human models. The tested and improved solutions can then be brought to the real user-studies, providing more effective adaptation to humans. This brings up our third research question:

Research Question 3. *"How can we develop reliable human models that simulate a great diversity of human behaviors including their erroneous behaviors in work environments?"*

For proper 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. It is hard to expect humans in real user studies to convey all the diverse behaviors, for example, due to the novelty effect they face or due to the constrained environments (Leite, Martinho, and Paiva, 2013). To the best of our knowledge, there exists almost no research that considers the whole range of dynamics in human intentions and actions for anticipatory decision making in real-world settings. It is, therefore, an open task for us to transfer the results from our planned HRC simulation targetting a wide range of human behaviors into a real-world experiment and validate it by means

of user studies. 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 (Hoffman, 2019b). Our intuition is that an evaluation setup should emulate real conditions instead of constraining human intentions that eventually leads to the assumptions mentioned in Section 1.1.1. A collaboration setup needs to enable flexible planning, foster also the unanticipated and uncommon human behaviors, such as, lost motivation and attention, tiredness, unwillingness to collaborate and changing preferences, and so allow for a fluent collaboration with flexible planning to compensate such erroneous behaviors. As a result, the robots face a more realistic and challenging environment to anticipate and adapt to, instead of a constrained one with mostly predefined roles. This formulates our last research question:

Research Question 4. *"How can we design a real collaboration setup and a task without making assumptions on human intentions so that they invoke unanticipated human behaviors including human errors in order to properly evaluate our robot's short- and long-term adaptation goals?"*

1.2 Approach

Our goal is to develop solutions that answer our research questions in the context of HRC. In this section, we address our approaches to each of the four problems defined under Section 1.1. Since we also aim for a holistic solution that is applicable as a human-in-the-loop autonomous system, we ensure the integration of each approach as part of our final framework. Our framework (the big picture of the thesis), the integration of our approaches, and how they are handled in other chapters are summarized in Chapter 3). In this section, we set out our approaches that answer our research questions.

1.2.1 Short-Term Adaptation

Our first objective is to address Research Question 1 that results after removing the common *assumptions* mentioned in Section 1.2.1. We propose a Partially-Observable Markov

Decision Process (POMDP) with two stages of the human state of mind anticipation. In the first stage, the planner incorporates the variability of the human's state of mind during the collaboration, which in our case is the human's task-related *availability, intent (motivation)* and *capability*. Then, in the second stage, through these first anticipated states, it tries to estimate if the human needs help and whether the robot should intervene. While doing so, the planner's goal is to increase the efficiency and the reliability of the collaboration, ensuring the safety and the autonomy of the human partner. We call our novel model design Anticipatory Partially-Observable Markov Decision Process (A-POMDP) that effectively handles unanticipated conditions:

(1) When the human's intention is irrelevant to the assigned task and maybe unknown to the robot, e.g. motivation lost, another assignment is received, becoming tired.

(2) When the human's intention is relevant to the task, but the human does not want the robot's assistance, e.g. because of the human's changing emotional states or the human's task-relevant distrust for the robot.

Our contribution lies in the ability of our model to handle these unanticipated conditions. Our goal is to demonstrate that anticipating and taking into account such human variability increases the overall efficiency (increased success rate over a shorter time) and the naturalness (fewer warnings received from the human, hence less intrusive robot behaviors and possibly increased willingness to collaborate) of an HRC. In order to exemplify and validate our approach, we focus on a smart factory environment in this study. We consider an HRC scenario at a conveyor belt for the task of inspecting and storing various 3D-printed products. At this phase of the study, we create diverse human behaviors using different types of simulated human workers and let them interact autonomously with our robot planner. The details of our methodology, implementation, simulation experiments and initial evaluation of our short-term adaptation approach are given under Chapter 4.

1.2.2 Long-Term Adaptation

Our second objective is to address Research Question 2 toward an extended adaptation to long-term human characteristics. As mentioned in Section 1.1.2, we believe that it is very difficult to design and/or learn a single intention-aware model for a person a robot is collaborating with, let alone for different human characteristics, due to the growing complexity of such models. We present a novel anticipatory policy selection mechanism built on top of existing intention- and situation-aware models for an extended adaptation of robots to various human types. Toward our short-term adaptation in Section 1.2.1, we design an A-POMDP that adapts to a human's short-term changing behaviors, modeling their availability, intention (motivation) and capability as a latent variable. Our focus here is on a robot's adaptation to human long-term behaviors, i.e, human types. We create a policy library by randomly constructing different robot models based on our existing A-POMDP model design. Through this random generation, we are agnostic to specific human types and behaviors modeled.

Our contribution is the Anticipatory Bayesian Policy Selection (ABPS) mechanism based on Bayesian Policy Reuse (BPR) (Rosman, Hawasly, and Ramamoorthy, 2016), which selects a policy from the library in a short time and converges to a reliable and nearly optimal policy after as few interactions as possible. The selection is based on a human's estimated long-term workplace characteristics, such as level of expertise, stamina (or fatigue), attention and collaborativeness*, that correlate to the policy performance. Instead of modeling known human types as a latent variable, we estimate human types that are unknown to the robot from the observed human behaviors using Bayesian belief estimation. To our knowledge, this is the first time such a policy selection mechanism has been proposed complementing intention-aware planning approaches in HRC, providing a fast and reliable anticipatory decision-making for both long-term and short-term adaptation to unknown human types (through ABPS) and their changing behaviors through each selected policy.

*We use this term to indicate the level of a human's will to collaborate that may change due to, for example, task-relevant distrust of the human to the robot.

Our goal is to show that integrating such a policy selection mechanism handling long-term human adaptation contributes positively to the efficiency and naturalness of the collaboration, when compared to the best intention-aware model in hindsight running alone. Through this integration of a long-term adaptation mechanism on top of the models covering short-term adaptation, we obtain a novel autonomous framework for a social cobot toward its extended human adaptation. At this phase of the study, we consider the same simulated HRC scenario at a conveyor belt, where different types of simulated humans, responsive to both robot actions and changing environment, collaborate on the task autonomously with our robot implementing ABPS. We present the details of this line of work, including methodology and early evaluations in Chapter 5.

1.2.3 Evaluation of Adaptation

Our third objective is to address Research Question 3, given in Section 1.2.3, for evaluating our target adaptation skills for cobots in simulation. In this thesis, we follow the conventional way of development and validation in robotics solutions, which start with simulations. Toward the purpose, we devise a simulation environment and human decision-making models as Markov Decision Processes (MDPs) that simulate a diversity of realistic human behaviors including the ones leading to human errors in a work environment, i.e., as we call unanticipated behaviors. The simulations enable our decision-making approaches in Section 1.2.1 and Section 1.2.2 to train on large scale data and to be rigorously tested under relatively harsh conditions under a greater uncertainty. To our knowledge, our human model design is unique in the sense that it provides reliable responses yet with a greater diversity through sampling. We use the simulation environment and the human models in the early evaluations of Chapter 4 and Chapter 5. In these chapters, we also briefly discuss their development and how they are used. Later, more details of the simulation environment and the test architecture are given in Section 6.4 of Chapter 6.

For a broader impact, our fourth and the last objective is to integrate short- and long-term adaptation mechanisms into one framework and evaluate all of our approaches

this time in the real world through user studies with real humans. In doing so, we address Research Question 4 and devise a collaboration setup that reflects contingent and diverse human behaviors; in other words, it provides a rather unconstrained and unlimited human intention space. Moreover, we design a cognitively exhaustive task of sorting colored cubes continuously flowing on a conveyor belt and placing them to 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 a turn-taking collaboration in a task, allowing a human and the robot to flexibly replan task allocations and freely take turns based on their estimate of the partner's behaviors. As a result, it allows for evaluating broader adaptation skills leading to the validation of the thesis work. To the best of our knowledge, this is the first time an anticipatory robot decision-making solution is tested on such a large diversity of human behaviors and characteristics.

In addition, throughout the thesis work we focus on nonverbal communication due to the following reasons: *i)* It has been argued that the majority of human-human communication is nonverbal (Calisgan et al., 2012); *ii)* A fluent coordination is often reached through nonverbal communication (Hoffman, 2019b); *iii)* In a factory environment verbal communication is often not efficient due to the noise pollution and that the workers usually need to wear hearing protection (Calisgan et al., 2012). The environment design, system integration, autonomous human-in-the-loop setup, robot cognition, deployed architecture, experiment design and evaluations with user studies are all given in Chapter 6.

1.3 Contributions

Section 1.2 sets out our novel approach in addressing each of four research questions. In this section, we briefly summarize our contributions for the readers.

1. *Short-term adaptation:* In addressing Research Question 1, our contribution lies in the ability of our decision models, A-POMDP, to effectively handle following unanticipated conditions below. This work also appears in (Görür et al., 2018).
    - When the human's intention is irrelevant to the assigned task and may be unknown to the robot, e.g. motivation lost, another assignment is received, becoming tired.
    - When the human does not want the robot's assistance, e.g. because of the human's changing emotional states or the human's task relevant distrust for the robot.
2. *Long-term adaptation:* In addressing Research Question 2, our contribution is an anticipatory policy selection mechanism, ABPS, which selects a policy from the library in a short time and converges to a reliable and nearly optimal policy after as few interactions as possible. The selection is based on a human's estimated long-term workplace characteristics including a human's expertise, stamina, attention and collaborativeness. To our knowledge, this is the first time such a policy selection mechanism has been proposed complementing intention-aware planning approaches in anticipatory robots. This work also appears in (Görür, Rosman, and Albayrak, 2019).
3. *Evaluation of adaptation through an HRC simulation:* In addressing Research Question 3, our contribution is the human decision models, developed as MDP, that simulates a human's decision process covering a large scale of diversity in human workplace behaviors including unanticipated ones as mentioned. To the best of our knowledge, this is the first time an anticipatory robot decision-making solution is rigorously tested on such a dynamic human environment for a greater adaptation.
4. *Evaluation of adaptation through user studies*: In addressing Research Question 4, our contribution is our novel collaboration setup with the task design that does not constrain a human participant's intention space, puts cognitive load and invokes erroneous human behaviors like a distraction. It emulates long-term behaviors of

human in a lab environment in short times, and so allowing for more realistic interaction. To the best of our knowledge, this is the first time an anticipatory robot decision-making mechanism is tested on such a greater diversity of human contingent behaviors and characteristics for a more realistic evaluation of its adaptation skills during user studies.

1.4 Summary of Results

Each of the technical chapters, Chapter 4, Chapter 5 and Chapter 6, has their evaluations, where the first two evaluates the subcomponents (short- and long-term adaptation solutions) in simulation. Chapter 6 deploys our complete framework into the real world conducts experiments with real people, repeats and validates the findings in the previous chapters and validates the thesis work. In general, we define 10 hypotheses for the evaluations and we have successfully proven each of them as stated below:

1. We show in simulation that anticipating and adapting to such human variability in short-term, i.e., unanticipated human behaviors including erroneous ones, increases the overall efficiency (increased success rate over a shorter time) and the naturalness (less intrusive robot behaviors) of an HRC. (covered in *Hypothesis 1*)

2. We show in simulation that a single intention-aware robot model is limited in its adaptation to various and changing human characteristics in long-term. Our ABPS mechanism provides broader adaptation to various and changing human types leading to more efficient and natural collaboration, while maintaining fast and reliable convergence to the best policy. (covered in *Hypothesis 2*)

3. We show that our real robot effectively communicates through its gestures (non-verbally), expressing its intentions (internal states and decisions) so that they can be correctly understood by humans. (covered in *Hypothesis 3*)

4. We show that the collaboration setup we have designed puts a cognitive load on the humans and the degree of cognitive load differs between the task types. As

a result of the cognitive load and/or long working hours, unanticipated human behaviors, such as, getting tired, losing motivation and attention, are successfully invoked during a task. (covered in *Hypothesis 4*)

5. We show through user studies that a robot's fluent collaboration with a human contributes to increased performance in a cognitively challenging task when compared to a human working alone. (covered in *Hypothesis 5*)

6. We show through user studies that our A-POMDP model covering and adapting to a human's unanticipated behaviors (short-term adaptation) contributes to a more efficient collaboration (covered in *Hypothesis 6*); a more natural collaboration (covered in *Hypothesis 7*); has a higher perceived collaboration, trust, and positive teammate traits (covered in *Hypothesis 8*); is able to show a better adaptation (covered in *Hypothesis 9*), when compared to a robot model that does not handle such behaviors.

7. We show through user studies that our integrated system with A-POMDP and ABPS provide a fast and reliable adaptation to both short-term changes in human behaviors, including unanticipated behaviors, and long-term changes in human characteristics, while it is perceived to have high collaboration skills, positive teammate traits and trust. (covered in *Hypothesis 10*)

1.5 Structure of the Document

The remaining of the thesis is divided into six chapters. After introducing our motivation, problems and research questions, we detail our literature survey that paves the road in defining our problems and shaping our work (in Chapter 2). Then, we give our methodologies in facing our research questions that divide the main body of the work into three topics, i.e., short-term adaptation, long-term adaptation and evaluation of adaptation. In the meantime, we aim for an integrated solution holistically bringing these three topics together. We detail the interaction between these parts in Chapter 3

under Section 3.1, which describes the big picture of the thesis. In the same chapter, we then summarize the building blocks, how they are covered in the following chapters and our continuous integration in Section 3.2. In particular, Chapter 3 serves as a summary of our methodology and evaluations, and how they are distributed within the thesis.

We handle our short-term adaptation goals in Chapter 4 that details the motivation and the methodology behind our A-POMDP model design. As a proof of concept, we run evaluations on simulation under the same chapter, showing the effectiveness of such design toward an improved short-term adaptation (in Section 4.3). In Chapter 5, we grasp the long-term adaptation goals toward satisfying our anticipatory robot's adaptation to various human characteristics, i.e., types, in a work environment. The chapter details the methodology behind the ABPS mechanism. Again as a proof of concept, we run evaluations for this component on simulation (in Section 5.3). The last chapter focuses on the evaluation of the adaptation capabilities of our complete framework (Chapter 6). Even though we have experiments under each chapter, this chapter repeats the experiments of Chapters 4 and 5, which are in simulation, and validates the thesis work through user studies. We first give our novel collaboration setup and task designs that together allow for a greater diversity of human behaviors toward testing our adaptation goals (in Sections 6.2.1 and 6.2.2). Then, we detail the deployment of our framework as a human-in-the-loop autonomous robotic system (in Sections 6.2.3 and 6.2.4). Finally, we conduct user studies to validate our complete framework and the simulation results with real humans (in Section 6.3). Since this chapter focuses on the evaluation of adaptation, we also detail our simulation environment and the simulated human models under this chapter to keep the consistency (in Section 6.4). Finally, in Chapter 7 we provide concluding remarks and future works.

2 Literature Survey *

This chapter presents a literature survey on anticipatory social robots and social human-robot collaboration. In the following sections, we first set out some guidelines for an efficient and natural HRC from literature findings (in Section 2.1). Then, we give the existing studies on anticipatory robots and examine their short-term and long-term human adaptation capabilities under the light of our research questions (in Section 2.2). Finally, we investigate the current approaches to evaluating such anticipatory robots, analyze the assumptions they make and the negative effects of these assumptions on the progress of the HRC field (covered in Section 2.3).

2.1 Steps Toward Efficient and Natural HRC

Recent engineering achievements on robotic technologies have led to robots that are robust enough to be put in close interaction with humans. However, the fact that robots have only recently been deployed out of their lab environments leaves it controversial whether or not their capabilities will be satisfactory enough to be accepted in the long-term within proper human environments, e.g., workplaces, homes. Doubts are driven by the fact that the *novelty effect*, the first response to new technology, also exists in HRI and the behaviors and the attitudes of humans towards robots change negatively as it wears off in the long-term (Kidd and Breazeal, 2008). To support the longevity of robot usage in close interaction with humans, robots need to be social actors achieving social intelligence (Dautenhahn, 2007).

*The content of this chapter mostly appears in our previous studies (Görür and Albayrak, 2016; Görür, Rosman, and Albayrak, 2019; Görür et al., 2017, 2018).

Studies in HRC have generated significant results on low-level (functional) planning for robots having safe and productive physical interactions with humans. However, as it has been recently stated in (Lasota, Fong, and Shah, 2017), alongside safe control and motion planning, significant importance should still be given to developing more generalized predictive and anticipatory planning solutions for increased safety and efficiency of a collaboration (Thomaz, Hoffman, and Cakmak, 2016). For this purpose, robots need to create cognitive representations of interacted humans by processing sensory inputs as well as learning from previous interactions (Bodenhausen and Hugenberg, 2011). In the case of social robots assisting humans in different tasks, cognitive representations should comprise understanding the personal needs and preferences of individuals and adaptively responding to meet those needs (Tapus, Mataric, and Scassellati, 2007). Then, such representations should be used to adaptively guide a robot's decisions during human interaction.

As indicated in (Hoffman et al., 2014), these decisions of robots, in general, need to strive for exerting responsiveness behaviors listed as understanding, validating and caring for interacted humans in order to achieve a secure attachment between cared-for individual and the robot. Moreover, it is emphasized that assisting a human should be in a personalized way through the use of memory of learned preferences, guiding the robot planning (Kirsch et al., 2010). Previous work on reasoning over human mental states has mainly focused on visual perspective-taking and belief management in understanding the world from the interacting person's point of view. A Theory of Mind (ToM) approach, the ability to attribute mental states such as intentions, beliefs, and desires to others, have received significant attention (Berlin et al., 2006; Hiatt, Harrison, and Trafton, 2011; Scassellati, 2002; Trafton et al., 2013). Utilization of this information has been shown to improve human-robot teamwork significantly, leading to a more effective and natural collaboration (Hiatt, Harrison, and Trafton, 2011; Trafton et al., 2013). More recent approaches have focused on reverse engineering ToM, where they show that a human's intentions and plans can be inferred by observing the human's actions (Baker and Tenenbaum, 2014; Holtzen et al., 2016). However, these are mostly limited to the

recognition of human states and are yet to extend to adaptively making decisions based on these states.

It has been stated that robots have deficient capacity to reliably adapt to a human's changing affective and motivational states (to empathize) in long-term (Irfan et al., 2019; Leite, Martinho, and Paiva, 2013). This suggests that robots are still far from creating such complex cognitive representations of interacted humans; as a result, they mostly depend on human commands or well-structured scenarios to initiate their interactions. The lack of adaptation to various changing human needs fails to keep users engaged over repeated and long-term interactions (Kidd and Breazeal, 2008). This is also valid in HRC. The majority of HRC is following command and response patterns, and not giving importance to fluent coordination that is stated to evoke appreciation and confidence. It is stated that a robot's acceptance in a work environment collaborating with humans depend heavily on their fluent collaboration capabilities, i.e., well-synchronized coordination of robot behaviors with a human's, often without a verbal interaction, through mutual adaptation (Hoffman and Breazeal, 2010). This is only possible through understanding various human behaviors and their individual needs, preferences, and plans in the context of collaborating on joint tasks and adapting to coordinate reliably (Hiatt, Harrison, and Trafton, 2011). Whereas, such an adaptation of robots to dynamic human needs is an open-ended task that spans longer intervals of time, which is stated to be non-trivial and best achieved with the role of memory in constantly modeling the current state of the world and the cared-for humans (Kurup and Lebiere, 2012; Trafton et al., 2013).

Studies on social robotics highlight the missing integration of cognitive information about the human (intentions, beliefs, needs, etc.) into the highest-level decision-making of robotic architectures toward their natural human adaptations (Devin and Alami, 2016; Görür and Erkmen, 2015). For this purpose, ACT-R/E cognitive architecture applies ToM approach in its reasoning process (Trafton et al., 2013). Although the estimation results are remarkable, the application is limited to a given scenario covering very limited human behaviors, and so with a limited adaptation. This is majorly due to

the problem that the architecture's learning and adaptation skills are mostly limited to its subsymbolic layer making it deficient in adaptively generating and selecting various symbolic decisions (high-level knowledge) in response to changing stimuli. Additionally, the highly complex structure of such memory-centered cognitive architectures practically limits a robot's conformation to social human interactions within real environments (Baxter, 2016). In general, such an open-ended task, in our case the adaptation of assistive robots in their long-term human collaborations, requires both a cognitive process for constructing knowledge and a metacognitive process for monitoring, controlling, and regulating the learned knowledge, i.e., high-level reasoning, (Sun, Zhang, and Mathews, 2006; Thiede, Anderson, and Therriault, 2003).

Moving from the outlined points, the extend of adaptation to human behaviors is still very limited for social robots to be accepted as long-term assistants in human environments. We believe that a metacognitive process integrating human mental state estimation is what is missing in conventional robotic architectures. A metacognitive process needs to utilize human mental states (to empathize) in controlling robot's cognitive tasks while always striving for exerting responsiveness behaviors and assisting humans with their goals and plans (e.g., preferences). Moreover, the architectures should reduce their system complexity and should be compatible and adaptable to various assistant tasks and human types. To our knowledge, there is no such an autonomous robotic framework integrating all of the listed findings above.

In summary, we define a set of high-level guidelines, moving from the findings above, in developing such a robotic architecture for social robots in their long-term, natural, and reliable human collaborations. Our goal is to remark on the challenges the collaborative robots need to face in the future. The *guidelines* are given below.

1. Implement a framework that is flexible and generic enough to easily adjust to different scenarios and people.

2. Ensure the adaptability of robot behaviors to various changing human behaviors and different human characteristics. This requires adapting both to short-term

changes in a human's mental state (e.g., though day at work) and to long-term personal habits and preferences.

3. Recognize human mental states (e.g., ToM) to anticipate the true need and preferences of the cared-for human. Such states include emotional and motivational (intentional) states used to interpret the human's changing beliefs and plans. For more fluent coordination in an industrial environment, non-verbal communication should be supported.

4. Use the estimated human mental states in meta-level robot decision-making, coordinating low-level cognitive processes of the robot. Metacognitive processes assess the overall success of a cognitive system and regulate its process accordingly. In the case of social collaborative robots, the assessment shall be based on changes in the recognized human mental states (see *Guideline 3*). A robot shall understand these changes and make new decisions, such as re-planning, which assures its adaptability (see *Guideline 2*).

5. Personalize the interaction: Learn several decision models including stereotyped plans and the personalized plans for collaborating individuals, construct a memory of such models, and select the most appropriate one for the collaborating human. Everyone has different social preferences and reacts differently to similar responses, which may require different robot strategies to handle.

6. Reward the robot decisions that a human can anticipate and approves toward establishing a secure attachment. Particularly, a robot shall follow responsiveness behaviors that always have internal goals of validating and caring for the user (Hoffman et al., 2014). A secure attachment leads to an increased will to collaborate and trust in the robot. Therefore, a success criterion for the metacognitive evaluation of a cognitive process (see *Guideline 4*) shall be toward getting the approval of the user to robot actions. Here, we generalize a term called *collaborativeness* that encapsulates both a user's will to collaborate with a robot and their trust in the robot.

7. Empathize and respond autonomously: decisional and functional autonomy. Understanding, adapting and planning shall be connected in a closed-loop manner constructing full autonomy in the robot. This supports a robot to engage in an intuitive and long-term interaction (Leite, Martinho, and Paiva, 2013).

8. Make mutual plans with humans toward satisfying efficient and fluent collaboration, in addition to thriving for naturalness. The process of mutual planning with humans (e.g., a task allocation) encapsulates all the other guidelines above.

Following these guidelines, we offer our conceptual robotic framework enabling an extended human anticipation and adaptation of collaborative robots, which is the focus of this thesis (covered in Chapter 3).

2.2 Anticipatory Decision-Making in HRI

In this section, we focus on existing studies on anticipatory decision-making in HRI, which is a field of research that partially or fully targets the *guidelines* we give in Section 2.1. A handful of studies that consider human anticipation in robot decision-making vary widely in their degree of human anticipation and adaptation capabilities in robots. Nearly all in the context of HRC have shown that anticipatory decision-making enables more efficient collaborations. It is stated that to ensure the long-term usability of robots and their fluent collaboration, a robot adaptation needs to cover both the short-term changes for individual differences and the long-term changes for allowing engagement over repeated and long-term interactions. Moving from this, in our literature analysis, we divide our findings and discussions into short- and long-term adaptation of anticipatory robotic solutions.

2.2.1 Short-Term Human Adaptation in Anticipatory HRC

Short-term adaptation is defined as the quick adaptation of robots to a dynamic environment without much prior information needed. Moving from our literature analysis covering HRC domain, this covers the following possible collaboration cases: *i)* tasks

last short and/or come with changing requirements, e.g., changing plans in mass customization; *ii)* collaborating people exert dynamic behaviors due to, e.g., their internal states like intentions and/or their quick adaption to the case in *i)* (Andriella, Torras, and Alenyà, 2019; Hoffman, 2019b; Sanelli et al., 2017). Such dynamic behaviors also include possible human errors that may directly affect the performance of a collaboration (Hiatt et al., 2017). As a result, the anticipation of such cases mentioned above is crucial for a cobot to adaptively plan and lead to more efficient collaboration. We call any dynamic human behaviors that may lead to a mistake, a task failure, or a dropped efficiency in a collaboration task *unanticipated* human behaviors.

Anticipatory decision-making studies, in general, target the gap between the estimation of human mental states and their explicit use in adaptive high-level (shared task-level) plan generation and execution. Several approaches have been proposed in that line, where, for example, a robot estimates its human partner's belief over the state of a shared table cleaning task in (Devin and Alami, 2016) and another adapts to a human's knowledge level in a cooking task in (Milliez et al., 2016), all to adapt to the changing human-robot work division during the tasks. Although these approaches inspire our study, they assume that belief estimation is a fully observable and deterministic process. There have been several studies applying stochastic planning approaches to robot decision-making during social HRI. For example, a robot car uses a POMDP to decipher the intention of the human driver and adapts to it in a driving task in (Broz, Nourbakhsh, and Simmons, 2011, 2013), a robot is guided by a user while it uses a Mixed Observability Markov Decision Process (MOMDP) model to anticipate unstable guiding and take some control in a shared autonomy task in (Nikolaidis et al., 2017), and a robot with an anticipatory motion planner serves a human based on her desires anticipated from her gaze in (Huang and Mutlu, 2016).

Anticipation is thereby mainly used to enable more efficient interactions (Hoffman and Breazeal, 2007; Mainprice, Hayne, and Berenson, 2015; Nikolaidis et al., 2013). It is additionally utilized to minimize the robot's interference with the human workspace (Mainprice, Hayne, and Berenson, 2015). Several studies motivate their efforts in HRI

with increased efficiency by improving the fluency of joint-action (Hoffman and Breazeal, 2007; Nikolaidis et al., 2013). Most of these models introduce human intentions as a latent variable in a decision-making model, such as POMDPs, which causes a great complexity with an increasing number of human intentions anticipated and handled. As stated by Hiatt et al. (2017), for a reasonable convergence time, such a design conventionally has to limit the human intention space and systemic errors a human can make. 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, e.g., in (Baker and Tenenbaum, 2014; Devin and Alami, 2016; Görür and Erkmen, 2014, 2015; Holtzen et al., 2016; Huang and Mutlu, 2016; Milliez et al., 2016). Also, they further assume that humans always accept the robot's assistance when offered, e.g., in (Devin and Alami, 2016; Huang and Mutlu, 2016; Penkov, Bordallo, and Ramamoorthy, 2016). As a result, robot decision policies generated by such complex models have been developed and tested under constrained environments with rather limited interactions (Baker and Tenenbaum, 2014; Bandyopadhyay et al., 2013; Broz, Nourbakhsh, and Simmons, 2013; Chen et al., 2018; Devin and Alami, 2016; Huang and Mutlu, 2016). 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 (Albrecht and Stone, 2018; Leite, Martinho, and Paiva, 2013).

There has been a handful of studies that removes such assumptions on human intentions and behaviors. Contingencies in human actions have been partly considered, e.g., in (Hiatt, Harrison, and Trafton, 2011; Koppula, Jain, and Saxena, 2016); however, all actions are still assumed to be toward fulfilling a task, possibly in a way that differs from the expected plan. For instance, these actions may occur due to a human's subgoals, different plans toward a goal, or unawareness of the changing situations as in (Hiatt, Harrison, and Trafton, 2011). Therefore, all these reviewed studies still assume the human is committed to a given goal and focus the collaboration more on the task execution level rather than the higher-level problems like the *two conditions* mentioned in Section 1.2.1. Leite et al. (2013) states that humans' diversity and various mental states emerge as sig-

nificant factors that impact human actions. Particularly, 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 in (Ji, Lan, and Looney, 2006). 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, inspired by the aforementioned intent detection POMDPs, as a complementary solution to the existing motion planners and as an alternative to the existing high-level decision-makers, that additionally incorporates such unanticipated human behaviors. Our novel A-POMDP mechanism is intended to provide an extended short-term adaptation to changing human behaviors, including the unanticipated ones, and changing plans, which is covered in Chapter 4.

2.2.2 Long-Term Human Adaptation in Anticipatory HRC

Long-term adaptation focuses on behavioral changes on the robot over long periods based on the characteristics of the environment; therefore, it requires a memory of previous interactions (Andriella, Torras, and Alenyà, 2019). Moving from our literature analysis, in HRC, long-term adaptation is needed to cover the following possible collaboration aspects: *i)* when a collaboration is required over a repetitive task; *ii)* when long-term collaborations with different people are expected (Leite et al., 2014; Leite, Martinho, and Paiva, 2013; Sanelli et al., 2017). Hence, as indicated by Leite et al. (2013), such an adaptation should meet an interacted human's long-term changing preferences, habits, needs, and trust, which also most probably differ from one person to another when collaborations with multiple people are required (e.g. in the case of work shifts in a factory environment).

As mentioned, we were unable to find any studies that consider human behaviors as freely stochastic in an unconstrained environment. The assumptions we state in Section 2.2.1 also has a negative impact in the long-term adaptation as it limits various long-term traits like preferences and collaborativeness (e.g. driven by changing trust

of human in a robot). In following our short-term adaptation goals, we conceptualize an anticipatory POMDP for the robot that removes those assumptions and handles a human's unanticipated behaviors after the human's changing availability, motivation, and capability in a collaboration task. Even though we believe that the modeling of such behaviors is essential, it still does not scale up to the long-term traits, for example, our POMDP model handles only the basic type of behaviors (i.e., tired, distracted, incapable) as a latent variable. In other words, it is not able to adapt to different humans but instead acting proactively against one person's short-term changing behaviors. Similar modeling schemes are used in scenarios with pre-assigned roles for the collaborators during the experiments (Cherubini et al., 2016; Nikolaidis et al., 2013); however, a large proportion of tasks in factory environments would benefit from a more generic and flexible planning approaches that do not assume pre-assigned roles (Koppula, Jain, and Saxena, 2016). Therefore, the adaptation of a robot to such changing roles is yet to be expected, but very difficult to scale with a single model that can only handle certain human characteristics.

Towards incorporating more variety in the anticipated human characteristics, some studies have proposed complementary solutions to be built on top of a robot's intention-aware planner to foster high-level strategies. In (Nikolaidis et al., 2015), 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 MOMDP model to decide on the robot actions. The number of types in this study is a limiting factor, where each different type is considered as a partially observable state. This limitation is again majorly due to MOMDPs struggling to scale to more states when each type is introduced as a latent state variable. It has been recently stated that when POMDPs are used to optimize spatio-temporal assignments of robots, accurate system models are needed to evaluate both actions and rewards, which are often unavailable or fail to anticipate and adapt to various conditions in long-term missions (McGuire et al., 2018).

To overcome the limitation of a Markov decision process in its larger-scale adaptation, Bandyopadhyay et al. (2013) build several such robot models with varying reward and

transition functions to handle different tasks. In other words, the robots are given the ability to explore different policies and trade-off toward higher interaction and task quality. However, the study is limited to analyzing different policies to govern such varieties in humans in the context of pedestrian-robot vehicle interaction leaving out the autonomous selection of an optimal one. Our approach to this problem brings together the idea of generating many such reliable Markov models as in (Bandyopadhyay et al., 2013) to construct a policy library, and the idea of estimating human types on a meta-level (e.g., in (Nikolaidis et al., 2015)) as a complementary solution to the intention-aware models, and goes beyond them to offer a fast and reliable policy selection mechanism as part of a closed-loop robot system. Our policy selection replaces the conventional method of using hierarchical or more complicated POMDPs, as it acts as a discretization of the POMDP models instead of modeling types as a latent variable. This allows us to deal with the problem in a more computationally efficient way, and to handle unknown human types while mitigating the need to learn (expensive) response policies on the fly.

For the policy selection in the context of adaptive social agents, some studies have developed decision trees, e.g., in (Kamar, Gal, and Grosz, 2009) or Bayesian models, e.g., in (Pöppel and Kopp, 2018), selecting from a limited number of policies. Towards broader adaptations, Mcguire et al. (2018) propose a Contextual Multi-Arm Bandit (CMAB) approach in an assistant selection mechanism for a robot. Although this approach has proven to be useful in adapting to human capabilities and constraints in a simulation, the exploration factor of a CMAB would be very dangerous and frustrating for a human collaborator in the real world. In addition, in policy or reward learning algorithms, the learning rate is very difficult to tune and the response time is considerably high for any interaction in real-time. Therefore, to satisfy fast learning rates in HRC, which is related to how fast the human behaviors are changing, the studies mostly assume limited human intention space, e.g., in (Nikolaidis et al., 2015; Ramakrishnan, Zhang, and Shah, 2017). Particularly, when human workers have their shift changes or when a human drastically exerts different behaviors (e.g., loss of attention, fatigue, or injuries in workplaces (Gombolay et al., 2017; Ji, Lan, and Looney, 2006)), learning a new reward

function or a policy would take time which is very costly, especially in collaboration scenarios. Moreover, we still need to have an accurate reward and transition model which, in the end, needs to be applicable to all humans being interacted with, and yet again is not realistic to find. In such cases, it is better to reuse an already trained model rather than spending too much time on training a new one (Rosman, Hawasly, and Ramamoorthy, 2016).

In an online HRC, a robot's autonomous, reliable, and fast response is a direct influence on the fluency and the naturality of the human-robot teaming (Hoffman and Breazeal, 2007). Toward more efficient and safe HRC, we believe the BPR algorithm proposed by Rosman, Hawasly, and Ramamoorthy (2016) is the best fit for our problem. BPR has been shown to perform better than a multi-arm bandit (fast and reliable policy selection) in online adaptation tasks when faced with greater uncertainty about the description of the task. It considers *a priori* information leading to less exploration and so less unreliable responses of the robot during operation. In our solution to the long-term adaptation, the ABPS mechanism, we have updated BPR to incorporate anticipation of a human's uncertainty in her long-term behaviors and to be a generic and complementary solution to the existing intention-aware planning solutions for an increased adaptation in real-time, in particular, complementing our A-POMDP model design by selecting a policy from a library of such models, each of which already handles the short-term adaptation. Even though the ABPS is agnostic to any labels of human types and robot policies, for our domain we generalize some characteristic features of humans in workplaces, inspired from the research in (Gombolay et al., 2017; Ji, Lan, and Looney, 2006; McGuire et al., 2018), 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 (covered in Chapter 5).

2.3 Evaluation of Adaptation in Anticipatory Robots

Robotic solutions, in general, are evaluated on simulation environments reflecting the real-world conditions and on the real setups after system deployment. Each field of

robotics has its own evaluation methods, and metrics to measure its performance and validate the systems. In this section, we first investigate the experiment setups in the field of HRC (in Section 2.3.1). Then, we analyze the simulation of HRC in Section 2.3.2. Finally, we examine the most commonly used metrics for the performance analysis of the robotic solutions in the domain (in Section 2.3.3).

2.3.1 Experiment Setups in Evaluating HRC

Working cooperatively in an industrial environment requires effective task allocations, communication strategies, and the design of the environment useful to pursue robust and effective teamwork (Calisgan et al., 2012; Tan et al., 2009). All of these are very challenging to achieve for collaborative robots that need to coordinate autonomously with humans due to the highly stochastic and dynamic nature of humans. The challenges are mainly observed during the experiment setup design for the user studies, in developing a robot cognition mechanism to understand human behaviors and in synchronizing and coordinating a robot's decisions with the human responses. The latter also includes a robot's expressiveness in exerting understandable actions by humans. Therefore, in this section, we examine how the studies face these challenges and point out the existing problems.

2.3.1.1 Setup and Scenario Design

The scenarios involving anticipatory robots closely collaborating with humans are so far either restricted to simulation or to experiment setups that fall short to resemble a realistic industrial environment with tasks that span longer working hours. Due to the implicit assumptions made by many studies, which limit a human's intention space and collaboration preferences, robot decision-making solutions have been developed and tested under constrained environments with rather limited interactions, e.g., in (Baker and Tenenbaum, 2014; Bandyopadhyay et al., 2013; Broz, Nourbakhsh, and Simmons, 2013; Chen et al., 2018; Devin and Alami, 2016; Huang and Mutlu, 2016; Nikolaidis et al., 2015, 2017). We were unable to find any studies that consider the whole range of

dynamics in human intentions and take human behaviors as freely stochastic in an unconstrained environment. Even though it is crucial, that results in robots not being able to reach a level of design that allows effective management of such faulty or unexpected human behaviors (Honig and Oron-Gilad, 2018). Additionally, the current experiment setups for HRC often require a human and a robot executing a task in a turn-taking manner or with a predefined task allocation according to their skills (Chao and Thomaz, 2012; Hoffman, 2019b). This may also limit flexible planning and so the fluency of the coordination.

Moreover, HRI studies test their solutions with human involvement, running evaluations using the feedback from the human subjects. Hence, the validation of the robotic solutions relies heavily on the subjective analysis of the participants in a user study. Given that the robots usually face a certain amount of people in a lab environment, the diversity of the human behaviors a robot needs to adapt to is also limited. In addition, the other factors, such as *novelty effect*, confine the human behavior space even more to a certain set of repeated actions (Leite, Martinho, and Paiva, 2013). This lack of diversity the robots face in the lab environments may lead to unreliable robot responses when faced with unknown human behaviors, decreasing the reliability of such robots in the wild. To overcome this uncertainty, the researchers usually need to run long-term and large scale user studies in order to observe a variety of human characteristics and their changing behaviors, e.g., human trust, motivation, preferences, and expertise, as in (Kidd and Breazeal, 2008; Leite, Martinho, and Paiva, 2013). Nevertheless, such experiments are very costly and non-practical, and so very limited.

It is, therefore, an open task to design experiments to consider a wide range of human behaviors including, in our case, erroneous and previously unanticipated human behaviors during their collaboration with a robot. For the HRC studies, the experiments should resemble harsh working conditions by introducing a high workload that invokes many unanticipated human behaviors alongside their routine task-related behaviors. In this thesis, we close this gap by implementing an experiment setup and designing a collaboration scenario in a lab environment that, to some extend, removes the aforementioned

assumptions and invokes the unanticipated human behaviors, such as lost motivation, lost attention, task-related mistakes, tiredness, etc. For this purpose, we design a task that creates a cognitive load on the human subjects, which leads to such behaviors to be observed in a rather short-term collaboration (covered in Sections 6.2.1, 6.2.2). Furthermore, the collaboration task in our experiments is not enforced as a turn-taking collaboration; on the contrary, a human and a robot take initiative to change and adapt to task allocations on-the-fly by observing each other and the environment towards maximum efficiency and safety. Finally, for a greater diversity of human characteristics and behaviors, we train and rigorously test our solution on a simulation environment, where we simulate humans with many different combinations of robot responses, behaviors, and characteristics (covered in Section 6.4).

2.3.1.2 Robot Cognition for Fluent Collaboration

The communication strategies during a collaborative work require natural awareness of the work environment and implicit and explicit exchange of information between the collaborators (Calisgan et al., 2012; Gleeson et al., 2013). Hoffman (2019) indicates that a robot's acceptance in a work environment collaborating with humans depend heavily on the fluency of their coordination, which is defined as the well-synchronized coordination of the behaviors of human and robot, often communicated through nonverbal signals. In a factory environment, there are mostly strict regulations like human workers needing to put on noise protectors. Such circumstances make verbal communication very difficult, whereas there are also many studies that point out the importance of nonverbal communication in coordinating teammate's actions (Breazeal et al., 2005; Mavridis, 2015b; Tan et al., 2009).

In this regard, it is pointed out that gestures are an effective way of communication in a noisy environment and they are easier to execute without preventing the flow of the coordination (Gleeson et al., 2013). There have been many studies using various machine learning approaches, including K-Nearest Neighbor, Support Vector Machine (SVM), Random Forest, Hidden Markov Model (HMM), to recognize human gestures

(Liu and Wang, 2018; Roitberg et al., 2014; Yamato, Ohya, and Ishii, 1992). Calisgan et al. (2012) analyze the human behaviors in a human-human turn-taking task with the intention of identifying non-verbal turn ending and starting cues. The authors emphasize that a robot should interpret a human's state even without any explicit actions from the human to coordinate accordingly. The results show that humans tend to give nonverbal turn-taking cues, while hand cues make up 96.3% of all of them. They extract the most common cues, they call the natural cues, and a human collaborator expects their partner to understand for a natural collaboration. According to the authors, the most commonly observed natural cues are eye gazing, laying a single hand on the table, laying hands on the table, crossing the arms, tying hands together, exerting several hand gestures, stepping back, and shuffling (Calisgan et al., 2012).The teamwork gestures include not only hand signs but also head and eye contact. Waving hands, thumbs-up, raising a finger, gesturing with the head are some of the most used gestures in teamwork communications (Breazeal et al., 2005). Other studies also point out the importance of head gestures as a good indicator of erroneous human behaviors that could hint for a possible disaster (Baker et al., 2004; Cheng, Park, and Trivedi, 2007).

In this thesis, we are inspired from these findings and construct our robot's observation vector to recognize human actions that drive the estimation of human states. For both the simulation and the real experiments, the observation vector holds the information of when a human idles, acts on a task, makes special gestures for turn-taking, gets distracted, loses attention, and is tired from the detected human hand and head gestures. Our goal is to catch the hints about human plans on a shared task and a human's possible erroneous behaviors. For the recognition of those features, we implement a state-of-the-art hand gesture recognition algorithm with semantic perception using Gaussian Mixture Model Hidden Markov Model (GMMHMM) (inspired mostly from (Ramirez-Amaro, Beetz, and Cheng, 2015; Roitberg et al., 2014)) and a head gesture detection system (covered in Section 6.2.3).

2.3.1.3 Natural Coordination of Robot and Human Behaviors

A high-level of coordination between a human and a robot is required for an efficient HRC. This is only possible if the collaborators could understand each other's changing behaviors and mental states, adapt to this changes, be well-accustomed to each other in terms of knowing each other's needs and preferences, show flexibility in taking different roles if needed and adjust their actions and plans according to changing requirements (Green et al., 2008; Hiatt et al., 2017; Hoffman and Breazeal, 2004; Huang and Mutlu, 2016; Villani et al., 2018). Only then the partners can reach a reliable synchronization of their behaviors. Our goal is to follow all these requirements, to some extent, to be able to reach high-level coordination and to enable an extended human adaptation in the robots.

Such synchronization of actions is a non-trivial problem due to the variety and uncertainty in human behaviors. In our case, since we do not want to limit a human's intention space, do not follow a turn-taking collaboration and leave the task allocation and planning flexible in a rather unconstrained environment, we face with even greater uncertainty in humans and this brings about a number of challenges in physical interaction design. We assume a variety of different human characteristics with different task-related preferences, and thus changing frequencies of robot responses need to be considered. In order to enable natural interactions, human activity recognition, and decision-making models need to cover a wide range of timing expectations for reliable responses that differ from one person to another and even during a single interaction (Mitsunaga et al., 2008; Wilcox, Nikolaidis, and Shah, 2012). In such a dynamic environment, a human activity recognition should update according to a window size that covers a single action period of a person, on average. In deciding the frequency of update patterns and interruptive calls, we are inspired from the study in (Ramirez-Amaro, Beetz, and Cheng, 2015) on synchronized recognition of human activities with semantic perceptions in real-time (covered in Section 6.2.3.1).

The legibility of robot intentions and motions plays a crucial part in such coordination. For that, a lot of research concentrates on the design of the intention of expressive robots.

One of the major focus in this field is the design of robot arm motions that express a robot's intent and are easily understandable by humans (Bodden et al., 2016; Dragan and Srinivasa, 2013; Dragan, Lee, and Srinivasa, 2013; Holladay, Dragan, and Srinivasa, 2014; Lasota and Shah, 2015). Legible motions allow humans to faster predict the robot intentions (Gielniak and Thomaz, 2011), and thereby, enhance the collaboration (Lichtenthäler and Kirsch, 2016). More recent work has found that unpredictable and fast robot movements tend to result in longer task durations in a cooperative quality control task since it increases the perceived risk of the robot, the human collaborator's anxiety as well as her workload (Koppenborg et al., 2017). They argue that a robot operating at a high level of autonomy would be more unpredictable for humans, and as a countermeasure, it is suggested that a robot communicates its intended movements by either visual signs or the sound (Koppenborg et al., 2017).

Bodden et al. (2016) compare different types of robot arm motions with regard to how accurate and fast a human can predict where a robot arm is intending to reach. They calculate trajectories that minimize a specific objective function per motion type; however, it is found that a motion design that fits all kinds of robot arms remains an open task. A recent study by Dragan et al. inspects the effect of different robot motion types on the fluency and execution time in a collaborative beverage preparation task (Dragan et al., 2015). A functional motion is designed to reach a given position and avoid collisions, whereas a predictable motion is described as a motion that the collaborator expects and understands the intention behind. The experiments have shown that the functional motion leads to a longer task execution times and should, therefore, only be used in cases where a task doesn't require coordination between human and robot or no close collaboration is needed. In our study, we design robot motions inspired by these studies. In general, we intend to have a motion that is both legible and predictable as we work on a close proximity HRC with advanced coordination. As such a design is tailored to the setups, we design some alternatives and conduct case studies to be able to find the best robot motions for our setup (covered in Section 6.2.5 and evaluated in Section 6.3.1).

2.3.2 Simulating HRI

The evaluation and validation of robots with human interaction capability rely heavily on the subjective ratings of the invited people participating in a user study. That means, the solutions only encounter a limited amount of people and their behaviors due to mostly short-term interactions and the constrained environments. As a result, benchmarking robotic decision models or several other AI approaches that involve human interactions is very difficult due to the natural diversity of human behaviors that are largely missed during a robot's development and testing phase. There have been many examples of human interactive robots that showed a great success in the lab through user studies, yet failed in the wild when faced with a greater diversity of previously unanticipated, and so unknown human behaviors (Hoffman, 2019a; Tulli et al., 2019). Research fields that involve human interaction, such as, autonomous cars, assistive robotics, collaborative robots, etc., are not making the same advances as the other robotics systems due to the lack of the human models/simulations that the system can use as ground truth to train on and rigorously test with.

Toward covering a bigger range of dynamics in human behaviors and decreasing the uncertainty a robot would face in the wild, the researchers usually need to run long-term and large scale user studies for their human-aware systems (Leite, Martinho, and Paiva, 2013). Nevertheless, such experiments are very costly and non-practical. We believe that, just like the other fields of robotics, robotic solutions for HRI also need to be rigorously tested on simulation first, which samples a great diversity of human behaviors with short- and long-term dynamics using accurate human models. The solutions can then be brought to the real user-studies, providing more effective adaptation to humans. One of our goals in this thesis is to ensure this mass-scale training and testing of an HRC on simulation.

Human behaviors in a work environment have been extensively studied. The researchers agree that the human actions toward their goal may be imperfect (Hiatt et al., 2017). Such imperfections may be a result of many dynamic cognitive, emotional, or physical conditions of humans. There are studies that analyze the behaviors of hu-

man workers operating on repeated tedious tasks. In (Ji, Lan, and Looney, 2006), the authors examine the fatigue levels of human workers and examine the effects leading to and possible impacts of tiredness. Askarpour et al. (2017); (2019) show an effort to model human behaviors during a robot collaboration in a work environment, including erroneous human behaviors and their effects on the safety of the collaboration. The authors describe that such errors are mostly due to inattentiveness and the lack of training. Moreover, there are also second-order causes that are driven by the lack of human skills and vigor. They point out the importance of covering such conditions to avoid hazardous situations. In their studies, they attempt to model such behaviors to simulate possible effects and generate important insights for designing a safe HRC; however, the models are not used to test collaborative robots by sampling non-deterministic human behaviors and generating simulated actions in response to the robot behaviors.

To our knowledge, human simulations with such generative models are not covered and used in training and testing robotic solutions in the HRI literature since covering the highly complex nature of human behaviors in a single modeling scheme is non-trivial. Simulating such a human has been shown to be accurate using a MDP to generate a policy for a human agent (Bandyopadhyay et al., 2013). The authors use these models to simulate pedestrian behaviors in an autonomous car's decision-making process yet with a very limited diversity of human behaviors constrained to certain road crossing scenarios. In this thesis, we target this gap by simulating stochastic, responsive, and reliable human behaviors covering many aspects of a human's possible collaboration performance. In our attempt, we are inspired by the studies on human behaviors in a workplace, e.g., in (Gombolay et al., 2017; Ji, Lan, and Looney, 2006; McGuire et al., 2018), and the literature findings on the erroneous human behaviors, e.g., in (Askarpour et al., 2019). We develop a design that also meets our expectations on responding to a robot's actions and the state of a task while reflecting a human's changing internal states and goals. We assume that a human worker optimizes an objective function to reach her goals. However, following our statement, this may also be an internal goal irrelevant to an assigned task and may reflect a goal imperfectly, e.g. leaving her place for a short break

and losing her attention. We model many different human types for our collaboration scenarios and randomly run generated models to reflect changing and unknown levels of human expertise, stamina, attention, and collaborativeness. The models reflect them as actions in a 3D factory environment, which are observations for the robot obtained from 3D human body joints. Hence, a realistic simulation of a HRC is realized (covered in Section 6.4).

2.3.3 Evaluation Metrics and Methods in HRI

There have been many different metrics used in evaluating social HRI solutions. Most experiments are user studies, where the participants experience different environmental conditions mostly instructed to them beforehand. Most of the reviewed experiment setups collect objective data during the experiments in addition to the subjective data in the form of questionnaires. In the studies that design repeated interactions, e.g., collaboration on repeated tasks, the surveys are often split up into a task related questionnaires and a general questionnaire, with the former being asked after each task to evaluate only that episode of an experiment and the latter concerning the overall experiment asked at the end, e.g., in (Huang and Mutlu, 2016; Koppula, Jain, and Saxena, 2016; Nikolaidis and Shah, 2013). The general questions, or post-hoc surveys, mostly contain open-ended questions to find out a participant's overall impressions of a robot (Huang and Mutlu, 2016; Nikolaidis et al., 2013). They are also asked to verify the correctness of the objective measures or to check the consistencies between objective and subjective results. For instance, the authors in (Nikolaidis et al., 2015) ask post-hoc questions to ensure that a participant's task-related preferences remained constant during a task.

The majority of experiments evaluate the questionnaire statements or answers on the basis of a 5-point Likert scale, e.g., in (Koppula, Jain, and Saxena, 2016; Nikolaidis and Shah, 2013) or 7-point Likert scale, e.g., in (Hoffman and Breazeal, 2007). In the analysis of the results, Analysis of variance (ANOVA) is most often used in order to check if different conditions are providing statistically significant effects, e.g., in (Bodden et al., 2016; Chang et al., 2018; Dragan et al., 2015; Huang and Mutlu, 2016; Takayama, Dooley,

and Ju, 2011). It is often followed by Tukey Honest Significant Difference (Tukey-HSD) post-hoc tests to further narrow down which conditions are distinct from one another and to what degree. In our user studies and evaluations, we are inspired from all these existing studies and follow their ways of designing experiments, protocols, questionnaires, and analyses.

For the metrics used in social HRI evaluations, it has been recently stated that the field is still in need of generally accepted measures, especially for the social collaborative robots (Hoffman, 2019b). After many attempts, in the literature, we have observed many commonly used metrics that highly inspire our study. Starting from the objective measures, the majority of the experiments measure time in the form of a task's execution and completion or a robot's response time to assess the effectiveness, performance, and efficiency of a collaboration (Huang and Mutlu, 2016; Koppula, Jain, and Saxena, 2016; Nguyen et al., 2011). The time information also evaluates a team's fluency by measuring concurrent motion times or human idle times in a collaborative task (Hoffman, 2019b; Nikolaidis et al., 2013; Nikolaidis and Shah, 2013). Furthermore, the studies that involve HRC also measure the success of collaboration in the forms of robot rewards gathered or the final outcome of a task.

In terms of the subjective measures, a number of studies that focus on anticipatory robotic solutions take into account the prediction or detection accuracy of the human intentions by asking the participants about their mental states or internal goals, e.g., in (Huang and Mutlu, 2016). Furthermore, the effectiveness of collaboration is widely asked in the HRC studies. In (Nikolaidis et al., 2013), the participants were asked to evaluate the effectiveness of a robot's performance. In (Huang and Mutlu, 2016), the questions go on to measure the robot's purposefulness and awareness toward a task, as well as the participant's perceived intelligence of the robot. Moreover, the trust, one of the most important metrics in the domain, has been studied to have an impact on the effectiveness of a collaboration; hence, a human subject's trust in a robot is generally evaluated in the surveys (Chen et al., 2018; Hoffman, 2019b; Nikolaidis et al., 2013).

Another subjective measure commonly evaluated is the collaboration behaviors of a robot in response to its human partners. In (Hoffman and Breazeal, 2004), a robot's teammate traits are formulated under three characteristics, namely, its intelligence, task-related trustworthiness, and commitment to a task. Moreover, the Working Alliance Inventory (WAI) (see in (Horvath and Greenberg, 1989)) has been adopted to HRI in measuring the collaborative relationship between a helper and a help recipient in a "bonding" and "following goals" dimensions (Hoffman, 2019b; Kidd and Breazeal, 2008). The aspects like mutual trust and respect and confidence in a teammate's abilities are covered under the "bonding" dimension whereas the "goal" dimension concentrates on the existence of common goals that both team members agree on. Regarding the perceived collaboration, the fluency of collaboration is also discussed as an important measure (Hoffman, 2019b; Hoffman and Breazeal, 2007). Furthermore, a human's will to collaborate and how satisfactory a robot collaboration was are also asked to evaluate a robot's collaboration skills in a team (Hoffman, 2019b; Huang and Mutlu, 2016).

Finally, how natural a collaboration is perceived by the humans is also asked as a subjective measure, as an extension to the general perceived collaboration skills of a robot. In (Koppula, Jain, and Saxena, 2016), a robot's level of naturality is derived from a robot's timely responses from the perspective of the participants. Furthermore, the robot's adaptation to a human's task-related preferences are evaluated, both regarding how well the robot could understand the preferences as well as how well it could subsequently incorporate the preferences into its own action decisions (Koppula, Jain, and Saxena, 2016). In our user studies, we are inspired from all these objective and subjective measures that are often used. In addition to the common objective measures, e.g., task success and robot rewards, we also observe the human responses and a human's contribution to success to evaluate a robot's teamwork capabilities and the naturalness in a collaboration. For the subjective analyses, we mostly inspire from these studies and ask to evaluate the naturalness of collaboration, the reliability of the robot, the participants' trust in the robot, the perceived collaboration and teammate traits of the robot (detailed under the objective measures of Sections 6.3.1, 6.3.2 and 6.3.3).

3 Concept and Architecture

In this chapter, we give the big picture of our entire development phase, clarify the connection between the main building blocks of this thesis, i.e., Chapter 4, Chapter 5 and Chapter 6, and elaborate how together they build our final system architecture while realizing our objectives stated in Section 1.2. The thesis follows a hierarchical development process, where each of the following chapters builds on top of its preceding work through continuous integration. In Section 3.1, we give our overall framework that we call Framework for Anticipatory Behaviors in Robots toward Interactive human Collaborations (FABRIC). This is our autonomous robotic architecture as a final product of this thesis applied to a real setup. Then, we give our workflow to guide the readers through our continuous development process in building this framework (in Section 3.2). We break down the components of the framework and summarize their development, unit tests and evaluations under their corresponding chapters.*

3.1 Anticipatory Robot Decision-Making Framework

Our framework is designed as a human-aware system with an extended human anticipation and adaptation in guiding robot decisions. The architecture, shown in Figure 3.1, is developed following the *guidelines* we set out for developing autonomous robots toward their long-term, natural and reliable human collaboration (see in Section 2.1). Our framework models humans with their characteristics and behaviors and utilizes these models in intention-aware high-level decision-making in the *meta-level*, which then regulates the

*A conceptual version of this framework appears in our previous study in (Görür and Albayrak, 2016)

low-level cognitive processes of a robot. Such an architecture ensures a robot's long-term assistance, in our case during their collaboration with humans, by providing an adaptation toward personalized interactions. We name the architecture FABRIC, an acronym for "Framework for Anticipatory Behaviors in Robots toward Interactive human Collaborations". It is a lightweight design with human-in-the-loop decision-making in real-time. A robot integrating FABRIC selects and executes the best decision model, i.e., a policy, for the interacted human. 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.

FABRIC consists of two main parts: *Meta-cognitive level* and *cognitive level*. The *cognitive level* is reactive and it integrates sensory-motor skills of a robot that is able to function without a meta-level. It comprises of the *sensing, memory, actuating* components. This level deploys sensory skills for a robot in recognizing outside stimuli and actuator skills for executing plans in forms of robot actions. The *meta-cognitive level*, on the other hand, further processes semantically represented stimuli and generates high-level decisions in forms of rules, goals and new plans for the cognitive part. This functionality is a non-blocking process for the *cognitive level* that runs in real-time. Our short- and long-term adaptation goals for the robot is realized in the *meta-level* through the *policy selection* and the *decision-making* components, where the latter acts as a mediator between the two levels as shown in Figure 3.1. The *meta-cognitive level* infers human mental states and the state of the environment, generates decisions adapting to short- and long-term human behaviors and shared goals, evaluates the success of the current plan and, if necessary, interrupts and informs the cognitive process about the new decisions generated.

The functionality of the *cognitive-level* starts with the *sensing* component that consists of *low- and high-level sensing*. *Low-level sensing* maps the environment, detects the objects and the presence of a cared-for human and extracts human signals (e.g., in our case a human's hands and head gestures). Afterward, the *high-level sensing* takes the detected observables as a feature vector, and with the help of recognition algorithms,

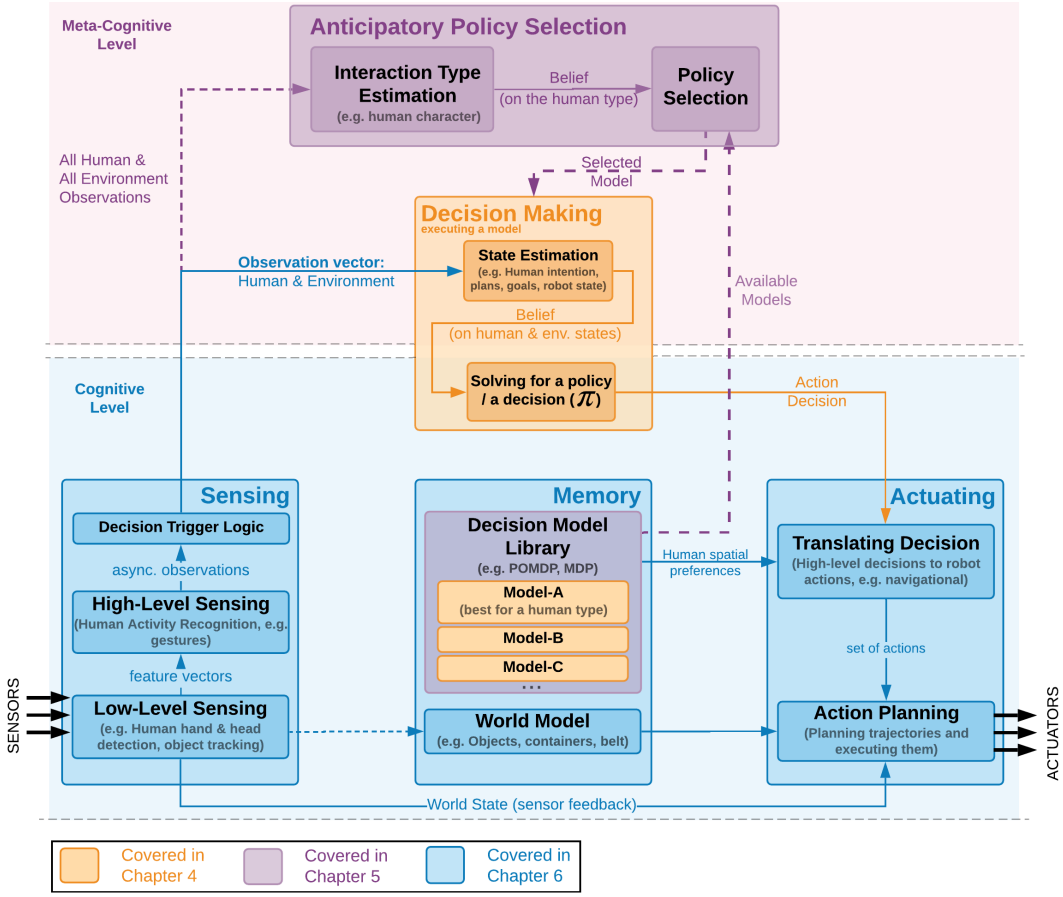


Figure 3.1: Our solution, a Framework for Anticipatory Behaviors in Robots toward Interactive human Collaborations (FABRIC), for robots' improved short- and long-term adaptation to humans toward their personalized and safe collaboration

e.g., a Human Activity Recognition (HAR) algorithm is integrated to recognize human gestures, it semantically describes the interaction environment. For instance, it concludes that a human grasped the red-colored object and put it onto the green container. Finally, the recognized human actions, objects, and the environment are forwarded to the *meta-level* as observations for the decision-making algorithms. After a decision is generated from the *meta-level*, the *actuating* component processes it to generate motion commands for the robot actuators. This component semantically describes the decisions as a related action or a set of actions and goals, for instance, the robot arm should navigate to the green-colored cube and grasp it. Finally, the *action planning* buffers the semantically

described actions coming after each new decision and prioritize 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 robot reaches the goal stated by the current action command.

The *memory* component stores the learned models, i.e., the experience, of different contexts for the robot's future operations. The *world model* component, as shown in Figure 3.1, 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 the information on the location of all the static objects like the containers and the conveyor belt on our collaboration setup (see in Section 6.2.1). Moreover, there are also spatial preferences of humans taken into account by the *actuating* component for personalized physical interaction. Since the proxemics is not the focus of our study, we manually input such preferences as general safety precautions; however, it is possible to have individual requirements (e.g. do not approach the user-X more than 30 cm). Most importantly, the *memory* block stores our novel decision-making models in the *decision model library* as different decision strategies, which are currently designed as POMDPs and MDPs (see in Section 5.2.2.1). The reason for having many decision models follows our discussions that rather than a single adaptive decision model, sometimes a robot may need to follow completely different decision-making strategies, i.e., policies, for a more reliable and faster adaptation. In our case, these models are generated using our A-POMDP model design (in Section 4.2.2). Each of them has a unique set of strategies that are selected to be the best strategy for a unique human type, which is assumed unknown to the robot. In conclusion, the *sensing, actuating and memory* blocks are crucial for the embodiment of our system. These components are built both for our simulation tests and the real setup deployment for the robot's autonomous human-in-the-loop interaction (see in Chapter 6).

Our primary contributions in this thesis lie in the *meta-cognitive level.* As shown in Figure 3.1, collected sets of observations are processed in the *policy selection* component to first estimate the type of the interaction (e.g., the type of a human character

the robot is interacting with, the type of a collaboration task and its requirements, the environment) then to select the most suitable policy for the estimated context. This component, with the *decision model library*, holds the history information about the previous interactions and retrieves the best decision-model to execute during an interaction (e.g., an A-POMDP model). We call this process the *long-term adaptation* as it assumes the context might change in the long-term due to, for example, a human's changing expertise, changing task requirements or a different human worker to collaborate with (covered in Chapter 5). A selected decision model, A-POMDP, is then forwarded to the *decision-making* component to solve and run for an optimal policy. During a collaboration in a single task, this component receives momentary observations from the *sensing* component, estimates the current state of the human and the environment, and selects the optimal decision toward the goals of the running model. Since the *decision-making* component estimates and adapts to rather short-term changes during an interaction (e.g., during a collaboration task a human may get tired or lose her motivation, both of which require the robot's assistance), we call this process the *short-term adaptation* (covered in Chapter 4). In summary, our long- and short-term adaptation solutions, when integrated through our framework, enable fast and reliable online adaptation to various human characteristics and their various behaviors including their risky and unanticipated behaviors leading to human errors.

3.2 Methodology and Workflow

This section summarizes our workflow in designing and developing FABRIC through a continuous integration process. We also set out how we distribute the development into the following chapters. In this thesis, we follow the conventional way of implementation and validation in robotic solutions. This consists of unit tests through rigorous simulation tests, followed by continuous integration and system deployment on a real setup. Then, we continue our improvements through the case studies and finalize our work after the evaluations with a user study. The remaining chapters, Chapter 4, 5 and 6, follow a conceptual order of this continuous development and integration process.

Short-Term Adaptation Below we summarize the steps we follow during the implementation phase of the work in Chapter 4 and how it contributes to FABRIC given in Figure 3.1.

1. Designing a base decision model, our A-POMDP, that implements the short-term adaptation capability. In this phase, we follow our discussions on handling the unanticipated human behaviors like lost motivation, lost attention, getting tired, being incapable, unwillingness to collaborate (see the *two conditions* in Section 1.1). We call a robot running this decision model a *proactive robot*.

2. Implementing a robot decision model that does not handle such unanticipated human behaviors for comparison purposes. We call a robot running such a decision model a *reactive robot*.

3. Implementing the *decision-making* component given in Figure 3.1 for the real-time capability of solving and executing a policy from a A-POMDP.

4. Implementing a human decision model that simulates human behaviors in a work environment with various characteristics, intentions, and goals using an MDP. This is for simulation tests of our robot models as a unit test. The robot models are agnostic to the human models, that is, no data generated from the human models are fed into the training of the robot models.

5. Designing and developing a simulation environment. We develop a 3D simulation of a human and a robot collaborating on an assembly line. The decisions generated from human and robot decision models are actuated on the 3D models in the simulation environment. The collaboration happens on a pick-and-place task on the continuously running conveyor belt carrying products.

6. Implementing the first version of the *sensing* component in Figure 3.1 with a simple HAR system that recognizes simulated 3D human actions.

7. Developing the *actuating* component that translates the generated robot decisions into the simulated robot actions, interfacing our simulation environment. We then

integrate this component with the *decision-making* component for a human-in-the-loop autonomous system. This completes the first version of FABRIC for simulation except the *policy selection* and the *memory* components.

8. **Validating A-POMDP**: Proactive vs. Reactive robot (see in Section 4.3).

   - We compare the collaboration performance of the proactive and the reactive robots in the simulation environment interacting with a vast range of simulated human behaviors. This enables rigorous testing of the robot decision models and their adaptation capabilities. Our evaluation metrics are the efficiency and the naturalness of the collaboration.

   - At this phase of the study for the unit tests, the robot models are hand-coded addressing the points discussed. In order not to bias the robot models, they are trained agnostic to the state transitions inherent in the human models.

   - We are able to dynamically change human behaviors, e.g., stubborn, distracted, tired, by sampling random but goal-oriented dynamics for the human collaborator using a Monte Carlo sampling.

   - Such changes are then executed on the simulated human during the scenario randomly. By doing so, in each scenario, the robots interact with a human with changing levels of stubbornness, tiredness, and distraction. This induces more occurrences of the aforementioned unanticipated human behaviors over the course of the collaboration as the number of task repetitions increase.

   - We show the importance and applicability of our extended short-term adaptation implemented through our A-POMDP models. We demonstrate, through simulations, that handling such unanticipated human behaviors that may occur in the short-term increases the efficiency and the naturalness of the collaboration when compared to a decision model that does not take such behaviors into account.

Long-Term Adaptation After the validation of A-POMDP design, we focus on the long-term adaptation capability of our robot. This is covered in Chapter 5 that builds on top of Chapter 4 findings. Below are the steps summarizing the content of this chapter and how it contributes to our overall framework.

1. Constructing the *decision model library* component. We follow our discussions in Section 1.1 on the need for different strategies when collaborating with different humans or their long-term changing characteristics, such as expertise, stamina, pensiveness, and collaborativeness. Different intrinsic parameters in an A-POMDP yield a unique robot behavior model, which is defined as the best robot for a unique human type. Thus, we randomly create many different models as such to construct the library.

2. Developing our Anticipatory Bayesian Policy Selection (ABPS) mechanism that consists of two parts: 1) Bayesian belief estimation that estimates the characteristics, i.e., a type, of an unknown human as a distribution of the known human types, 2) Bayesian policy selection mechanism that runs on the *decision model library* and selects the best policy against the estimated human type. The policy selection is trained and runs on top of the A-POMDP models in the library.

3. Implementing the *policy selection* component in FABRIC that integrates our ABPS approach and allows it to run in real-time. Thus, we obtain the first complete version of FABRIC running in simulation.

4. **Validating ABPS**: ABPS running on many A-POMDP models vs. the best performer A-POMDP model in hindsight running alone in simulation (see in Section 5.3).

    - We first test the performance of the ABPS, then validate the necessity of such a selection algorithm for the long-term adaptation that leads to a more efficient and a natural collaboration. For this purpose, we compare our architecture in Figure 3.1 with and without the *policy selection* component in the simulation.

That is, we compare the performance of ABPS robot with the *decision model library*'s best policy in hindsight running alone.

- At this phase of the study, we train the estimation and the selection mechanisms of ABPS over randomly generated, a wide range of human decision models. For this purpose, we create many different simulated human types by randomly changing the intrinsic parameters of the human decision model we create in the previous chapter. We also use the same simulation environment and collaboration scenario.
- Many scenarios are executed using randomly created human types unknown to the robots allowing rigorous tests on the system. Our results from simulations show that the ABPS estimates and picks a decision model in a short time and leads to a more efficient and a natural collaboration in the long-term when compared to the best intention-aware decision model in hindsight running alone.
- Hence, we show the necessity of such a policy selection mechanism in long-term collaborations and the applicability of our integrated framework, FABRIC, in simulation.

Evaluation of Adaptation, System Deployment and User Studies After the core components are rigorously tested in simulation, which is a rather more dynamic environment with vast amount of different human types and their dynamically changing behaviors that are hard to observe in a user study, our goal now is to show the applicability of our system in the real world and validate our ideas this time with real human interactions. Chapter 6 improves our framework to deploy and test on a real robot and prepares a novel setup and experiment design toward a more realistic collaboration environment. Below are the steps summarizing this chapter's content:

1. Designing a collaboration scenario and a task that is similar to the simulated scenario.

2. Designing an experiment setup that together with the task invokes unanticipated human behaviors, such as, lost attention, lost motivation, etc. This is expected to lead to the human errors that we want our robot to handle.

3. Improving the *sensing* component in Figure 3.1 with the real-time HAR capability this time for the real setup. This step includes human gesture and object recognition algorithms and the observation vector generation from the real setup for the decision-making.

4. Developing the *actuating* component that translates robot decisions into actions on our robot hardware, and designing expressive robot actions that are able to reflect a robot's decisions to the interacted humans for a fluent communication.

5. Developing a *world model* for the interaction setup and the collaboration scenario.

6. Adjusting the A-POMDP model design and the ABPS mechanism to be better suited for a real-time interaction with a real human. This step trains the models using the simulation data that covers many different scenarios at first, then improves the trained models with the real human data collected during the case studies.

7. Deploying the complete FABRIC architecture on our robot.

8. **Validating the core components developed in the previous chapters and the full system through user studies** (see in Section 6.3).

    - *Experiment-0*: It includes an initial case study to test the system setup, experiment design, task, and collaboration design and robot expressiveness. Using the outputs of this experiment, we validate our novel setup that it puts a cognitive load on humans and invokes human errors as expected. Additionally, using the data collected we adjust the setup and our A-POMDP base model and our ABPS system to better synchronize with the humans in real-time in a *non-turn-taking* collaboration (see in Section 6.3.1).

- *Experiment-1*: The experiment is to validate our A-POMDP model through a user study. It repeats the experiments in Chapter 4; therefore, the *policy selection* component is excluded in this experiment. We show the similarity between the results obtained from the simulation and prove our extended short-term adaptation capability and its necessity, this time with real humans (see in Section 6.3.2).
- *Experiment-2*: The experiment is to validate the applicability and the long-term adaptation capability of FABRIC. We show that our framework is a human-in-the-loop autonomous system that is capable of adapting to and interacting with humans in real-time without any turn-taking needed, i.e., a more fluent collaboration and a high-level coordination. Also, we validate that the ABPS provides adaptation to the long-term changing characteristics of humans, such as their expertise.
- Our final analysis shows that our framework contributes positively to the efficiency and the naturalness of a collaboration. Also, it yields high perceived collaboration skills and teammate traits and trust.

4 Short-Term Adaptation: A-POMDP *

In this chapter, we focus on the short-term adaptation of a cobot to unanticipated behaviors of humans and their changing preferences during a collaboration task (the decision-making component of our framework, FABRIC, in Figure 3.1). We propose a novel stochastic robot decision-making model to anticipate a human's state of mind, and so plan accordingly during a human-robot collaboration task. Our model is a partially observable Markov decision process (POMDP), we call it Anticipatory POMDP (A-POMDP), anticipating a human's state of mind in two-stages. In the first stage it anticipates the human's task related availability, intent (motivation), and capability during the collaboration. In the second, it further reasons about these states to anticipate the human's true need for help. 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 robot, e.g., motivation is lost, another assignment is received, onset of tiredness, and 2) when the human's intention is relevant but the human doesn't want the robot'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 robot. Our results show that integrating this model into a robot's decision-making process increases the efficiency and naturalness of the collaboration.

*The content of this chapter was originally published as 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), March 5-8, 2018, Chicago, IL, USA. Association for Computing Machinery, New York, NY, USA, 398–406. DOI:https://doi.org/10.1145/3171221.3171256 (Görür et al., 2018).

4.1 Introduction

In this chapter, we implement a POMDP with two stages of human state of mind anticipation, covering our short-term adaptation goals stated in Section 1.2.1. We call our novel models anticipatory partially observable Markov decision process (A-POMDP). The ultimate goal of the robot is to estimate a human's unanticipated dynamic behaviors that could lead to human errors, and correctly interpret if the human needs help and whether the robot should intervene. While doing so, the planner's goal is to increase the efficiency and the reliability of the collaboration, ensuring the safety and the autonomy of the human partner.

Our contribution is the previously unanticipated human conditions that we model in two stages. In the first stage, our model anticipates if human has lost motivation, lost attention, is incapable to fulfill a task or tired. Moving from these findings, in the second stage our model further reasons about to anticipate if the human needs help and would accept the robot's asssistance offer. Our goal is to demonstrate that anticipating and taking into account such human variability increases the overall efficiency (increased success rate over a shorter time) and the naturalness (less warnings received from the human, hence less intrusive robot behaviors) of an HRC.

In order to exemplify and validate our approach, we focus on a simulated smart factory environment. We consider an HRC scenario at a conveyor belt for the task of inspecting and storing various 3D-printed products, as illustrated in Figure 6.26). The details of the simulation environment is given under Section 6.4. At this phase, we train and evaluate our approach in simulation, then we repeat the same experiments this time on a real setup with real humans in Chapter 6. We model different types of human workers and let the models interact autonomously with two different robot planners, one of which incorporates the *conditions (1)* and *(2)* (see in Section 1.2.1) through this two-stage anticipation (Section 4.2). Finally, we present our results and analysis of the effects of our robot model on the overall productivity and naturalness of collaboration (Section 4.3).

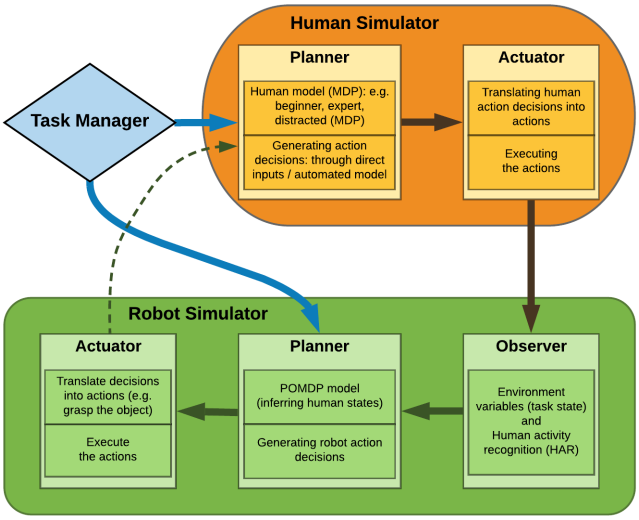


Figure 4.1: System flow of the simulated human and robot interaction

4.2 Methodology

4.2.1 Overall System Flow

In our general HRC framework, we let both the robot and the human intervene and assist each other when needed, where the robot's ultimate goal is to estimate correctly and non-intrusively when to intervene and assist the human. To facilitate this, we present a system flow showing how the simulated humans and the robot interact with each other. The architecture consists of three main building blocks, as depicted in Figure 4.1: Task Manager, Human Simulator, and Robot Simulator. Each HRC scenario starts by receiving a task of inspecting and storing 3D-printed products (see Figure 6.26) assigned initially either to the human or to the robot. The human takes an action decision, as elaborated in Section 4.3.1, which is then actuated within the simulation environment in our setup (more details on the simulation environment are given in Section 6.4). Although robot perception is not the focus of our contribution, we equip the robot with a human sensing capability, through recognizing distinct human gestures, to achieve a more realistic implementation (Section 4.3.2). These observations are then fed to our robot planner model to estimate the human's belief and take action decisions accordingly, as explained next.

4.2.2 Robot Models

For this purpose of clearly showing the advantages of using our anticipatory planning approach, we design two robot models where one is intended to handle the *two conditions* stated in Section 1.2.1 while the other cannot. We name these the *proactive* and *reactive* models, respectively. Our goal is to examine the additional effects of the stochastic interpretation of a human's need for assistance (*condition (ii)*) and the anticipation of a human's changing *availability, motivation* and *capability* (*condition (i)*) on the overall efficiency and naturalness of the HRC.

The reactive robot model is introduced as a base model for the comparative evaluation of our proactive robot. It is an MDP, as shown in Figure 4.2, with a tuple $\{S, A, T, R, \gamma\}$ S comprises the state of the human collaborator's need for help from the robot's perspective, the global success and failure states that define the result of the task (terminal states), the states of a new task assigned to the human or to the robot (initial states), and a state when the robot receives a warning from the human for any reason; A is the robot actions listed in Figure 4.2; T is the state transition probabilities; R is the immediate reward the robot receives; γ is the discount factor for delayed rewards. Positive rewards are acquired when the global success state is reached (the task has been accomplished by some agent). Negative rewards are assigned when the global fail state is reached, or when warnings are received from the human to encourage the planning to be less intrusive, i.e., the robot will not offer help unless it is deemed part of the optimal policy. We solve the MDP model for a robot policy, π, that is optimal with respect to the robot's expected total reward.

In the reactive model the states are directly observable to the robot through the list of observations listed in Figure 4.2. How these observations are obtained is given in Section 4.3. Toward our goal of examining the effect of handling *condition (ii)*, the state of *human needs help* is fully observable in the reactive model, reflecting the assumption mentioned in Section 1, namely that humans accept the robot's help right away. In other words, the MDP treats the human's need for help as a directly observable (deterministic) state. In general, the reactive behavior of the robot is expressed through the robot

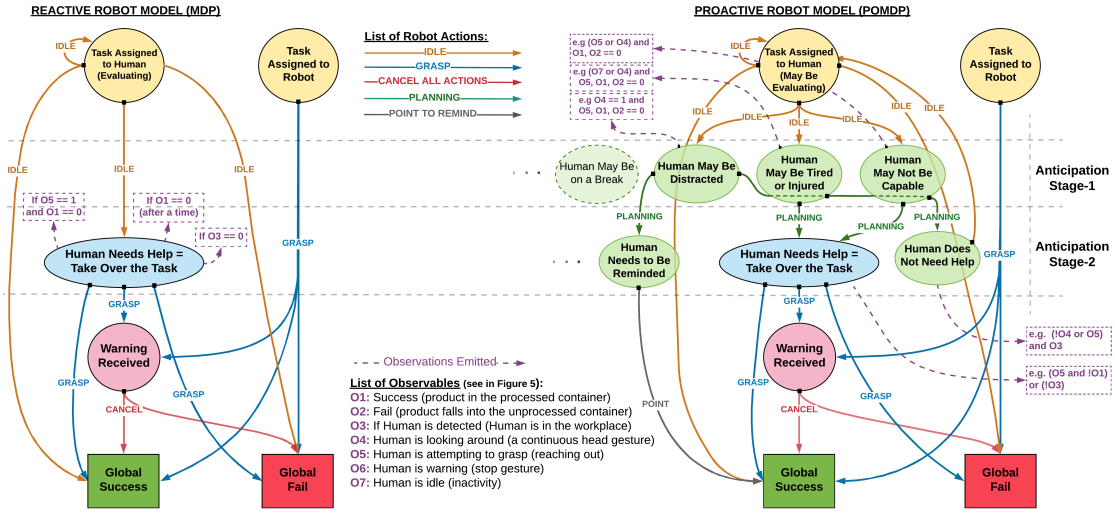


Figure 4.2: Reactive and proactive state-action connection models. The observables that are input to both of the systems are listed. The states added to the proactive model which incorporate the further reasoning (anticipation) layer of the robot are shown in light green.

directly taking over the task no matter what the human's actual internal state is (i.e., ignoring the *two conditions*). The robot assumes the human needs help deterministically when (i) a certain time duration has passed with no success achieved (i.e., leaving enough time for the robot to realize the task before its time limit is reached in a continuous process), (ii) the human is not detected around the work place, or (iii) the human takes a task-related action but no success is detected (e.g., the human failed to grasp and lift in our case).

The proactive robot model is a POMDP inspired by available human-intent based POMDP planners, e.g. (Broz, Nourbakhsh, and Simmons, 2013). The model is a tuple $\{S, A, T, R, \Omega, O, \gamma\}$. The five elements, S, A, T, R, γ, have the same interpretations as in the reactive model, while Ω is the set of observations as listed in Figure 4.2 and O represents the conditional observation probabilities. We also solve the POMDP model for an optimal robot policy, π. Both the proactive and reactive models share the same immediate reward assignments and receive the same observations from the world. As mentioned, we keep their differences to the level of handling the *two unanticipated con-*

ditions. These differences are twofold: (1) although they share the same rules for the detection of a human's need for help (as visualized in Figure 4.2), the proactive model is not deterministically bound to these rules through its partial observability (as opposed to the *assumption (ii)* in Section 1.1.1); (2) the proactive model has more belief states than the reactive one, as visualized in Figure 4.2, to be able to anticipate human's changing availability, motivation and capability contrary to the *assumption i).* For this purpose, in our implementation we model the human states of being distracted, tired and not capable of fulfilling a task. In the reactive model, such intermediate states of the human that give unanticipated behaviors are encapsulated under *human needs help* whereas those are all handled separately in the proactive case.

In the proactive model, we distribute these states to the two stages of anticipation. The *anticipation stage-1*, as shown in Figure 4.2, is an additional stage in the proactive model consisting of hidden human states of *availability, motivation* and *capability.* We believe that it is necessary for the robot to be able to account for the so-called unanticipated human behaviors given in *condition (1)* in Section 1.2.1. Having such a hierarchical anticipation phase also helps the robot to reason about if the human does not need help, see in the *anticipation stage-2* in Figure 4.2. We put this condition as another state which adds uncertainty to the robot's previous estimation in the *anticipation stage-1.* This gives the robot the ability to reason that it was actually wrong when the observation gathered from the human changes accordingly. This prevents the direct conclusion of taking over the human task when it should not, or when changes occur during the interaction. As a result, it is expected to decrease the number of warnings the robot receives from the human. For clarity and brevity of presentation, only the most prominent state transitions of the proactive model are visualized in Figure 4.2. For example, the transitions from the *anticipation stage-1* to the *Global Success* or *Global Fail* with the *Planning* action is not shown (i.e., the human has already ended the task or it failed). Similarly, we provide only some examples of the observations emitted from the proactive model states in the figure. From these examples, we point out that it is not trivial to estimate the states in the *anticipation stage-1* due to their similar observations emitted.

Such a hierarchical approach also provides some additional capabilities to the robot's reasoning that are expected to contribute to the fluency of the collaboration. For example, as depicted in Figure 4.2, the robot has some additional action decisions such as *planning* and *pointing to remind* as a result of the additional states in the proactive model. In the *planning* action, which is a necessary step for any other action to be taken, the robot starts to plan its motion (e.g., for grasping: find the grasping points and plan for moving the grippers) right after any state in the *anticipation stage-1* is estimated. In the cases where the human really needs help, this behavior is expected to save the robot a significant amount of time to execute the action, in our case grasping. In addition, after estimating that the human may be distracted, e.g., detected *looking around* action for a time, the robot may take the decision of pointing toward the object to draw the human's attention to it rather than directly taking it. The human states we include in our proactive model are intended to reflect on the general possible human states in a work environment. As shown in Figure 4.2, such human states and so robot capabilities can be extended by adding more states to the *anticipation stage-1* and some relevant robot decisions to the *stage-2*.

Some general remarks that apply to the robot models are provided below.

i) The models are designed generically to comply with various HRC scenarios. For better clarification we use *grasp* as the action to achieve our specific task but in general any action that a task requires can replace it, making the model adaptable to different use-cases.

ii) Both models start with the task assignment states. If the task is assigned to the human, the robot first remains idle and observes the human (see idle action in Figure 4.2).

iii) Since our focus remains on the human states of mind that have a direct impact on the progress of the task, we do not consider the internal states of the robot (e.g., its battery level). We assume that the robot action of grasping always leads to global success, global failure or warnings from the human.

iv) The exact reason for a warning is hidden to the robot but it may be due to the human's task related distrust to the robot, the desire to remain autonomous, an incorrect estimation of the robot about the human's need for help, the robot realizing the task incorrectly, etc.

v) After a warning is received, the robot cancels its action for the safety and autonomy of the human.

4.3 Evaluation on Simulation

4.3.1 Human Models

As a proof of concept, we devise experiments where we control the degree to which the human displays these unanticipated random behaviors. Evaluating these behaviors with the corresponding robot responses would be difficult with real human subjects in experiments (in an uncontrolled environment) as they depend on the personality of the human and are likely to be observed over the course of longer interactions in the workplace (Ji, Lan, and Looney, 2006). In simulation, through modeled humans, we scale the experiments to emulate many different combinations of such behaviors to rigorously test our robot model's capability to estimate, avoid and respond to such cases (e.g., a human stubbornly rejecting the robot's help, getting tired fast, being distracted easily). We are agnostic to the exact implementation of the human models (i.e., the states, actions, transitions) while our goal remains creating use-cases where a human worker follows the aforementioned *two conditions* and occasionally performs unanticipated behaviors.

For our purposes, we implement an example human model using an MDP as shown in Figure 4.3. More details of this modeling scheme, along with its design, are given in Section 6.4.1.1. During the operation a human model first selects an action. Then, a state transition occurs randomly (if the task does not succeed or fail), reflecting normal and unanticipated human states of mind. In a normal behavior, the human first evaluates the task, grasps the product, and inspects and places it into the adjacent container (see Figure 4.4a). The rest of the state transitions reflect the unanticipated behaviors through

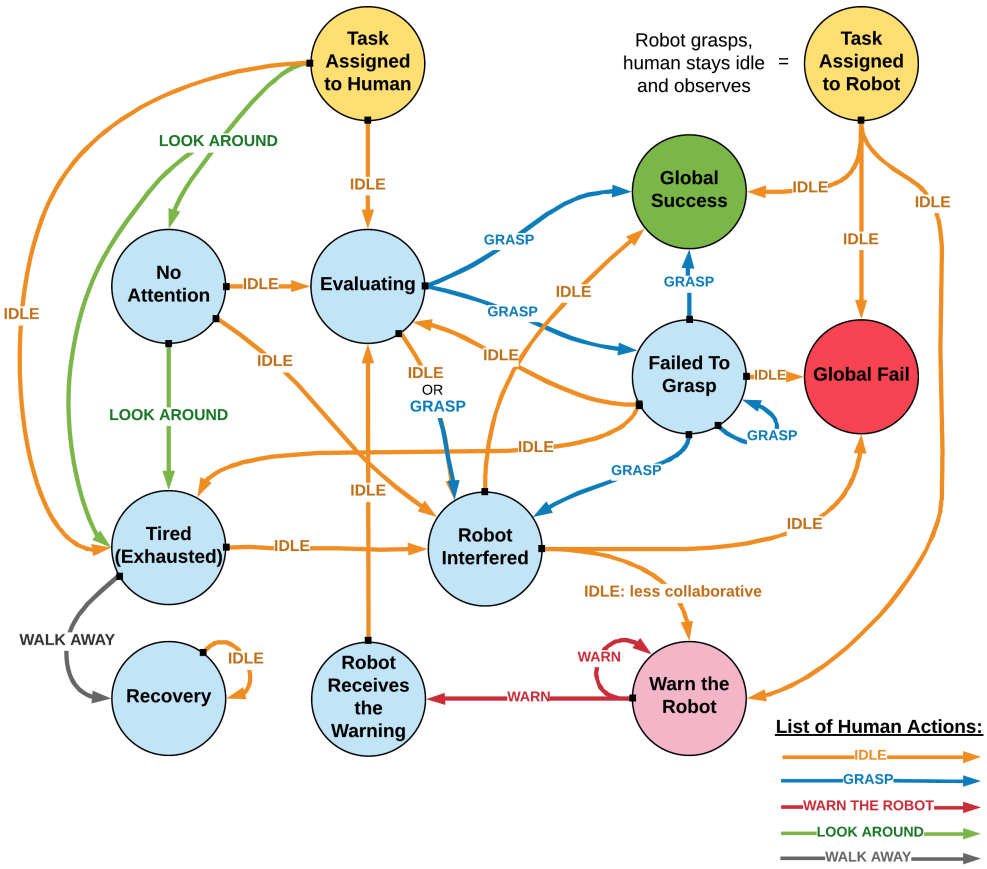


Figure 4.3: Human model: human state-action transitions

the given examples in Section 4.1. In the cases when the robot incorrectly estimates the human's need for help (detailed in Section 4.3.2), the human may transit to the *warn the robot* state, which may resemble the trust in the robot of the human in a given task. The warning state is followed by a human action of gesturing to stop the robot action (see Figure 4.4d). More details on the simulation environment that runs and executes both the human and the robot models, enabling a synchronized collaboration, are given in Section 6.4

Keeping the same scheme given in Figure 4.3, different state transition probabilities and rewards lead to different human types. We use four of them at this phase of the study for a proof of concept, referred to as *i*) normal, *ii*) stubborn, *iii*) distracted and *iv)* tired. Referring to the states shown in Figure 4.3, the type-*i* model is most likely to

succeed after a short period of evaluation of the task. The type-*ii* is more likely to fail the task, in our case the human fails to grasp, e.g., it is too heavy to lift for that person. In this case, the robot should intervene to assist before a global fail occurs. However, the human may want to try once again, iterating between evaluating and failing phases. The type-*iii* human mostly looks around resembling lost attention, or stalls more in the evaluating phase. Finally, type-*iv* is observed also through stalling in the evaluating phase and most likely ending up being exhausted, which then forces the human to leave the workspace for a recovery. The similarities in different types are expected to result in making anticipation difficult for the robot since, for example, staying idle or looking around could indicate the human states of *evaluating, no attention* or *exhausted*.

4.3.2 Experiments

We implement the proposed architecture in the Robot Operating System (ROS) and simulate our smart factory scenario with a conveyor belt by using the MORSE environment, utilizing the available human and PR2 robot models (as shown in Figure 4.4 and detailed in Section 6.4). All scenarios consist of several sequential task assignments for this purpose of simulating long-term collaboration. Each task of product inspection and storing starts with an initial assignment to either the robot or the human based on the product's weight and fragility. We consider only the cases where the task is assigned to the human, in order to keep our focus on anticipating the human's states and need for assistance. If the product is eventually inspected either by the human or by the robot, it is placed into one of the inspected-product containers (green containers in Figure 4.4) resulting in the *Global Success* state of all Markov models. Since the collaboration is a continuous process, we set a maximum allowed processing time for each product inspection, t_{max}. The conveyor belt waits when a package is between the human-robot team for t_{max} or for the package to be successfully processed, before starting a new task. If the product is not processed before t_{max}, it falls into the uninspected-product container (red container in Figure 4.4a) interpreted as a *Global Fail*.

At this phase of the study both the reactive and proactive robot models are hand-coded addressing the points discussed. In order not to bias the robot models, they are trained agnostic to the state transitions inherent in the human models, i.e., no data generated by the human models is used in training the robot models. During the experiments, we measure the state estimation accuracy using the human models as the ground truth. Our goal at this stage is to match the ground truth and do the tests against rather unexpectedly acting human models unknown to the robots.

The observations the robot receives 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 (listed in Figure 4.2). A state-of-the-art HAR module, inspired by existing studies, e.g., (Roitberg et al., 2014), has been implemented to recognize the constrained and distinct simulated human gestures from the body joints available. These are: the human is attempting to grasp (see Figure 4.4b), warning the robot (a special stop gesture shown in Figure 4.4d, idle (inactive), walking away (see Figure 4.4c) and looking around (from the body pose) in Figure 4.4e. The robot then uses these observations in both of the reactive and proactive models to estimate the next state (the human belief). We note that as our focus is in showing the effect of handling such unanticipated human behaviors on the collaboration performance, we use the same observation conditions for both models and our insights from the performance comparisons are agnostic to the HAR component that is used.

In each scenario consisting of several tasks, a human is first created as the normal type (referring to the human types introduced in Section 4.3.1), while the other types (i.e., the policies of the stubborn, distracted and tired MDP models) are then executed on the human during the scenario randomly, but becoming more likely as more tasks are assigned. By doing so, in each scenario the robot models interact with a human with changing levels of stubbornness, tiredness, and distraction. This induces more occurrences of the aforementioned unanticipated human behaviors over the course of the collaboration as the number of task repetitions increase. Additionally, the human may warn the robot when the robot estimates the human's need for help incorrectly. It would

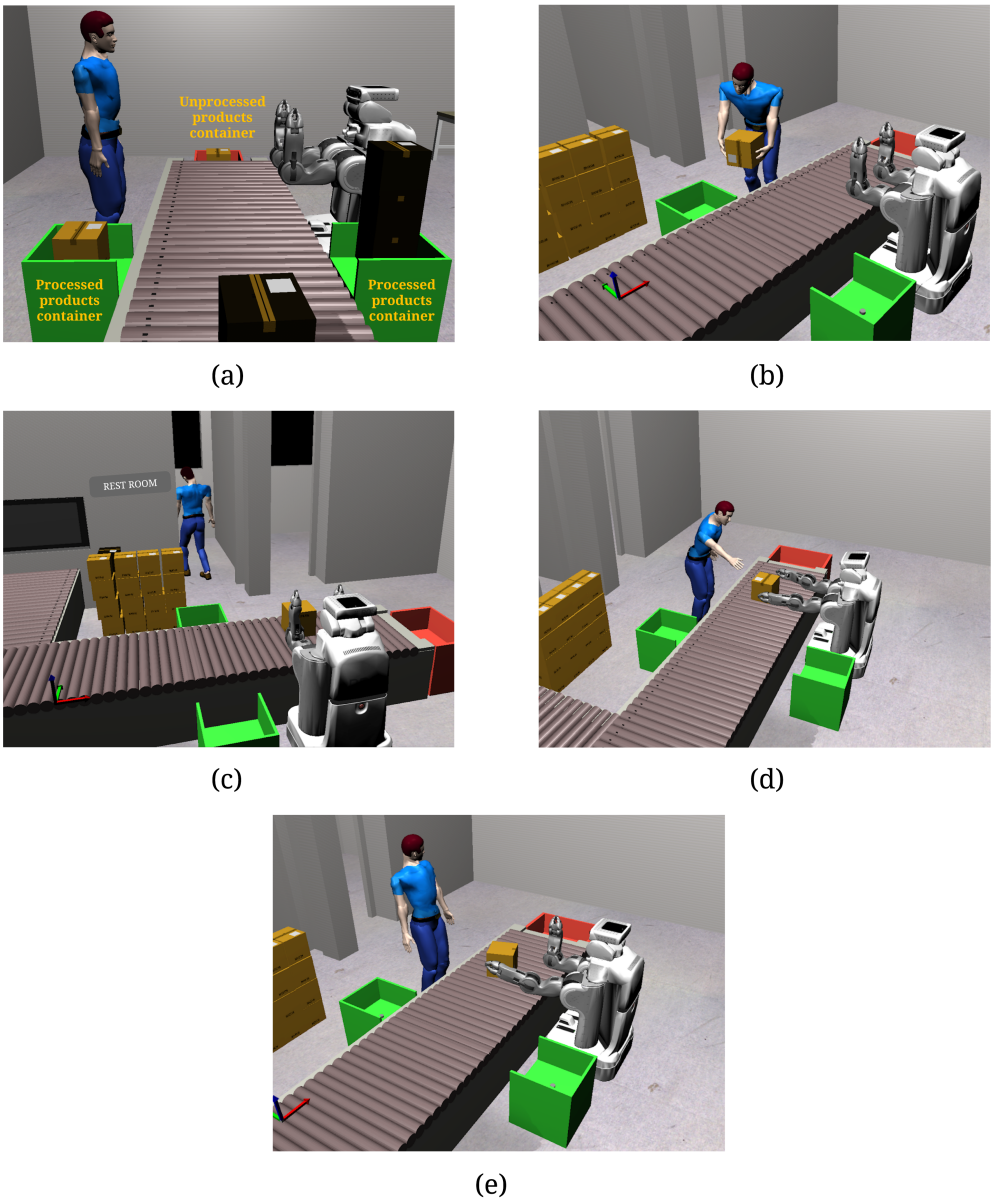


Figure 4.4: Our HRC scenario. (a) Idle human and robot, with containers shown; (b) Human grasps the product; (c) Human goes to the rest room (walks away); (d) Robot takes over the task and human gestures to stop the robot (warns the robot); (e) Distracted human while the robot is pointing out to remind.

be correct in the case when the human is tired. In a distracted case the human approves the help proportional to the time of the distraction. Moreover, when the human tries and fails to lift the object, it is more likely to approve the robot's taking over the task

unless the human is still trying to grasp. These cases are hard coded and the next state observations are input to the human MDP model executor by the system. The decision of a next state is random if there is no update of the task status or an action from the robot. This allows us to observe various scenarios of the humans randomly transitioning between the states.

Our hypothesis is given below:

Hypothesis 1. *Anticipating and taking into account such human variability, including unanticipated human behaviors, increases the overall efficiency (increased success rate over a shorter time) and the naturalness (less warnings received from the human, hence less intrusive robot behaviors) of an HRC.*

To evaluate if our proactive model can support *Hypothesis 1*, we gather the objective measures below during our experiments:

- **Human state distributions:** To show on average which states the human models have selected (based on the state transitions given in Figure 4.3) and how the state transitions change over time.

- **Estimation accuracy:** To compare how accurately both models estimate the human's true need for help, and to show how our proactive robot model performs in estimating human belief states in general (taking the interacted human models as the ground-truth).

- **Success rate:** The comparison of the success rates of a task with a human alone, a human and a reactive robot collaborating, and a human and a proactive robot collaborating.

- **Rewards gathered:** To show the change of overall rewards gathered over time, and to compare how many warnings the robot receives from the human (also in line with the wrong estimations of the need for help) in reactive and proactive mode. This hints at the naturalness of the collaboration.

- **Time to finish the task:** To compare the average time taken to complete a task with the reactive and the proactive models.

We use an updated version of the DESPOT online POMDP planner (Ye et al., 2017) to solve for both the POMDP and the MDP models, and execute them in real-time while interfacing the ROS environment through the *Planner* components as demonstrated in Figure 4.1.

4.3.3 Results

We run 5 different scenarios for each of the reactive and proactive robot models. Each scenario consists of 40 sequential task assignments. As mentioned, the humans exhibit the *two unanticipated conditions* listed in Section 1.2.1 randomly but with increasing likelihood over repeated task assignments.

This is also shown in Figure 4.5, which plots the average time a human spent in the given states over the simulation steps (i.e., task assignments). In particular, concerning *condition (1)*, as more tasks are assigned the humans stall more in the thinking phase (depicted as *evaluating*), are more distracted (depicted as *no attention*), are more likely to fail to grasp, become tired, or lose motivation. Such states are all summed and depicted as the *sum of all unanticipated behaviors* in Figure 4.5. Concerning *condition (2)*, as more unanticipated behaviors are performed the robot is increasingly convinced of the human's need for help and takes over the tasks. This leads to more warnings from the human in the case of wrong estimates (depicted as *warning the robot* in Figure 4.5).

Table 4.1 shows the results obtained from the experiments with the reactive and proactive robot models whereas Figure 4.6 demonstrates the change in the measured values over the simulation time. As is demonstrated in the figure, in the first 10 tasks there is no significant difference between the two robot models. The tasks are mostly succeeded by the humans conveying normal behaviors regardless of the robot models (see Figure 4.6b); therefore, the robots in these tasks constantly receive the maximum rewards (Figure 4.6a) and the tasks are completed efficiently (Figure 4.6d). However, as more

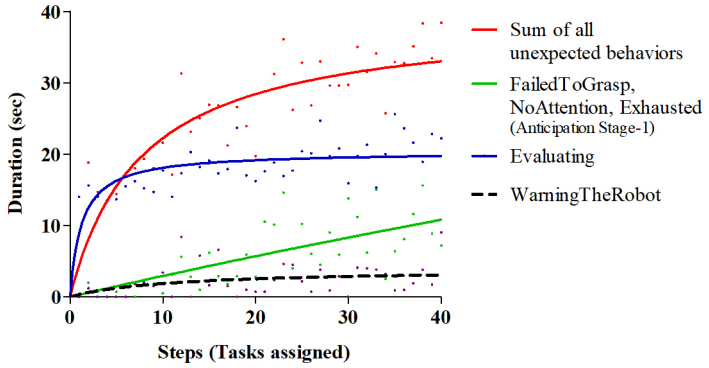


Figure 4.5: 2D scatter data of the states of the human models over the course of the tasks assigned (referring to Figure 4.3). A nonlinear curve fit over the data shows the overall trend of the average time spent in these states.

tasks are assigned, the average success rate falls and the differences in the performance of the two models are enhanced. This is directly in line with the increased likelihood of observing the unanticipated human behaviors over the simulation time. Before beginning the comparison of the two models, we show in Figure 4.6b that the success rate of such a human model alone is worse in the long-term than a collaboration with any of the robot models. Thereby, we underscore the importance of such robots collaborating with humans in tedious tasks.

The negative rewards are acquired from the task failures and the warnings received from the human, the former having more impact than the latter. The positive rewards are received only when the task succeeds. Therefore, the change in the accumulated rewards is directly in line with the changes in the success rate and the number of warnings received (Figure 4.6b and 4.6c, respectively). After the 30th task in Figure 4.6b, the average success rate keeps decreasing in the reactive case whereas it stabilizes in the proactive model to a rate of around 87% (see Table 4.1). Additionally, in the reactive case the robot continues to receive warnings during the collaboration while in the proactive model this amount is kept low. These two together contribute to a gradual decline in the increase of the accumulated reward in the reactive case (see Figure4.6a). On the other hand, as

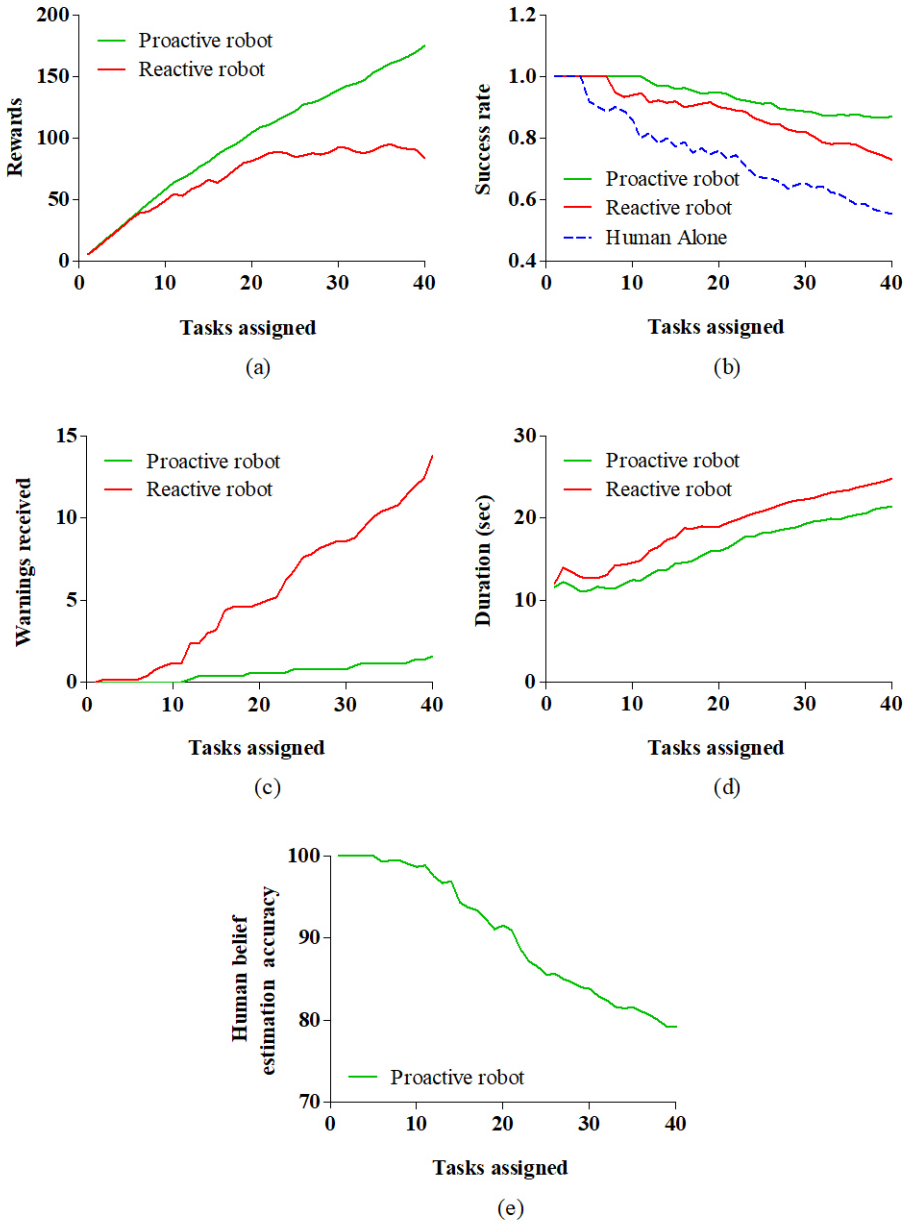


Figure 4.6: Comparison of the proactive and reactive robot models over the task assignments: cumulative (a) rewards acquired; (b) moving average of the success rates; (c) number of warnings received; (d) moving average of the task durations in seconds; (e) moving average of human belief estimation accuracy of the proactive model. Each result obtained from proactive and reactive robots are averaged over 5 trials.

shown in the same graph, the accumulated reward for the proactive case is affected much less by the changing human conditions.

The proactive model copes well with the unanticipated human behaviors mainly due to the model's significantly better performance in estimating the human's true need for help, as indicated by the *anticipation stage-2* estimation accuracies in Table4.1. Recall that the robot's anticipation of the human beliefs is divided into two stages, where *stage-1* involves reasoning about conditions such as the human being tired, distracted or incapable, and *stage-2* is to further estimate the human's need for help possibly as a result of the *stage-1* anticipation. In the reactive model the robot takes over the task through the given deterministic conditions without anticipating the *stage-1* conditions (see Section 4.2.2). Therefore, it is often either too late to judge the human's need for help or the robot's interference ends up with the human warning the robot, which results in the poor *stage-2* estimation accuracy of the reactive model (44.69%). As mentioned in Section 4.3.2, if the robot intervenes when the human is still trying to grasp, is only distracted for a short time, or is still evaluating, then the human stochastically rejects the robot's offer for help in line with *condition (2)*. Since these conditions are anticipated by the proactive model in the first phase, the proactive robot stochastically ends up estimating that the human may not yet need help or that the human needs to be reminded when she is distracted, rather than directly taking over the task based on the continuous observations received.

The average estimation accuracy of the proactive robot in the *anticipation stage-1* is about 79.26%, which has a direct influence on the *stage-2* estimation that is about 71.37%. Although we use a simulation environment and the human has a finite set of states and actions, we observe that it is still nontrivial for the robot to estimate the states given under the *stage-1*. As we demonstrate in Figure 4.6e, the average human belief estimation accuracy of the proactive model decreases over the task steps, being subjected to the higher frequency of unanticipated human behaviors. The model often confuses the states of *human may be distracted* with *human may be tired*, as expected, which is due to these two conditions causing similar human actions, i.e., looking around and staying idle longer. Such wrong estimations lead to task failures since, for example,

the robot points out the object to remind rather than offering help when the human is really tired.

Table 4.1: Proactive Robot vs. Reactive Robot Final Results

Type	**Reactive**	**Proactive**
Rewards (Avg.)	2.10	4.37
Number of Warnings per Scenario (Avg.)	13.8	1.6
Anticipation Stage-1 Estimation Accuracy	Not Applicable	79.26%
Anticipation Stage-2 Estimation Accuracy	44.69%	71.37%
Success Rate	0.73	0.87
Task Duration	24.77*secs*	21.40*secs*

We also measure the average task durations in both models (whether it succeeded or failed). As shown in Table 4.1 the proactive model improves the time a task takes over the reactive case by about 14%, i.e., 3.37 seconds. This can be thought of a significant change especially in an industry with mass production. The improvement is again driven by the hierarchical anticipation in the proactive robots. First of all, a successful estimation of a state in the anticipation stage-1 allows the proactive robot to estimate the human's need for help faster, whereas in the reactive case this needs to wait until observing the human failing to grasp and giving up. Secondly, the robot starts the *planning* in the *stage-1* (see Figure 4.2) while it still observes for the *stage-2*, the estimation of which then leads to a direct execution of the *grasp*. In the reactive case, however, even though the robot's taking over the task is approved by the human, the robot still needs to plan for its grasping motion execution.

Finally, our basic HAR system works with approximately 10% noise, which emerges from the misclassification of human actions, especially for warning the robot and grasping gestures producing similar observations (see Figure 4.4). To understand the effect of this noise on the human belief estimation accuracy, we also tested providing the actual human actions as direct inputs to the robot, thereby taking HAR out of the loop. In this case the estimation accuracy of the proactive model increased, as expected, to around 85%.

For an actual deployment in a noisy real-life environment, where the observations of the hidden human states are not limited to a small set of human actions as in our simulation, belief estimation accuracies are likely to be lower, mainly due to lower accuracies in the HAR system.

In conclusion, we show that the common human conditions of *(1)* intention being irrelevant to the assigned task (e.g., due to motivation loss or tiredness), and *(2)* disapproval of the robot's help (e.g., no trust in the task, not needing/desiring the robot's help) lead to a drastic decrease in the overall success rate of the collaborative tasks. Since these conditions are expected in HRC scenarios, we show in simulation that a cobot that can anticipate and handle these conditions yield more efficient (i.e., increased success rate, lower task duration) and more natural (i.e., less intrusive) collaboration when the collaborating human demonstrates these behaviors. We show the feasibility and effectiveness of this concept through our two-stage anticipatory decision-maker, A-POMDP, hierarchically and stochastically reasoning about such human behaviors. Hence, we support *Hypothesis 1.*

4.4 Conclusion

This chapter examines the effects of a robot's anticipation and response to the unanticipated human behaviors that could lead to human errors, which may be observed in a collaboration with repeated and tedious tasks. In this chapter, as a proof of concept of our novel anticipatory modeling scheme, we call A-POMDP, we use simulated human models in our experiments, which provide a more robust setup to test such various uncommon and unanticipated conditions rather than expecting real human feedback in a rather longer term user studies. We are aware of the possible biases in the experiments which could be introduced by the simulated humans due to their limited reflections (limited actions) of the hidden human states 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 expected to be observed in a real human. Our model has generated promising results, encouraging us to move towards validating these results with real hu-

man experiments through a similar model design but training with real human data, which is covered in Chapter 6. We consider real humans' trust building in the robot over the course of the collaboration, which in the simulation setup we assume random, thereby creating a more difficult experimental case. The results we obtain in the user studies are inline with our findings in simulation (see in Section 6.3.2); hence, showing the reliability of our approach and the results obtained in this chapter.

5 Long-Term Adaptation: ABPS *

In this chapter, we focus on the long-term adaptation of a cobot to various human characteristics, implementing the policy selection component and the decision model library component of our framework, FABRIC (see Figure 3.1). As a key component of cobots working with humans, existing decision-making approaches try to model the uncertainty in human behaviors as latent variables. However, as more possible contingencies and characteristics are covered by such intention-aware models, they face slow convergence times and less accurate responses. For this purpose, we present a novel anticipatory policy selection mechanism built on existing intention-aware models (e.g. our A-POMDP models), where a robot is required to choose from an existing set of policies based on an estimate of the human's long-term characteristics. Each of these intention-aware models anticipates and adapts to a different human's short-term changing behaviors. Our contribution is the Anticipatory Bayesian Policy Selection (ABPS) mechanism which selects from a library of different response policies that are generated from such models, and converges to a reliable policy after as few interactions as possible when faced with unknown humans. The selection is based on the estimation of the human in terms of long-term workplace characteristics that we call types, such as level of expertise, stamina,attention and collaborativeness. Our results show that incorporating this policy selection mechanism contributes positively to the efficiency and naturalness of the collaboration, when compared to the best intention-aware model in hindsight running alone.

*The content of this chapter was originally published as 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), May 13-17, 2019, Montreal QC, Canada. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 77–85. (Görür, Rosman, and Albayrak, 2019).

5.1 Introduction

In this chapter, we implement our novel anticipatory policy selection mechanism, ABPS, to realize our objective of an extended long-term adaptation to humans (in Section 1.2.2). In Chapter 4, we design our A-POMDP model that adapts to a human's short-term changing behaviors. Our focus in this chapter is on a robot's adaptation to human long-term behaviors, i.e, human types. We create a policy library by randomly constructing different robot models based on our existing model design. Through this random generation, we are agnostic to specific human types and behaviors modeled. Then, ABPS selects and converges to an optimal policy from the library in short time based on a human's estimated long-term workplace characteristics, such as level of expertise, stamina (or fatigue), attention and collaborativeness. Instead of modeling known human types as a latent variable, we estimate unknown human types from the observed human behaviors using Bayesian belief estimation and correlate them to a policy performance. To our knowledge, this is the first time such a policy selection mechanism has been proposed complementing intention-aware planning approaches in HRC, providing fast and reliable anticipatory decision-making for both long-term (to unknown human types) and short-term adaptation (to their changing behaviors through A-POMDP models).

Our goal is to show that integrating such a policy selection mechanism contributes positively to the efficiency (e.g. time to finish a task, success rate) and naturalness (e.g. a human's increased willingness to collaborate) of the collaboration, when compared to the best intention-aware model in hindsight running alone. At this phase, we conduct experiments using the same simulated HRC scenario detailed in Section 6.4, i.e., at a conveyor belt for the task of inspecting and storing various products with different weights. We then repeat the experiments on a collaboration task with real humans in Chapter 6. In simulation, different types of modeled humans, responsive to both robot actions and changing environment, collaborate on the task autonomously with our adaptive robot decision-making mechanism implementing ABPS (in Section 5.2). We present our experiments and analysis on our policy selection through its effects on the efficiency and the naturalness of the collaboration (Section 5.3).

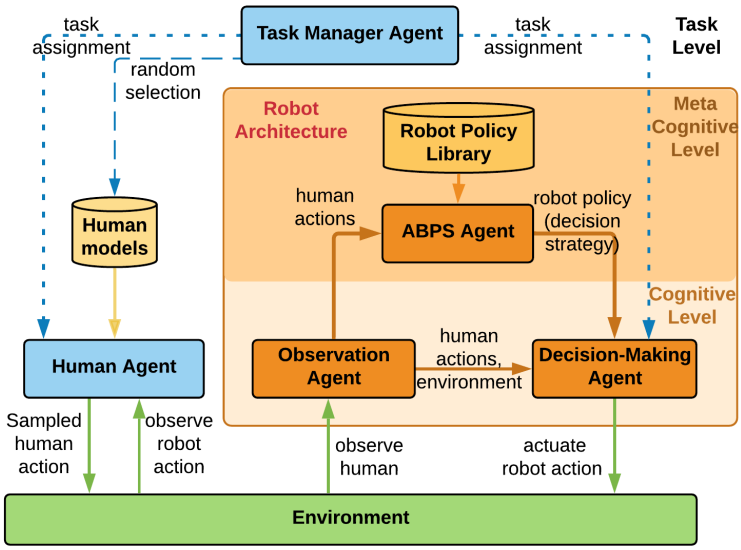


Figure 5.1: Anticipatory Bayesian Policy Selection (ABPS) mechanism integrated on top of our A-POMDP human intention-aware models, running against simulated human models.

5.2 Methodology

5.2.1 Overall Robot Decision-Making Flow

Figure 5.1 gives an overview of how the decision-making process takes place in a human-robot collaboration, when interacting with our simulated human models. This elaborates the interaction between the policy selection, the decision model library and the decision-making components of FABRIC in Figure 3.1 In the upper layer, tasks are created and assigned to either the human or the robot. In the meta-cognitive level the *ABPS agent* selects one decision strategy among a *policy library* according to the anticipated type of the human partner. Following the selection, a decision strategy is forwarded to the cognitive level for the *decision-making agent* to execute. In our implementation, each decision strategy is a robot policy comprised of optimal actions for each possible belief over the world states and generated when a POMDP robot model is solved for maximizing expected rewards (see in Section 5.2.2.1).

The cognitive level of the system in Figure 5.1 has been the focus of similar studies, which involves a robot's decision-making agent acted upon one precomputed anticipatory

model, in our case a A-POMDP model detailed in Chapter 4. In this work, we focus on the *meta-cognitive level*. It includes the policy library constructed from such handcrafted Markov models (detailed in Section 5.2.2.1) and our ABPS mechanism consisting of human type (belief) estimation (Section 5.2.2.2) and policy selection with an exploration heuristic for a quick adaptation to a class of human types (Section 5.2.2.3).

5.2.2 Anticipatory Bayesian Policy Selection (ABPS)

Our approach is based on the Bayesian policy reuse (BPR) algorithm (Rosman, Hawasly, and Ramamoorthy, 2016). The definition of ABPS is given with Definition 1.

Definition 1 (**ABPS**). *An ABPS agent is equipped with a policy library* Π *to act appropriately in the context of some human types and tasks in HRC domain. The agent is presented with a human collaborator having an unknown type in a known task, which must be solved within a limited time and small number of trials. The goal of the agent is to select policies from* Π *for the new and possibly unknown human type, over which it has a belief distribution* $\beta(.)$*, while minimizing the total regret in a limited time. Minimizing the regret in this domain is defined by increasing the task success rate and decreasing the amount of warnings received from a human collaborator, relative to the best alternative from* Π *in hindsight.*

ABPS measures the similarity between an unknown human type and previously known types to identify which policies may be the best to reuse. In this case, a collaborating 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. The similarity of types is extracted from offline training with some known types and by utilizing this trained model online, constructing $\beta(.)$. The general algorithm of our robot's anticipatory decision-making is given in Algorithm 1. We first detail how a policy and the library Π is constructed in Section 5.2.2.1. The observation signals and the observation model for the human type belief update (see *line 7, 8* of Algorithm 1) are detailed in Section 5.2.2.2. Then, the policy selection step and the construction of the performance model used in this step (in *line 4* of Algorithm 1) is described in Section 5.2.2.3.

Algorithm 1 Anticipatory Bayesian Policy Selection (ABPS)

Require: Human type space $\mathcal{T}$, robot policy library Π, an observation vector for observed human behaviors σ in an observation space Ω, an observation model to match observables to known human types $P(\Omega|\mathcal{T},\Pi)$, utility as accumulated discounted reward obtained from running a policy U, a performance model $P(U|\mathcal{T},\Pi)$, number of tasks K, exploration heuristics $\mathcal{U}$.

1: Train offline for performance and observation models.
2: Initialize a belief: β^0 uniform distribution from the prior $\mathcal{T}$.
3: **for** task IDs $t = 1...K$ **do**
4: Select a policy $\pi^t \in \Pi$ using β^{t-1} and performance model, $P(U|\mathcal{T},\Pi)$, using $\mathcal{U}$ in Equation (5.3).
5: Apply selected policy π^t to the task and the human.
6: ***wait***(until the task t is completed)
7: Obtain observations σ^t from the human and the environment emitted during the task.
8: Update belief β^t using σ^t by belief update function in Equation (5.1).
9: **end for**

5.2.2.1 Policy Library Construction

To generate many policies for the policy library, we use the anticipatory robot model design (see the proactive robot model in Figure 4.2) as a base, which we have shown to perform well in this context in Chapter 4. Recall that, the goal of this state machine is to first anticipate a human's unanticipated behaviors, i.e., her task related availability, motivaiton and capability. Then, moving from these estimations, the robot anticipates whether the human needs assistance or not from the robot's perspective. We generalize the states and the robot actions to comply with several collaboration tasks. The robot actions are to wait for human (*idle*), plan for assisting action (*planning*) and *assist* human as shown in Figure 4.2. Positive rewards are acquired when a task has been accomplished by any agent and negative rewards are for a task failure or when warnings are received from the human. The latter is to encourage the planning to be less intrusive, i.e., the robot will not offer assistance unless it is deemed part of the optimal policy.

The offline generation of different policies to construct the policy library is done by adjusting T and O: the state and observation probabilities of the A-POMDP model corresponding to different human types (see Section 4.2.2 for the POMDP design). Changes in T correspond to different transitions of a human's internal states, e.g. a robot policy assumes the human tires faster (related to the stamina-level) or the human needs assistance when she is not capable (related to the collaborativeness). Whereas changes in O

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 a beginner (related to expertise) depending on her type, both of which should be handled differently by the robot. Additionally, by adjusting O, we are able to make the model a partially observable, mixed observability or fully observable Markov decision process (POMDP, MOMDP or MDP, respectively). We randomly adjust the probabilities as mentioned above to generate various Markov decision models, each of which handles a unique human type, and solve for their optimal policies to construct our policy library Π. The main reason we move from a base model as in Figure 4.2 is to limit the arbitrary generation of robot policies to avoid overloading the space with unreliable candidates (Albrecht and Stone, 2018). This way we also show how we integrate ABPS to existing intention-aware models.

5.2.2.2 Human Type Belief Estimation

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, $\beta(.)$. In order to train such a model, we generalize some characteristic human features to approximate a human type. These features are inspired from (Gombolay et al., 2017; Ji, Lan, and Looney, 2006; McGuire et al., 2018) and are stated to be crucial to be known by a cobot. These are a human's expertise, attention, stamina-level and collaborativeness. The last term is a more general description of a human's acceptance rate of a robot's offer for assistance. The type space consists of many human types by adjusting the level of these features, e.g. a human with beginner skills, pensive, bad 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. More details on the simulated human types in type space τ are given in Section 6.4.1.1.

The human type estimation model is used by ABPS as *a priori* information, which we call the *observation model.*

Definition 2 (**Observation model**). *For a robot policy π, a 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 the 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* (detailed in Section 5.3.1.2). The observation signals are emitted by the collaborating 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 6-D boolean vector with the following observables similar to the observation set used in A-POMDP models (in Figure 4.2): $\{$*human is detected, human is looking around, human has taken a task related action and succeeded in it (e.g. grasping and lifting a package in our scenario), human has taken a task related action and failed, human is warning the robot, human is idle*$\}$. 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 finishes (see *line 6, 7, 8* of Algorithm 1). Finally, the type belief update is Bayesian, given by

$$\beta^t(\tau) = \frac{P(\sigma^t|\tau,\pi^t)\beta^{t-1}(\tau)}{\sum_{\tau' \in \mathcal{T}} P(\sigma^t|\tau',\pi^t)\beta^{t-1}(\tau')}, \quad \forall \tau \in \mathcal{T} \tag{5.1}$$

where β^{t-1} stands for 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.

5.2.2.3 Policy Selection with Exploration Heuristics

The policy selection process of the robot is based on an exploration heuristic called *expected improvement (EI)* (Rosman, Hawasly, and Ramamoorthy, 2016). As stated in

line 4 of Algorithm 1, this algorithm runs on another trained *a priori* model called the *performance model.*

Definition 3 (**Performance model**). *The performance model, $P(U|\tau,\pi)$, is a probability distribution over the utility, U, of a policy π when applied to human type $\tau \in \mathcal{T}$.*

The system utility, U, is the accumulated discounted reward received after a policy is run (see Section 5.2.2.1 for the immediate rewards a robot obtains during a task). All the combinations of known human types $\tau \in \tau$ and the robot policies $\pi \in \Pi$ are repeatedly run against each other offline to generate our *performance model.* 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 (5.2) and achieves the utility U^+.

$$\pi' = \arg\max_{\pi\in\Pi} \sum_{\tau\in\mathcal{T}} \beta(\tau) P(U^+|\tau,\pi) \tag{5.2}$$

Because the choice of U^+ directly affects the performance of the exploration, its selection is crucial to the performance of this exploration. The *expected improvement* 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 below.

$$\pi' = \arg\max_{\pi\in\Pi} \int_{U^\beta}^{U^{max}} \sum_{\tau\in\mathcal{T}} \beta(\tau) P(U^+|\tau,\pi) dU^+ \tag{5.3}$$

$$= \arg\max_{\pi\in\Pi} \sum_{\tau\in\mathcal{T}} \beta(\tau)(1 - F(U^\beta|\tau,\pi)) \tag{5.4}$$

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.

5.3 Evaluation on Simulation

5.3.1 Experiments

In this section, we briefly talk about the simulation environment we have used for our experiments in Section 5.3.1.1 and how human models are crafted towards simulating short and long-term changes in human behaviors and types (Section 5.3.1.1). After that, we describe the training phase to construct the library and the estimation models for ABPS (Section 5.3.1.2). Finally, we give the details of how we conduct the real-time experiments and our performance metrics in Section 5.3.1.3.

5.3.1.1 Simulation Environment

We use the same simulation environment in Chapter 4, given in Figure 4.4. The simulation environment allows our robotic system to run a long-term collaboration. Such long-term experiments make it possible for the robots to face many different changing human types and behaviors under various conditions. As a result, we do not have to be limited to constrained environments and human interactions. This helps us train very accurate models of the interaction, as well as run rigorous tests on our system facing and covering more uncertainties of humans. We run the same scenarios in Section 4.3.2, which consist of several sequential task assignments to simulate a long-term collaboration.

The task definition is also the same as in Chapter 4, that is, product inspection and storing. We, again, only consider the cases where a task is assigned to the human in order to keep our focus on anticipating the human's type and behaviors and correctly assess her need for assistance. The collaboration is when the robot correctly estimates the human's such need and helps with the task. As stated in Algorithm 1, a new policy is selected after each task is finalized.

We have modeled many different human types for our collaboration scenarios, using the same simulated human modeling scheme in Section 4.3.1. More details on the design of human models are given in Section 6.4.1.1. Before moving to the user studies, in our proof of concept experiments, we run randomly generated models to reflect changing

and unknown levels of expertise, stamina, attention and collaborativeness. The models reflect them as actions, which are observations for the robot obtained from 3D human body joints always available directly from the simulated humans (see Figure 4.4). A human chooses an action based on the collaborating robot's action, the state of a task, the human internal states and human internal goals, assuming that any human actions toward her goal may imperfect (Hiatt et al., 2017). During a task, the robot collects all the observations emitted from the human and the environment and averages them to construct our 6-D observation vector for the task (given in Section 5.2.2.2). The observation vector also has features of the human's success in the task. After each attempt the human makes at grasping, the robot associates the human's attempts with expertise.

The human models also govern the human's responsiveness to the interacted robot actions through a dynamic transition function that is changing over the course of the interactions according to the number of times the robot interfered in a task and the number of tasks handled so far (see how this is handled in Section 6.4.1.1). This leads to updates in a human model, and so their changing behaviors generated. An example to such responsive behaviors is that a human becomes less collaborative as their robot partner selects wrong policies, e.g., the robot takes over a task 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 (see Figure 6.28). Another example is that a transition to the state of being tired depends on the number of tasks already handled. This modeling scheme leads to human simulations exerting dynamic behaviors changing in response to the robot decisions and with a small random factor. In addition, by changing the state transition probabilities and rewards we create various human types with changing characteristics, e.g. *a beginner, tired and collaborative human*, a *mid-expert, high-stamina and less-collaborative human.*

5.3.1.2 Training Phase

We have generated many different robot policies to build Π, the library. During the generation, all robot model designs have random transition and observation probabilities assigned (see Section 5.2.2.1); therefore, they are agnostic to the state transitions inherent in the human models. In the end, 20 policies have been selected for their use in the experiments in order not to overload the policy library. This selection is based on how well they performed overall against many different randomly generated human models, i.e., discarding the worst ones, and how distinct their *performance models* are (see Definition 3) from the other policies, i.e., grouping the similar ones. 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 trade-off between these two is more obvious when it comes to non-collaborative 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 more 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 this purpose, we have crafted 16 different human types (known types), again with the goal of each generating as distinct set of observations (human actions) as possible toward a more heterogeneous distribution. Generating distinct observations means creating human types with extremes of the four features, 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. Increasing the number of known human types would yield less accurate type estimations with a higher convergence time. However, we note that since each of the 16 human models are stochastic, they still generate a diversity of behaviors after random sampling (see Section 5.3.1.1).

During the training phase we run each of the 20 robot policy against the 16 known human types for 50 sequential tasks. In total, human and robot models accomplished 16000 interactions (16000 task instances), which is very difficult to manage in real-life scenarios. The *performance model* and the *observation model* (see Definition 2) are constructed after this training phase.

5.3.1.3 Real-Time Experiments Phase

Our goal is to support the hypothesis below:

Hypothesis 2. *A single intention-aware robot model is limited in its adaptation to various human types. Our ABPS mechanism provides broader adaptation to various and changing human types leading to more efficient and natural collaboration, while maintaining fast and reliable convergence to the best policy.*

To explore the hypothesis, we conduct two lines of experiments and gather the objective measures below:

Experiment-1: *The goal of this experiment is to see the performance of ABPS in terms of how fast and reliably it converges to the best policy performance*. For this purpose, we compare an ABPS robot collaborating with a randomly created human type unknown to the robot, with the best performing robot policy for that type when collaborating with the same human. The best robot policy is picked from the library as the best performer in hindsight for the experimented human type. Both robots interact with the human for 30 sequential task assignments and this scenario is repeated 10 times for each. We measure the moving average regret of both of the robots and compare the change over time. A regret for a selected policy $\pi \in \Pi$ is:

$$R_{\pi}^{\tau} = U_{\pi*}^{\tau} - U_{\pi}^{\tau} \tag{5.5}$$

for a human type $\tau \in \mathcal{T}$ and the best policy $\pi* = \arg\max_{\pi \in \Pi} U_{\pi*}^{\tau}$.

Experiment-2 *The goal of this experiment is to show the contribution of our ABPS model to the adaptation capability of a robot through increased efficiency and naturalness*

in the collaboration. For this purpose, we compare the performance of ABPS robot with the library's best policy in hindsight running alone, when collaborating with various types of humans unknown to the robot and changing during the operation. The robot is unaware of this change, which might be thought of as a shift change in a factory environment. The best policy for this experiment is the overall best performer in the policy library, picked after the training phase when averaged over all the interactions. For this purpose of simulating unknown human characteristics, we randomly crafted 10 different human types offline. At every 30th task in a scenario (enough to let ABPS converge), the human type changes drastically to another unknown human type in a certain order, and the robot has to adjust its responses accordingly. The same order is repeated 5 times (300 sequential tasks in each scenario) for each strategy to average and smooth the human type characteristics and observe the long-term performance of the robots. We analyze the following for both of the strategies:

- *How the human state distribution and the average reward the robot collects change over time.* This shows the effect of such type and behavior changes on a single intention-aware model that introduces those changes as a latent variable versus the ABPS mechanism and its adaptation capability.
- *Success rate, the number of warnings the robot received from humans and the approximate time a task takes.* These are to compare the task efficiency and naturalness of the robots. We also analyze the trade-off between time and success rate as made by the policy selector to avoid human warnings.

5.3.2 Results

The results of *Experiment-1* are illustrated in Figure 5.2. ABPS naturally has a uniform belief distribution over the human types when first initialized, and has selected different policies (best performers of the library) until its belief estimation converges (as shown by higher deviations). It has already reached a very close performance to the best policy after the 6th task, by correctly selecting the same policy at that time (i.e., at the 6th

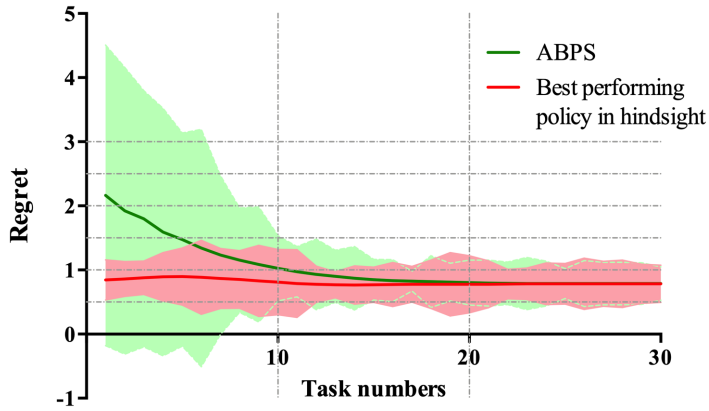


Figure 5.2: Moving average regret over time with error bars denoting the standard deviation, collected by ABPS and the best performing policy in hindsight for the experimented human type.

iteration). The difference between the moving average regrets of both strategies decreases to equalize after this point showing the fast convergence of ABPS (one-way ANOVA: $F(1, 58) = 0.017, p = 0.895$). It should be noted that a zero regret cannot be reached even by the best policy and there is a constant variance on the values. This reflects that the human's behaviors are constantly changing over the course of the interaction. In a real world setup we may have more stable behaviors from people; however, with this experiment we point out the adaptation performance of ABPS.

For *Experiment-2*, the results are shown in Figure 5.3 and Figure 5.4. We compare our ABPS with the overall best policy in hindsight in the policy library. To reflect the dynamic nature of our human simulations, Figure 5.3a shows the average duration of each different human state in one task, and how this duration changes over the task assignments. Within every 30 tasks we visualize a human's changing availability, motivation and capability which are generated by a single type. These behaviors are reflected under the human states of *failed to handle* the task, being *tired*, being *distracted*, *evaluating* (spending time to figure out how to achieve) and *warning the robot* (when a human does not want the robot's assistance) as shown in Figure 5.3a. The drastic changes of these states after every 30 tasks shows the different long-term characteristics, i.e., types, of human workers. For example, the human which took a shift between the 90-120th tasks

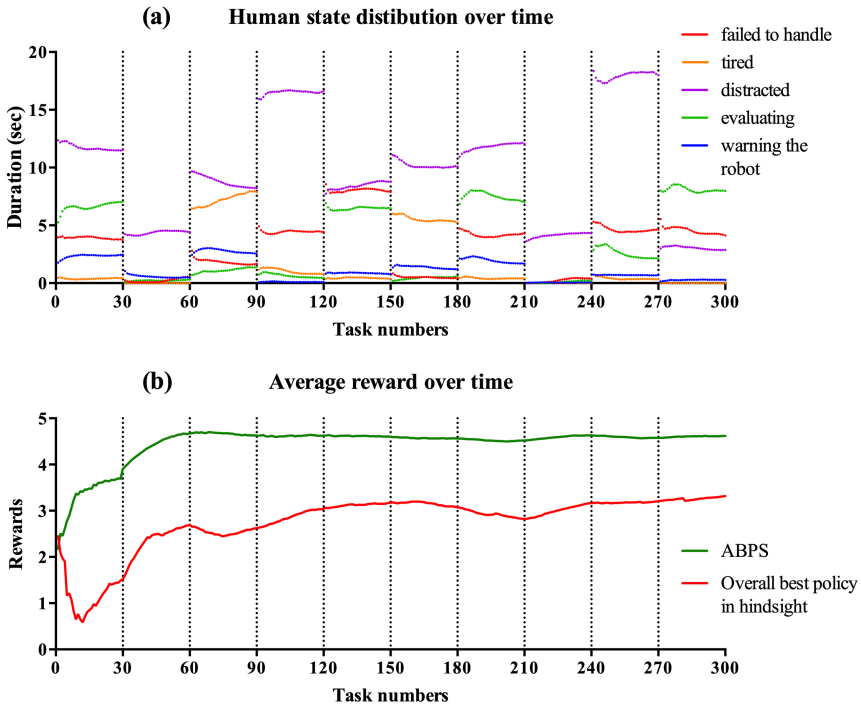


Figure 5.3: (a) Human state distribution over time showing the changes on human types and behaviors. (b) Moving average reward over time collected by ABPS and the overall best single policy in hindsight (best performer in the library).

is more of a distracted type whereas between the 210-240th tasks is a more expert human, with less failures and evaluating time.

We note again the dynamic nature of each human models and Monte Carlo sampling yield various behaviors as shown in Figure 5.3a, e.g. an expert human can also fail sometimes. This causes many fluctuations in the moving average rewards collected after each task as illustrated in Figure 5.3b. It is noticeable how the rewards are affected by different humans starting to interact with the robot (at every 30th task instance). In most of such cases, the overall best policy model is affected more negatively than ABPS. For example, between the 180-210th tasks the human type resembles more beginner and less collaborative behaviors than the others due to the number of warnings she made to the robot and the number of failures. Such a difficult human type causes a drop in the average reward the overall best policy collects, whereas our ABPS selects another policy (the best for that type) adapting to the situation that results in almost no change

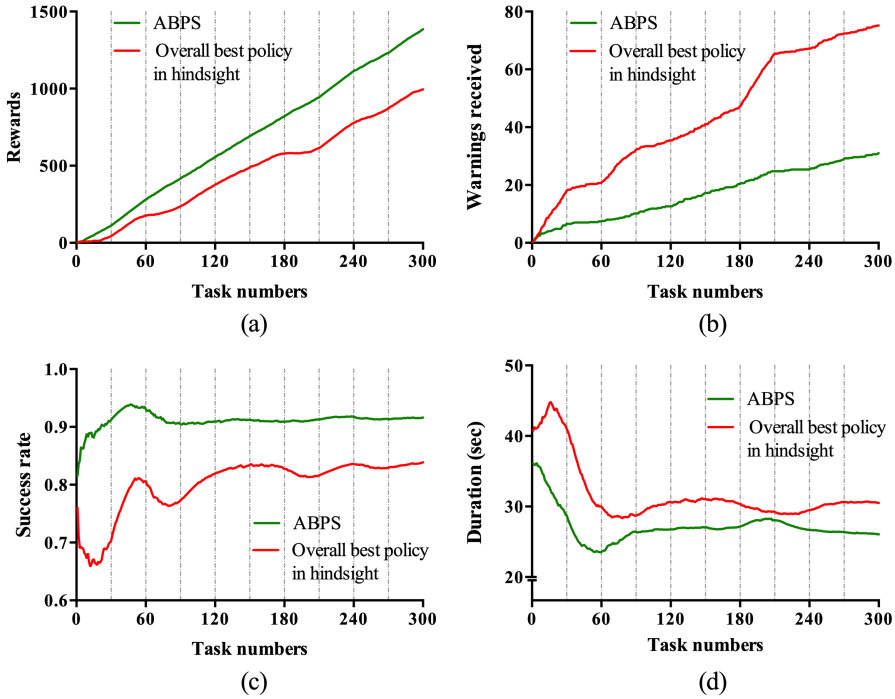


Figure 5.4: Comparison of the robot with ABPS and the robot with the overall best single policy in hindsight over the task assignments: (a) cumulative rewards acquired; (b) number of warnings received; (c) moving average of the success rate; (d) moving average of the task durations in seconds.

to the reward level. This policy has avoided possible warnings by not taking over the human's task directly but encouraging the human and waiting patiently for the human to handle it unless it is too late for the task. ABPS shows its necessity in such a realistic system, especially at the beginning of the experiment where the overall best policy is clearly not suitable for that type of human. In general, one-way ANOVA tests show that the accumulated rewards of ABPS is significantly larger (see in Table 5.1) with a mean difference of 39.2%. This and the almost stable reward level in Figure 5.3b shows that ABPS provides faster and more reliable adaptation to these difficult cases.

As shown in Figure 5.4b and Figure 5.3b, the warnings received by the overall best policy accumulate greatly, especially against the humans between the 60-90th and 180-210th tasks (the former is tired, the latter is a beginner and both are non-collaborative), whereas the ABPS robot has successfully adapted to the situations and accumulated

fewer warnings. This shows such an adaptation is necessary for the naturalness of the collaboration, which also affects the success rate of the system and finally the accumulated reward being the combination of both (Figure 5.4c and Figure 5.4a, respectively). During these task intervals, ABPS trades off the duration of the tasks with the efficiency and naturalness through the selected policies (see Figure 5.4d and Figure 5.4b). The human type is likely to exert slightly more non-collaborative behaviors and the policy selection of the ABPS favors avoiding interference, waiting for the human to succeed, collecting more rewards through higher success rates and *fewer warnings*. On the other hand, the best policy offered assistance and took over the tasks in general. This resulted in more warnings received by the best policy and lower success rates as the robot had to cancel its action after the warning. However, it led to faster times of completing the task, most of which were a failure. Despite such small trade-offs, ANOVA tests show that ABPS leads to significantly better efficiency (with 9.5% higher mean success rate and with 14.6% less mean task duration) and more natural (with 58.8% fewer warnings received) human-robot collaboration, as summarized in Table 5.1. All these analyses support **Hypothesis 2**.

Table 5.1: Final results from Experiment-2: ABPS vs. overall best policy ($\mu = mean$)

Type	$\boldsymbol{\mu_{ABPS}}$	$\boldsymbol{\mu_{BestPolicy}}$	**ANOVA**
Discounted rewards	4.62	3.32	$F(1, 598) = 74.11$, $p < 0.0001$
Total warnings	31	75.2	$F(1, 598) = 70.37$, $p < 0.0001$
Success rate	0.92	0.84	$F(1, 598) = 29.94$, $p < 0.0001$
Task duration	26.04 *secs*	30.49 *secs*	$F(1, 598) = 15.61$, $p < 0.0001$

5.4 Conclusion

We introduce our novel anticipatory Bayesian policy selection (ABPS), in an HRC setup as a complementary solution to the existing intention-aware robot decision-making models, e.g. our A-POMDP model, that provide rather short-term adaptation due to their scale of covered human behaviors. We examine the effects of our ABPS on a collaborative robot's adaptation to unknown human types and their changing behaviors in a long-term collaboration. Our results have shown that ABPS is a fast and reliable policy

selection mechanism for HRC scenarios. Having such a mechanism on top of a robot's intention-aware decision-making contributes positively to the efficiency and naturalness of the collaboration by providing a better adaptation, i.e., extended with long-term adaptation, to the collaborating human, when compared to the state-of-the-art robot decision-making models running alone. In our experiments, we have utilized our fully autonomous architecture we describe in Chapter 3, capable of running ABPS along with a robot's decision-making and observation agents, in our real-time human-in-the-loop simulation setup. The promising results we have obtained have encouraged us to move towards validating these results with the real human experiments, this time having the ABPS mechanism trained with the real and the simulated human data, which is covered in Chapter 6.

6 Simulation, Deployment and User Studies

In this chapter, we first implement the sensing, actuation, and world model components of our framework, hence completing FABRIC. Then, we transfer the results from Chapter 4 and Chapter 5, which have been evaluated on simulation so far, into a real-world experiment and validate them by means of user studies. Evaluating the adaptation capability of a social cobot relies heavily on the data collected from the people participating in a user study. This introduces a limitation on the diversity of human behaviors a robot faces during a lab study due to, e.g., constrained environments and a limited amount of participants. Hence, the full dynamics of humans are mostly overlooked in existing HRC solutions. Our goal is to remove such limitations on human behaviors and create a more realistic evaluation environment. For this purpose, we first design a novel simulated human model that samples a vast range of human behaviors. The simulation helps our solutions to train on a large scale data and to be rigorously tested. Another contribution is our novel experiment setup. Through the case studies, we show that the setup induces a cognitive load on the human participants and so invokes unanticipated human behaviors, e.g., lost motivation, human errors. After deploying our framework on the real setup, the first round of user studies proves our findings in Chapter 4, the short-term adaptation skills of our A-POMDP. We also show that our simulated human models are able to generate reliable human behaviors. Finally, for a broader impact, we integrate our complete framework, including ABPS, as a human-in-the-loop autonomous system, and show that the thesis work is applicable and able to provide extended long- and short-term adaptation to various humans.

6.1 Introduction

In this thesis, we follow the conventional way of validation in robotics solutions. Our first goal has been to train and rigorously test our robot models with a large scale data on a simulation that provides various dynamic conditions (as covered in Chapter 4 and 5). Toward the purpose, we devise a simulation environment and human decision-making models that simulate a diversity of human behaviors including the unanticipated ones (see in Section 1.1). In this chapter, our major objective is to validate our approaches and contributions in previous chapters in a real-world application with real humans. For that reason, we design a small-scale conveyor belt setup allowing autonomous human-robot collaboration (HRC) with our anticipatory decision-making framework integrated.

It is hard to invoke contingent human behaviors and human errors in constrained lab environments; hence, they are mostly overlooked in HRC studies. Contrarily, human work environments have been extensively studied to examine the conditions of cognitive and physical workload on humans and their existing effects on human behaviors (Hiatt et al., 2017). We take the first step to close this gap between an anticipatory cobot design and its deployment and tests in a more realistic and unconstrained collaboration setup. We design the setup and a shared task that together induces a cognitive load on humans to invoke a larger diversity of human intentions and behaviors in order to show the negative effects of harsh work environments on human workers and on the efficiency of a task. Then, our robot should recognize such situations and naturally coordinate with the human to compensate for possible human errors and to contribute positively to the task.

Physically interacting with humans in real-time in such an unconstrained environment brings about a number of challenges for an autonomous robot. There are a variety of different human characteristics with various task-related collaboration preferences; as a result, recognizing human actions and understanding human behaviors show a great diversity due to their unique reaction times, timing expectations and different ways of reflecting the same intentional and emotional behaviors. For example, one person may idle too long when evaluating a subtask whereas for another this would mean they need

assistance. 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 (Villani et al., 2018), and communication plays a decisive role in this. The robot should then be equipped with intention expressive gestures that effectively communicate a robot's decisions.

In this chapter, we first outline how we face these challenges and describe our setup and experiment design in Section 6.2. Then, we create several hypotheses from our discussions, give the objective and subjective metrics to evaluate these hypotheses, describe the experiment protocols, and discuss the results in Section 6.3. In the first round of experiments, we call it *Experiment-0*, we conduct case studies to test and validate our setup and our task design whether they are able to induce cognitive load on humans (covered in Section 6.3.1). After the validation of the setup, *Experiment-1* validates the findings in Chapter 4. That is, we repeat the experiments conducted in simulation (in Section 4.3) now with real humans to validate our anticipatory partially observable Markov decision process (A-POMDP) design (covered in Section 6.3.2).

The results prove that our design of A-POMDP increases the efficiency of the collaboration and the human's perception of the robot as a collaborating partner by providing a better short-term adaptation, including adaption to unanticipated human behaviors. The experiment results also validate the reliability of the simulated human models we develop (in Section 6.4.2). In *Experiment-2*, we integrate the whole system and confirm through another user study that our anticipatory Bayesian policy selection (ABPS) system, along with A-POMDP model designs, is able to provide an extended long-term adaptation to the different human characteristics (covered in Section 6.3.3). With that, we also prove the applicability, adaptability, and reliability of our final framework in Figure 3.1. Additionally, in this chapter we also detail our simulated evaluation environment that is used in the previous chapters, address how we simulate human behaviors, and validate its reliability using the real human data collected during user studies (in Section 6.4).

6.2 Methodology*

In this section, we first detail our collaboration setup with its multi-agent configuration (Section 6.2.1), the use-case scenario, and our collaboration task inducing cognitive load on humans (Section 6.2.2). Then, we give the design of our robot's cognition that recognizes human actions and the environment and triggers the robot decision-making to synchronize robot responses with human actions (Section 6.2.3). Next, we address the necessary adjustments on the robot decision models, A-POMDP, and the policy selection system, ABPS, to be compatible with the new task definition in the real-world (in Section 6.2.4.1 and 6.2.4.2). Finally, we give the design of the robot motions (e.g., gestures) for the explainability of the robot decisions toward fluent coordination (see in Section 6.2.5).

6.2.1 System Overview and Setup Design

An overview of the system architecture is depicted in Figure 6.1 and the real system setup is shown in Figure 6.2. Two cameras (one RGBD camera: Asus Xtion RGBD camera, one RGB camera: Dericam W1 720p webcam), three load sensors (weight sensors), and one infrared sensor provide sensory information to the system. We use Dobot Arm (with 3 Degree of Freedom (DOF)) and its conveyor belt kit * in our experiments. We use Raspberry PIs both for the robot arm and the conveyor system to interface the rest of the sensors and the PCs. We use a separate PC, the *robot observation PC* for the real-time recognition of the environment and the human actions. In order to reach a fast and real-time response, that is, providing a fast decision every time an action is recognized, we use another PC called *robot decision PC*. This is required due to the low computational power of the robot PC at hand. Such a distributed system is required as the observations should be processed fast and the decision making should not be delayed to avoid asynchronous robot responses to the human actions, one of the biggest issues in human-in-the-loop systems. Our distributed multi-agent system is running on ROS,

*Some content in this section also appears in our previous studies (Kargruber, 2019; Vi, 2019).
*https://www.dobot.cc/products/conveyor-belt-kit-overview.html

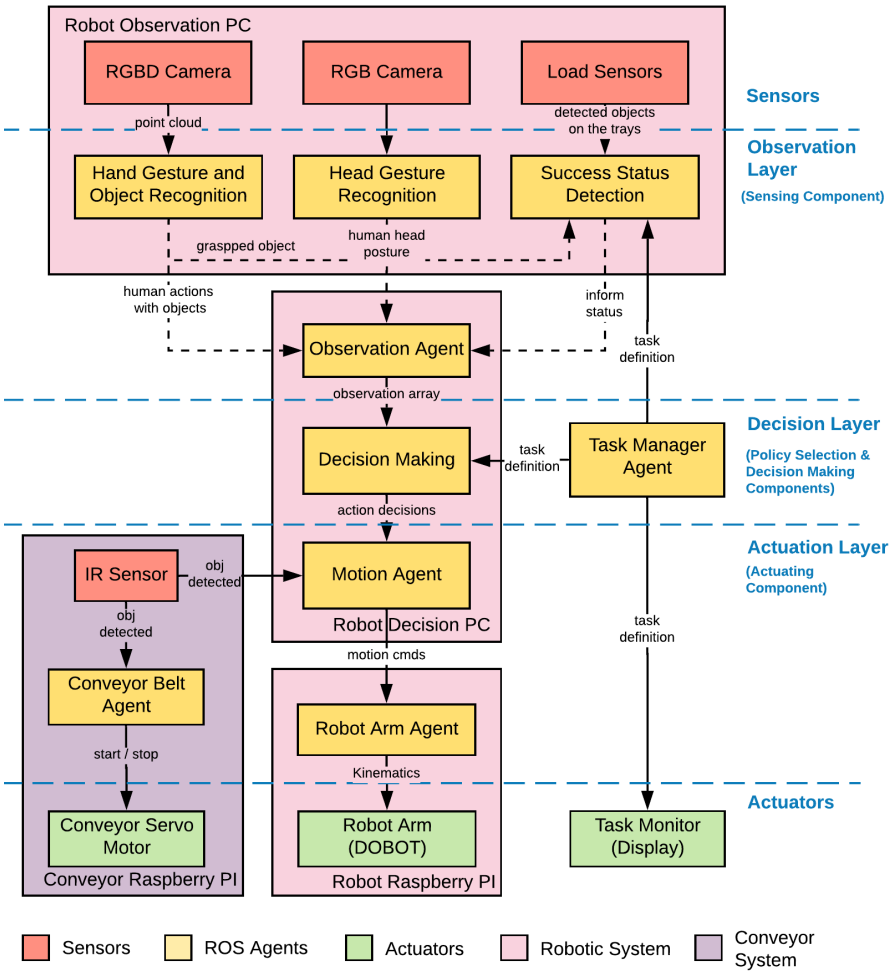


Figure 6.1: Real setup system architecture with the software and hardware components. The layers reflect how the components in our FABRIC framework (in Figure 3.1) is deployed.

which allows us to deploy the agents on different computers like cloud servers. This makes the system scalable to incorporate more observations and multiple HRC nodes in a factory.

There is an additional display to show the current task assigned and its status to a human participant during the task (*task monitor* in Figure 6.2). A task, in our case, is a pick and place job that a human and our robot sort colored cubes onto the containers based on defined rules. A human worker interfaces the system through a yellow glove and a safety helmet as shown in the figure. The *hand gesture and object recognition* and

head gesture recognition nodes then process the camera data to recognize human actions from these wearables and their interactions with objects and supply this information as observations to the robot's *observation agent*. In addition, the load sensors located under the container trays detect the objects the participants put on them. This provides more stable sensing. In the robot's case the sorted objects are tracked from their colors by the RGBD camera, in addition to the available information on which of the trays the robot is heading to. Using the information of a currently grasped object coming from the *object recognition* agent and the container that is placed on, *success status detection* block checks and informs to the *observation agent* whether the human sorted an object to the correct tray or not, stating the success of a task (see Section 6.2.2 for the task details).

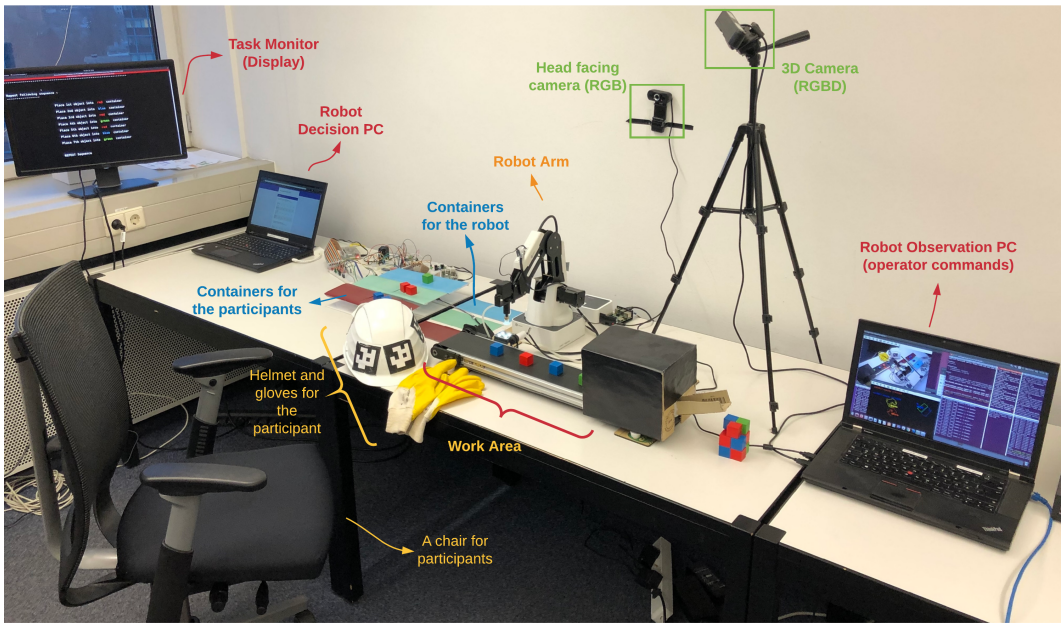


Figure 6.2: Real-world setup: Human-robot collaboration on a conveyor belt.

The *decision making agent* derives the observables from the extracted sensor information (detailed in Section 6.2.3) and calculates an optimal robot action. It is based on our A-POMDP models and ABPS as described in Chapter 4 and 5, respectively. It follows the decision-making process as given in Algorithm 1. A A-POMDP model is selected either by an experimenter (manual) or by the ABPS mechanism (automated),

and it runs until a task reaches to an end. A selected POMDP model is run using our updated version of the DESPOT online POMDP planner, in (Ye et al., 2017), to solve for POMDP models, and execute them in real-time while interfacing the ROS environment through a web socket connection. The generated action decisions by the *decision-making agent* are then sent to the *motion agent*. The abstract action decisions, such as, grasp the current object, are translated by the *motion agent* as the motion commands detailing where the arm should move for the action of grasping. Finally, such motion commands are processed by the *robot arm agent* for the kinematics to execute the action. The *conveyor belt* is coupled with the infrared sensor and runs independently from the rest of the system with a rule to stop it when a cube is detected by the switch. The orchestration of the system is done by the *task manager agent* node, which allows the system users (experimenters in our case) to define and initiate new tasks, informing the current task to the robot and to the display for a participant, resets and initiates most of the other nodes and collects all relevant statistics.

6.2.2 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 as shown in Figure 6.2. The experimenter selects a task of sorting the colored wooden cubes onto the containers based on their colors (with diffirent rules). Each cube sorting is called a subtask and a certain amount of subtasks sorted in a row defines a task. Following the selection, the task rules are displayed on the *task monitor* as shown in Figure 6.1. A task is visible only for a certain amount of time and then it disappears as the participant's job is to memorize it (i.e., a cognitively challenging task). The selected task starts right after the rules disappear. Then, the robot system (see the *robot decision PC* in Figure 6.1) is launched. A decision model is selected either directly by an experimenter (i.e., short-term adaptation only, e.g., a A-POMDP model runs as in Section 6.3.2) or autonomously selected using our ABPS system (i.e., short- and long-term adaptation as in Section 6.3.3). The conveyor belt then starts to move and transports a wooden cube towards the human and the robot. A

cube is available to be grasped and placed once it stops in front of the infrared sensor. The participant and the robot might now decide to grasp the object and place it on one of the containers according to an assigned task.

After each placement, either by a human or by the robot, the status of which cube is placed on which container is displayed live on the *task monitor*. However, it is not displaying if it is a successful placement or not. This is only for the participants to keep track of the task with many cube placements. Once a maximum number of cubes are placed in a task, the robot *decision making agent* and the *task manager agent* terminate. The task results are 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 in an experiment. It is calculated based on the requirements set by an experiment protocol, which is left configurable and informed to the participants in advance of each experiment.

Experiment controller: In an experiment, the following are configurable by the experimenter according to the experiment design and goals:

- Definition of a task and its difficulty: The sorting rules (which colors and cubes to be placed on which containers), the number of cubes to be placed to finish a task, the time the rules are available to a participant on the *task monitor* before they disappear.
- The scoring mechanism: The score as a metric for the success of a task and for the motivation of and the competition between the participants.
- Robot decision-making modes:
  - Short-term adaptation only (Manual Mode): A decision model (e.g., a A-POMDP model) is selected by an experimenter and runs for a task. This mode only activates the cognitive-level of our framework in Figure 3.1. This mode is used in *Experiment-1* in Section 6.3.2.
  - Long- and short-term adaptation (Automated Mode): ABPS runs for autonomous decision model selection. This activates all levels (including

metacognitive-level) in our framework. This mode is used in *Experiment-2* in Section 6.3.3.

Once an experimenter makes the selections above by filling out a configuration, they trigger to initiate a new task as shown in Figure 6.3. In the figure, the process of automated system configuration is shown, which is conducted by the *task manager agent*. The experiment configuration is parsed to obtain task and subtask conditions and rules. Then, the scores are reset, a robot decision-making mode and models (if short-term adaptation is selected) are informed to the robot *decision-making agent* and *robot arm agent* and *conveyor belt agent* are reset to reconfigure the hardware. This entire process, until a task terminates, is tracked autonomously by the *task manager agent* as given in Figure 6.3. The robot decision models terminate automatically when a global success (a task termination) is issued (see their design in Section 6.2.4.1).

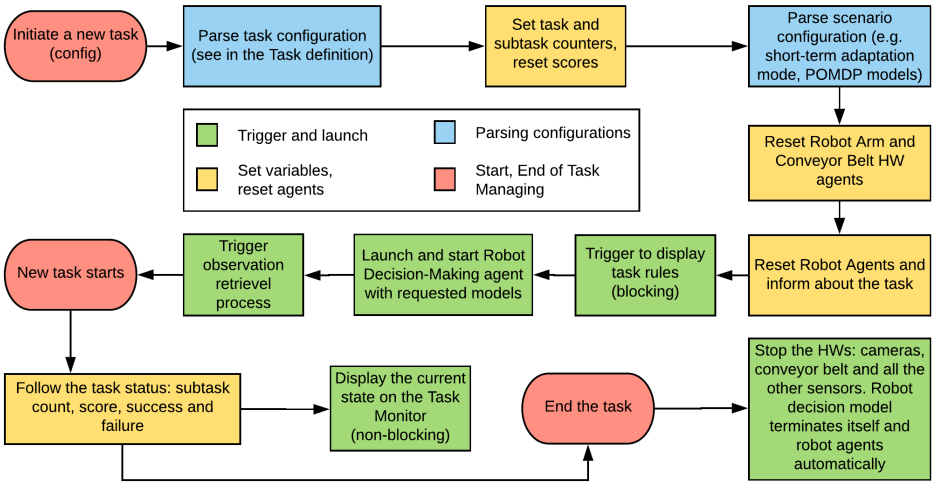


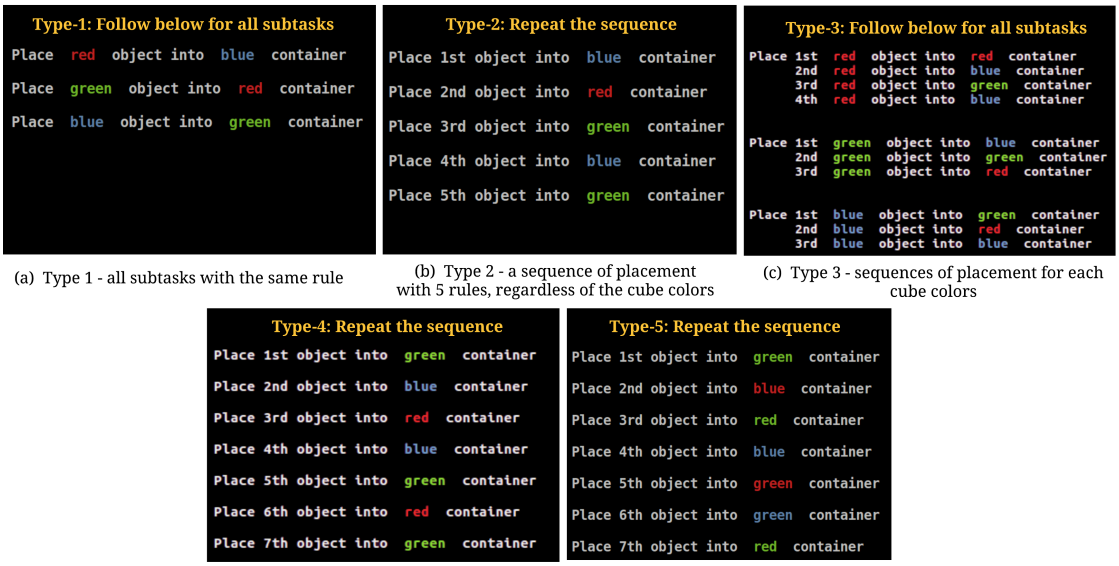
Figure 6.3: Task initialization and progress flow coordinated autonomously by the *task manager*

Cognitively challenging task design: The current experiment setups for HRC often require a human and a robot executing a task in a turn-taking manner with predefined roles ((Chao and Thomaz, 2012; Hoffman, 2019b)). A task in our experiments allows for a fluent collaboration, where a human and a robot should take the initiative to change and

adapt to task allocations on-the-fly by observing each other and the environment towards maximum efficiency and safety. For example, when a human fails with a task assigned to her due to an internal or an external reason, a robot should anticipate and detect the situation to offer its assistance. Our robot's main task is to correctly anticipate such cases for high-level coordination. Hence, we create an environment that provides harsh working conditions to invoke human errors or failures, i.e., a realistic and unconstrained work environment.

In Chapter 4 and Chapter 5, we model a task to be physically demanding in simulation, and a robot would take over when a human's physical abilities are exceeded. In user studies with real people, such a setup would be hard to prepare in a lab environment as we need a robot that can carry a weight a human can't. In addition, physical stress over the participants may cause physical damage if they push themselves hard to lift the weights. For this purpose, we still focus on a task of pick and place but for simplicity we create cognitive load on 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 incapability can be resulted also from a difficult cognitive performance, which is an easier and safer option in a lab environment.

We categorize tasks with their difficulty levels (i.e., cognitive load). More placement rules to remember would yield to a more difficult task since they are only displayed for a short amount of time at the beginning of a task and vanish afterward. Therefore, a task is cognitively demanding with varying extents according to the complexity of the rules and the amount of time they are displayed. We run case studies in order to evaluate and compare the cognitive loads these tasks induce on people (see in Section 6.3.1). Our goal is to compare and select tasks that are both challenging but with a difficulty still achievable by the people on average when they receive no help from a robot. In total 5 different task types have been implemented. Each type is inspired by mind games to sort colors with confusing rules (e.g. using Stroop effect (Stroop, 1935)). The types we have created and tested in our experiments are given in Figure 6.4.



(a) Type 1 - all subtasks with the same rule

(b) Type 2 - a sequence of placement with 5 rules, regardless of the cube colors

(c) Type 3 - sequences of placement for each cube colors

(d) Type 4 - a sequence of placement with 7 rules, regardless of the cube colors

(e) Type 5 - Type 4 with color mismatch in the text and the words, i.e. Stroop effect

Figure 6.4: Displaying task rules for 5 different task types. The cubes flow on the conveyor with a random order unknown to the participants.

In addition to the color rules and certain display time to memorize them all, participants face another challenge during the execution of a task. After each cube placement, as we call subtasks, the participants are informed to wait for audible feedback, i.e., a beep sound, before they move to the next cube (the robot also follows the same rule). This corresponds to a supervisor check of a placement. The *task manager* triggers the sound if it receives placement feedback (e.g., the load sensors under each of the container trays). Since the load sensors are slow in processing such light cubes, it takes a varying amount of time to output a detection with precision over 95%. If a rare occasion of failed detection occurs, the participant needs to repeat this subtask. This is expected to put an extra cognitive load on the participants due to a dynamic delay between each placement.

Motivating the participants to collaborate: The collaboration between a human and a robot is motivated by a scoring system that punishes wrong placements with negative rewards and assigns positive rewards for any correctly sorted objects, whether it is by a participant or by the robot. The same rewarding mechanism is also used in the robot decision models, i.e., POMDP (see in Section 6.2.4.1). Hence, it is a shared goal of the

team to maximize the collaborative score. 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 robot, especially when they are struggling with a task. Since the color of the cubes can be recognized wrong due to the changing lighting conditions or other classification errors, the robot may also make a mistake. This is informed to the participants without the success rates of the robot, which is above 95%. Whether to trust the robot or not depends on a participant, which usually builds up in humans in time after observing the robot behaviors. 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, they may trust the robot more, which we categorize as the long-term changing human characteristics.

One of our robot's goal is to adapt to a human's changing willingness to collaborate (i.e., collaborativeness as we call). Therefore, in addition to the changing trust in the robot, we induce competitive behaviors for a dynamic collaborativeness in the participants. For that, we assign all of the tasks to them, and the robot is there to assist if needed. Our intention is to avoid situations where a participant lets the robot permanently take over a task, which is against the idea of collaboration. To motivate them even further, we also assign bigger scores for a successful placement by a participant compared to the same placement achieved by the robot. Hence, if a human is not sure about a subtask, it is a strategy to let the robot take it over in a trusted collaboration, avoiding a negative reward by sacrificing a bigger reward.

All of this is informed to the participants before the experiments with an analogy from a real factory that an assigned task is better achieved by the assignee for increased quality. Since in our setup a robot can indeed perfectly achieve a task, to describe this analogy to the people we give an example that the objects to be placed are fragile and so a human would handle the placement better than a robot. We expect this to encourage the participants to take over as many of the placement tasks as possible in order to maximize the collaborative reward.

How to induce the unanticipated human behaviours?: The goal of this environment setup and the scenario is to invoke the unanticipated human conditions in a work environment like getting tired, lost motivation, failures, not willing to collaborate, etc. that are expected to be observed during monotonous work conditions. In Table 6.1, we summarize our methodology to induce such conditions (referring to Figure 6.2):

Table 6.1: Our experiment design choices to invoke dynamic human behaviors.

Human Behaviors	**Invoked through**
Mistakes (Failures)	Cognitive load: Cognitively demanding task rules that are hard to remember and only visible to a participant for seconds. Placement rules: Audible feedback is given to confirm if the placement is recognized. In the case of a no sound, even if the placement is correct, the subtask must be repeated.
Distraction	The *task monitor* is placed to the left-hand side of a participant so that when the human is checking the current status of a task, she is distracted and not paying attention to the continuously conveyed cubes.
Tiredness	The cognitive load accumulates through multiple tasks, where an experiment takes approximately 1,5 hours in total along with questionnaires.
Initial motivation to work	The scoring system: A competition is devised among the participants over their average scores. Negative rewards are assigned for misplacements and positive rewards for the correct ones. Training phase: We train participants and practice some tasks before the experiments to get familiar with the robot and the task for initial motivation.

Continued on next page

Table 6.1 – continued from previous page

Human Behaviors	Invoked through
Lost motivation	The same task type is repeated several times, which is expected to cause a participant to lose interest. The tasks themselves are also demotivating as they require focused attention and strong memory. The length of an experiment is also expected to cause a demotivating effect.
Will to collaborate with the robot	Robot's influence: Once a participant realizes that the robot works with high precision, they are expected to trust the robot. For example, when they don't remember the rules or lost the motivation, they may leave the task to the robot. This is also motivated through the punishments for misplacements.
Not wanting the robot to assist (non-collaborative behaviors)	The fact that a participant receives more rewards, due to the initial assignment, than the robot doing a subtask may play down the teamwork and increase competitive behaviors in humans. We observe this behavior through the warning gesture from the participants stop the robot from taking a cube. Misjudgments of the robot about a participant's need for help are also expected to invoke less collaborative behaviors in the participants, e.g., a human may take some time to remember a rule as there is no time pressure in a task, while the robot thinks they need help. Finally, we also inform the participants that the robot may also make mistakes.

The actual effect of the experiments depends on a participant's characteristics (e.g., how fast they get tired), the time of the day the experiment is conducted, and a participant's background (e.g., from computer science, experience with robots). We examine all these effects in a case study in Section 6.3.1.

6.2.3 Robot Cognition

To allow for real-time interaction between humans and robots, a robot needs to constantly perceive its environment and human actions toward fluent coordination. For this purpose, we gather visual information about the scene using 2 cameras. There is an Asus Xtion camera (an RGBD camera) for object and human action tracking subsystem. The other one is a webcam that supplies RGB video for the head gesture tracking system. In addition, load sensors are used to detect the objects placed on the container trays and a proximity sensor detects when a cube stops in front of the robot arm to pick (see all the sensors in Figure 6.2). All the data gathered from the sensors are used in *hand gesture and object recognition*, *head gesture recognition* and *success status detection* agents shown in Figure 6.1. Then, the processed information is forwarded to the observation agent for the synchronization of the heterogeneous sensor data, which are then forwarded to the robot decision-making block as an observation vector with semantic descriptions.

6.2.3.1 System Update Frequency

In scenarios where humans and robots work together, the timing of robot decision making is crucial for a timely response in a fluent collaboration. The stochastic nature of human behaviors makes it difficult for a robot to follow decision-making with a constant frequency. Poor coordination is likely to result in a collaboration that is neither effective nor perceived as natural. For this reason, the robot should balance timely decision-making with natural interaction speeds, catching human actions and the important changes in the environment to react reliably. For a fluent coordination, our robot *observation agent* follows a decision trigger logic as represented in Figure 6.5. In our setup, a robot 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 a no change in human actions or a long-time absence of an observation update for any reason. The *observation layer* in Figure 6.1 continuously runs to observe and recognize the environment. The *observation agent* synchronizes the sensory inputs to create an observation vector (see Figure 6.6), analyzes them in case a

decision should be triggered, and forwards them to the *decision-making agent* in case of a trigger. As shown in Figure 6.5, both a subtask update and a human action recognition can happen at any time (discontinuous) and may trigger a decision making including interruptive cases like a human warning the robot.

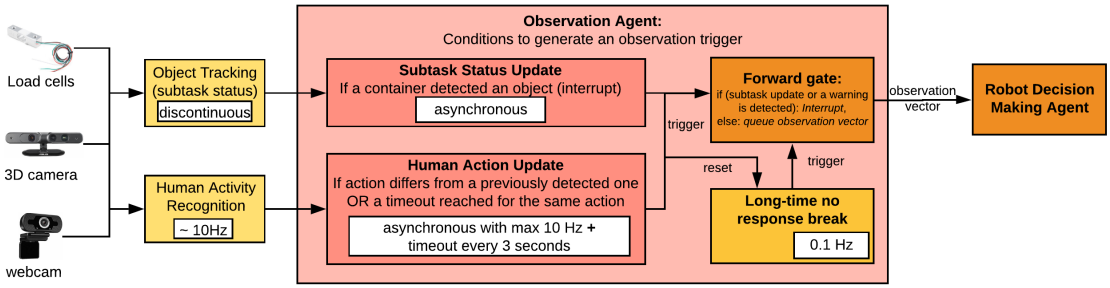


Figure 6.5: Decision trigger logic of our robot (handled by the *observation agent* in Figure 6.1)

The robot makes an action decision by subsequently searching a generated POMDP policy for optimal action. For this purpose, we run an online POMDP planner (see Section 6.2.4.1 for details) that significantly enhances the real-time performance. Even so, a decision generation, which follows the Algorithm 1, is still the limiting factor for the update frequency of the robot. The experiments have shown that generating an optimal action decision can take up to around 1 second with the computing power of the PC running it. This limits the decision making frequency to a maximum of roughly $1Hz$, whereas we are able to detect human observations with a frequency of $10Hz$. For that, we set some filters for the observations or queue them not to overload and crash the *decision-making agent* with many requests (see the trigger conditions and the forward gate in Figure 6.5). For more reliable decisions, we catch and respond to changing 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 case studies in Section 6.3.1.3, 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. Therefore, an update frequency of 10Hz would

be unnecessarily fast for the robot decision-making and would even be unreliable as the robot may think a human is taking the same action repeatedly even if it is still one action in progress. However, it is necessary to catch every new human action for a fast response whether they are queued or interrupting a new decision.

As mentioned, there are some interrupt events that trigger a decision update. As shown in Figure 6.5, when a cube is detected on a container, which is a discontinuous event that may occur at any time, an interrupt is generated that triggers immediate decision-making even if there is already a running process. Another interrupt event is when a warning action is detected from human activity recognition. According to a running robot decision model, i.e., an A-POMDP model, a robot may stochastically decide to cancel its actions when a warning is received from the human (i.e., obeying the warning signal). In such a case, it interrupts an ongoing action by stopping the robot arm and executing the cancellation protocol (see in Section 6.2.5). Thanks to our distributed multi-agent architecture, a decision-making process runs in parallel to an ongoing robot action being executed. Unless the new action decision is to cancel an ongoing one, all of the new decisions are queued as the next action in the robot's *motion agent* (as shown in Figure 6.1). Other than the interrupt conditions, when a decision-making process is already running and a new trigger arrives, the observation vector is queued to be forwarded after the currently running process ends. In all of these cases, a robot decision-making process always receives the same observation vector with the up-to-date information collected from subtask status (i.e., ongoing, failed, or succeeded) and human action update.

6.2.3.2 Observation Vector

The observation vector is the feature vector triggering the robot decision-making. The collection of observables used by the robot in this chapter is an extension of the one used in Chapter 4 and Chapter 5 (see the list of observables in Figure 4.2). In real user studies, we change the definition of a task. Instead of one heavy object placement, we create a task consisting of several small object placements to create a cognitive challenge.

Therefore, each placement is now a subtask, which introduces two new observables to the previously used vector. The new observation vector is shown in Figure 6.6.

Obs #	Observation
O1	Task success
O2	Task failure
O3	Is human detected?
O4	Is human looking around?
O5	Is human attempting to grasp?
O6	Is human warning the robot?
O7	Is human idle (inactive)?
O8	Has human succeeded in a subtask?
O9	Has human failed in a subtask?

Figure 6.6: Observation (feature) vector used to update robot decision-making system in the user studies

Among the observables, there are direct observations received from the *hand gesture, head gesture and object recognition* and *success status detection* agents in Figure 6.1. These observables, which are all binary, are shown in Table 6.2 along with their descriptions and how they are detected. A human is detected ($O3$) when a hand glove is visible for the HAR system and the human actions of attempting to grasp ($O5$), warning the robot to stop ($O6$) and staying idle ($O7$) are classified by the hand gesture recognition algorithm (detailed in Section 6.2.3.3). $O4$, looking around action, is detected by the head motion tracking through a marker on the safety helmet participants wear (in Section 6.2.3.4).

On the other hand, there are some observables that are derived from the direct observables received (shown in Table 6.3). According to the assigned task rules, we check the status of all of the containers (Oc) and the object of interest (Ob) to output a subtask success or failure ($O8$ and $O9$). With that, the robot tracks how the human is doing with the task (see in Section 6.2.4.1). Finally, task success and failure are calculated by counting the amount of successful and failed subtasks. When the total amount is reached to a value defined by a task definition (mostly 10 in our experiments), either $O1$ or $O2$ becomes true so that a robot decision-making model (i.e., A-POMDP) terminates. In

Table 6.2: Direct observations

ID	Observation	Description	Means of detection
O3	Is human detected?	yellow glove detected by RGBD camera	color detection and tracking by RGBD camera
O4	Is human looking around?	side marker of helmet is detected	head (helmet) tracking by 2D camera
O5	Is human attempting to grasp?	human grasps a detected object and/or moves it	GMMHMM gesture classification
O6	Is human warning the robot?	A warning gesture by human is detected	GMMHMM gesture classification
O7	Is human idle (inactive)?	Human is idle	GMMHMM gesture classification
Oc	container status	measuring load (participant containers) and detecting colors (robot containers)	load sensors and detection by RGBD camera
Ob	detected objects	detected cubes by color (green,red and blue)	color detection and tracking by RGBD camera

our experiments, the task status flags are just for the robot models to terminate. They are not used to measure the performance, which is calculated using the subtask results.

Table 6.3: Derived observations

ID	Observation	Description	Derived from
O1	task success	counting succeeded subtasks and comparing with a threshold	successful subtasks (O8)
O2	task failure	counting failed subtasks and comparing with a threshold	failed subtasks (O9)
O8	has human succeeded in a subtask?	human grasped and placed an object into a correct container	human grasp recognition (O5) and a container status (Oc) compared with the task rule
O9	has human failed in a subtask?	human grasped and placed an object into a wrong container	human grasp recognition (O5) and a container status (Oc) compared with the task rule

6.2.3.3 Human Activity Recognition: Hand Gesture

This section covers the recognition of human actions that are depicted as $O5$, $O6$ and $O7$ in Table 6.2. A state-of-the-art human activity recognition (HAR) module, inspired by existing studies, e.g., (Roitberg et al., 2014), has been implemented to recognize the constrained and distinct human gestures from the movements of the worn hand glove (see Figure 6.7). As we need to semantically describe the grasping action, i.e., the human has grasped a blue cube, we also develop an object detection algorithm. It detects the contours of objects using the border following algorithm (Suzuki, 1985) in OpenCV. First, a color filter is used to remove any objects but the colored containers, the colored cubes, and the yellow glove. The algorithm then finds the border contours. The center points of each contour are then estimated using a Kalman filter for smooth tracking of the objects. For the containers, they are also spotted initially using a color detection and their known distance to the camera, which remains constant throughout the operation. Using the distance of a tracked object and its location with respect to the camera, we detect if the object is put over a container with some distance comparisons. With the load sensors deployed under the containers, the placement detection works accurately.

Moving from the tracked location of the hand glove and spatial relation between the tracked objects and the hand, human actions of *idling (inactive), attempting to grasp* and *warning sign* are recognized by a GMMHMM classifier (e.g., in (Li, Lei, and Zhang, 2018)). An additional class of "undefined actions" is also added with some random irrelevant gestures in order to avoid false positives on the three important gestures. The feature vector for the training and classification consists of *a)* the distance between the hand and the object of interest, *b)* the hand velocity and *c)* the distance between the RGBD camera and the hand. The classification algorithm consists of a Gaussian Mixture Model (Gaussian Mixture Model (GMM)) and a Hidden Markov Model (HMM). A GMM operation is needed to create discrete clusters of the components in the feature vector, which are all continuous variables. The discrete values are then put into the HMM for the classification. We create a separate GMMHMM for every action of interest. Then, every time a feature vector arrives, all models generate a confidence value, and the one

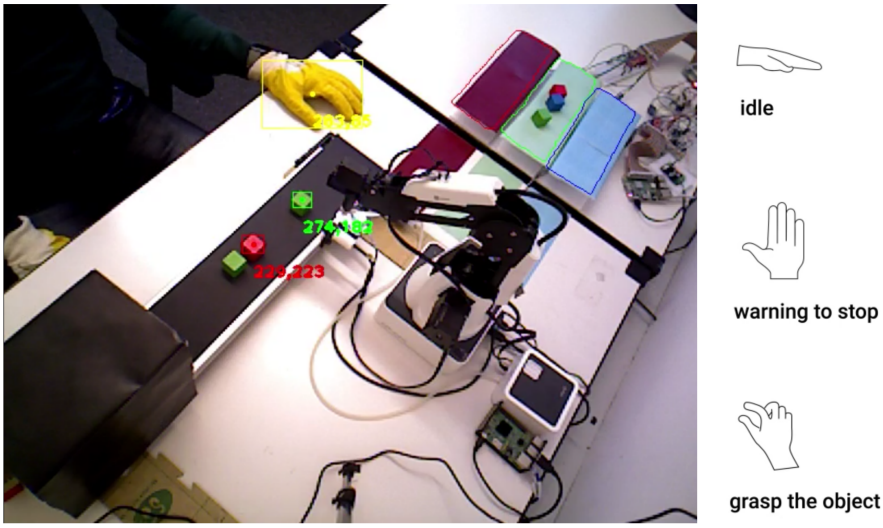


Figure 6.7: View from the RGBD camera: Objects and the glove are detected in the work space even when the objects are grasped.

with the highest confidence is taken as the current most probable action. For the training and testing we have invited five Colleagues and students to record a dataset with these actions performed. Such diversity of a dataset has resulted in an accurate trained model with over 90% accuracy running with $10Hz(fps)$ on a continuous video stream (online data), which is more than enough for our experiments.

6.2.3.4 Human Activity Recognition: Head Gesture

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 (observation $O4$). For this purpose, the ArUco markers are used on the safety helmet (see Figure 6.8). The black border of the markers allows for a robust and easy detection within an image. Additionally, besides its robust detection, by continually minimizing the reprojection error of the four corners of the marker, the pose within the image can be estimated ((Garrido-Jurado et al., 2014)). According to the projection on the work area and several tests with the people, we define margins to interpret if a human is looking around. The accuracy of this observable is very high; however, a misdetection is compensated through the need for sequentially observed

multiple looking around gestures in order to estimate that a human has lost her attention in an A-POMDP.

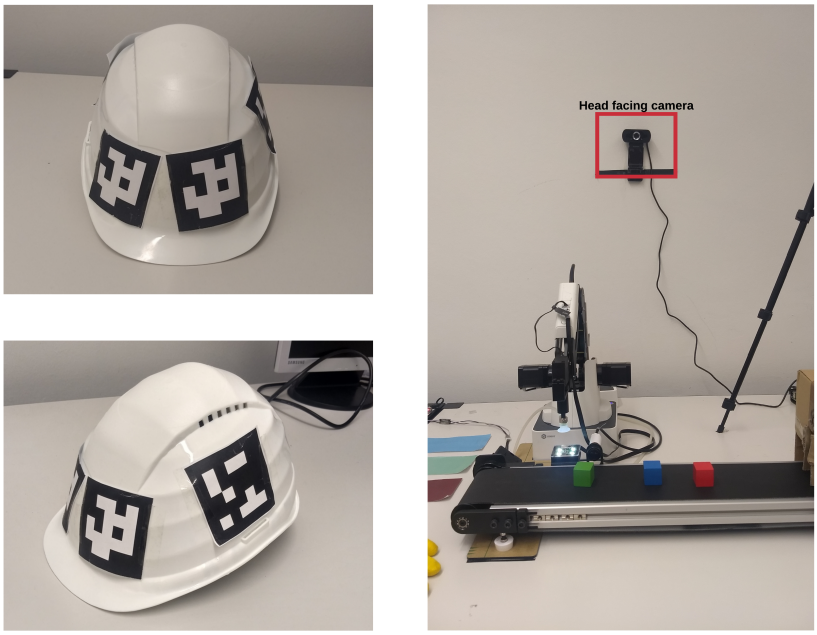


Figure 6.8: Helmet, markers and head facing camera for the head posture detection.

6.2.4 Robot Decision-Making for the User Studies

6.2.4.1 Adjustments to A-POMDP Models for the User Studies

This section summarizes our necessary adjustments on the robot decision-making to comply with the new task design and the new environment. In addition, we describe how we improve the decision-making components, i.e., A-POMDPs and ABPS, using the human data collected through the early case studies.

Updates on POMDP state transition and observation functions

The new collaboration task for the user studies requires multiple placement jobs as opposed to the previous task of one package placement in simulation (see in Section 6.2.2). This means a longer interaction time during one task and multiple succeeded or failed

placements, leading to more observations for the robot to assess and interpret a human's state. For this purpose, we add a new state and two new observations in our A-POMDP model, and we update the transition probabilities to favor a slower transition (i.e., a slower judgment) from one state to another as of now the tasks span a longer time (see our new A-POMDP model in Figure 6.9 and compare it with the version in Figure 4.2). The new state is a partially observable state and it is called *human doing ok*. It indicates that when the human is successfully achieving the subtasks (in our case the placed cubes), the robot will most likely end up in this belief state until the human repeatedly executes the unanticipated behaviors like making mistakes or acting slow. In such a case, the robot may transit to the *anticipation stage-1*. The new model also favors a transition from this stage back to the *human doing ok* state if the human keeps up with the task. Finally, the newly added observables reflect the success status of a subtask, as they are detailed in Section 6.2.3.2.

The new additions require updates on both the observation and the transition matrices of a POMDP. At first, the probabilities from the old model are redistributed to the new matrices with the goal to maintain a behavior of the robot closer to the previous one, e.g., cancel the action if a human warns. This process is also qualitatively tested with the simulation-in-the-loop, where we use our simulated human models. The probabilities are tweaked until the new robot policy approximates to the one generated by the old model design. In the second step, the probabilities are tweaked this time to adapt to the new timing of a task. Due to the multiple subtasks in and a longer duration of a task, the robot needs to avoid quick judgments as they may lead to unreliable responses. For example, sequentially recognized idling actions between two cube placements may be interpreted as "the human is tired" if the robot transits to that state quickly, whereas in reality the human may be waiting for a new cube to arrive on the conveyor. As a result, we favor more the probability of a state to transit back to itself rather than another one. However, we note that there cannot be a single probability (or a judgment speed) defined for a robot in its interactions with multiple people. This is covered in our robot's

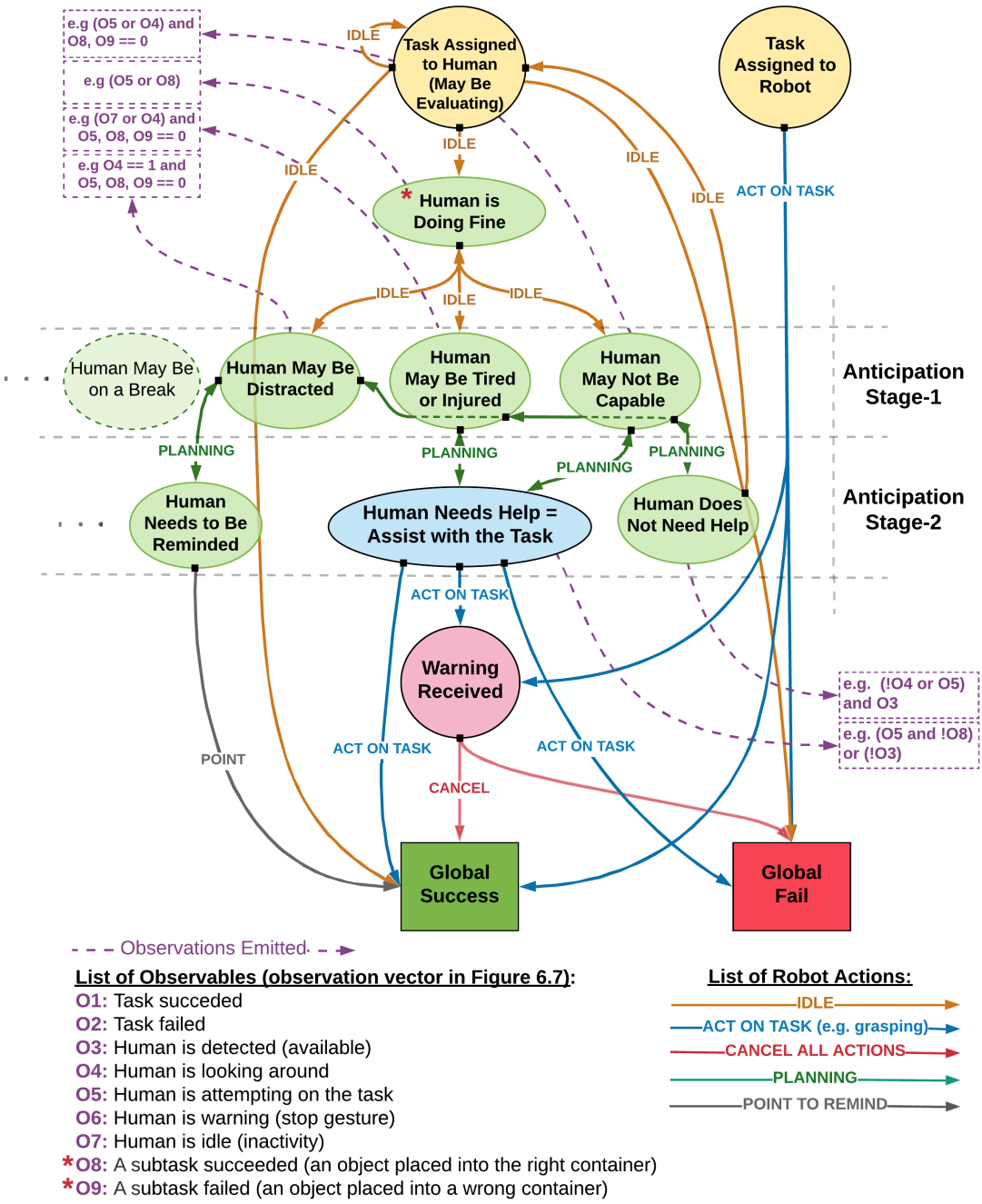


Figure 6.9: Anticipatory robot decision-making model A-POMDP) adapted from the one in Figure 4.2 for the real user studies and the collaboration environment. * marks the new additions; a new state added to track if human is doing fine with a task and the new observables, i.e., $O8$ and $O9$, are introduced to inform about the current subtask status.

long-term adaptation capability through ABPS. These adjustments are also made with simulation-in-the-loop.

Finally, we collect real human observations from the case studies and use this data to improve the intrinsic parameters of the model, i.e., the observation and transition probabilities. Since the experiments involve collaborations with multiple people, we obtain an average model, as we call the base model, that is used to generate many other feasible models (as needed by the ABPS policy library). Our findings from the early real human interactions and the update process are given under Section 6.3.1.3.

Rewarding mechanism We set a cooperative reward for the human-robot team to evaluate the performance of the robot during a task. Whenever an observation is received by the system, an immediate cooperative reward is assigned. As depicted in Table 6.4, there are rewards only for the observations of "a warning received", "a subtask success" and "a subtask failure"". The distribution of the immediate rewards favors correctly placing an object and sharply punishes wrong takeover decisions by the robot when it leads to a warning by the human collaborator. It should be noted that the same rewards are assigned for a subtask success or failure whether the human or the robot completes it. We define a discount factor to avoid longer waiting times before a placement takes place. The robot would want to take over if a human is stalling too long. The discounted reward function given below is the final value a robot model collects after a subtask.

$$R_T = \sum_{t=0}^{T} \gamma^t r_t \tag{6.1}$$

where R_T denotes the total discounted reward in a subtask, r_t is an immediate reward between time $t = 0$ and $t = T$ discounted by γ^t. Also, t is the time at which a new observation is received. It is reset to 0 at the end of each subtask (i.e., an object has been detected on a container). Finally, the total discounted reward in a task, given below in Equation 6.2, is the sum of all subtask rewards and it is a direct indicator of the performance of a robot model during an episode of a collaboration. As in the previous

chapters, we use our updated version of DESPOT solver in (Ye et al., 2017) to solve for and generate a policy from our A-POMDP models.

$$R_{task} = \sum_{subtask=1}^{n+1} (\sum_{t=0}^{T} \gamma^t r_t) \tag{6.2}$$

Table 6.4: Immediate rewards assigned during the experiment

Observation	**Immediate reward**
warning received	-3
sub task success	+6
sub task failure	-6
all others	0

6.2.4.2 Adjustments to the ABPS Mechanism

The updates on ABPS consist of the training of the observation and the performance models (recall in Section 5.2.2). We update the simulation environment according to the new task definition (see in Section 6.4) and run multiple simulations for the collaboration of several human types (run by the simulated human models) and the newly created policies in the policy library (using the new modeling scheme in Section 6.2.4.1). We repeat the same training experiments as in Section 5.3.1.2 with the 16 distinct human types configured using the modeling scheme in Figure 6.28 and the 20 best robot policies in the library in hindsight. The procedure of the creation and the selection of the best 20 policies are as in Section 5.2.2.1. The collaboration of each of the 20 robot policy and each of the 16 human types (see the types created and used in Section 5.3.1.2) is repeated for 90 sequential tasks. In total, we accomplished 28800 interactions (28800 task instances and 288000 subtask instances), which is very difficult to manage in real-life scenarios.

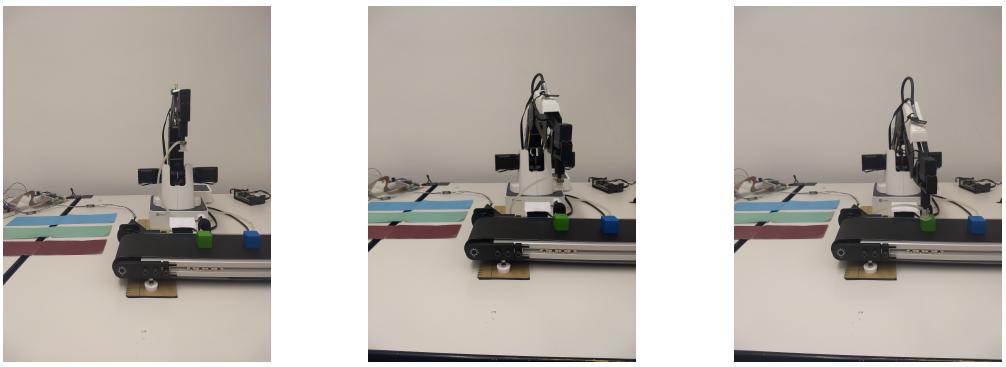
After the training in simulation, we update the models also using the real human data collected through the case studies in Section 6.3.1. Our update in this regard is on the observation model, i.e., a Bayesian estimation model. We do not update the performance model as it requires multiple collaborations of every participant with each of the robot decision-making models in the robot policy library, which would result in

days of experiments with only one participant. Before, we have used the generated observations from our simulated human models to cluster different human types based on their varying levels of expertise, stamina, attention, and collaborativeness. Now, we add the real human data labeled based on a participant's objective success rate, average idle time, and the number of warnings to the robot. The new dataset, consisting of both the reliable data and the simulation with much more variety, is used to retrain the observation model that is evaluated under Section 6.3.3.

6.2.5 Robot Motion Design for a Natural Collaboration

For a fluent and effective HRC, good communication between the team members is essential (Holladay, Dragan, and Srinivasa, 2014). The agents should effectively convey their intentions and also understand their partner's (Dragan, Lee, and Srinivasa, 2013). In industrial environments, verbal communication is often not possible due to the high noise levels; hence, the coordination needs to rely on non-verbal communication (Gleeson et al., 2013). For the sake of the reliability and validity of our experiments, we implement robot gestures that are understandable to the human participants and that the participants are able to anticipate the robot's intentions. Our goal is to design such gestures that effectively express the robot's internal states while keeping the motions simple, efficient, and nonintrusive to the human to execute on our robot arm with only 4 DOF. The robot actions are generated from the robot's decision-making models (in Figure 6.9) and they are idling, grasping a package, canceling the current action, planning for grasping and pointing to remind the human about the task. In our scenario, the human and the robot need to collaborate in close proximity but do not have to coordinate their actions to move simultaneously. However, the actions should be smooth and fast enough to catch up with changing human decisions and plans for fluent coordination. Following the findings in (Dragan et al., 2015), predictable and legible motions would be suitable in our case. Below we briefly describe our motion designs for each of the actions mentioned.

Point to remind human about the task: As stated in (Imai, Ono, and Ishiguro, 2003), eye contact is very effective for initiating communication between a human and a robot. It makes a human notice that the robot is going to express something. Since our robot is only an arm, we exert a similar behavior by rearing up and facing the human in order to draw the human's attention towards the robot (see Figure 6.10a). We argue that this design is effective as an attention taking gesture, which is an available option for industrial settings with robot arms. Afterward, the robot moves toward the object up and down as a pointing gesture. We expect that such a motion would redirect the human's attention toward the object of interest. As this motion can be designed in many ways, we come up with two different motions that are evaluated by the humans during our case study (see in Section 6.3.1.1). These motions differ in the heights of rearing up and the speed and amount of up and down motions.



(a) Standing up to take the attention

(b) Moving to the planning location

(c) Grasping the package before the placement

Figure 6.10: Snapshots from some of the robot actions.

Planning for a grasping trajectory: To be able to notify the human that the robot is now planning and soon it may grasp the package, we design two different positionings for the robot closer to the pickup point. The first version, we call it aggressive planning, interferes directly with the conveyor area and stops right above the package. This gesture makes it clear that the robot will go down soon and grasp it. In the second version the robot does not interfere with the work area, but rather aligns itself horizontally with

Table 6.5: Prioritization of robot actions.

Robot action	**Priority**	**Preemptable by**	**can preempt**
cancel	1	not preemptable	all other actions
grasping	2	only preemptable by cancel	all lower priority actions
planning	2	cancel	no preemption, blocking priority 3 actions
pointing	3	cancel, since it is a short motion	no preemption, blocking priority 2 and 3 actions
idle	3	cancel	no preemption

the package outside of the pickup location (see Figure 6.10b). These versions are also evaluated by the people during the case studies.

The other actions, i.e., grasping and idling, are more intuitive and easy to express to humans in an understandable way. In grasping, the robot plans a trajectory that always has the same pickup location (in front of the IR sensor as in Figure 6.10c) whereas the drop location changes according to the task rules. Our only concern in designing this motion is a possible collision with a human trajectory. For the idling, the robot goes to its home position, which is outside of the conveyor area. Our last action, i.e., canceling a robot action, is intended to express to the human that the robot received the human's warning and canceled its current operation. When this action is executed, the robot stops its motion immediately. If it is holding a package already, it drops the package back on the conveyor so that the human can pick it up. All in all, we introduce these actions to the humans in the training phase of *Experiment-1* and *Experiment-2*, whereas in *Experiment-0* we do not inform the humans not to bias them in their evaluations of the robot actions. Finally, in Table 6.5 we give the prioritization and queuing of all these actions in the robot's motion planner. It is the motion planner's job to prioritize the continuously generated actions from our decision-making components and make sure that a fluent robot motion is achieved no matter how frequent and in which order the action decisions are received. In our case, the canceling action has the highest priority as it terminates a robot's action once requested by a human.

6.3 Evaluation Through User Studies

There are three major experiments conducted under this section. *Experiment-0* is the base experiment that involves an initial case study with human participants to validate and fine-tune our novel interaction setup and the robot system for the following experiments (see in Section 6.3.1). This experiment acts as prototyping tests on the setup to check and analyze if our setup and the robot's collaboration is reliable, and how/if we can induce any cognitive load on the participants. In *Experiment-1*, our goal is to validate our findings in Chapter 4, which is the short-term adaptation capability of our robot handling unexpected human conditions. We repeat the same experiments conducted on simulation that compares our A-POMDP robot decision model design with a conventional reactive model (see in Section 6.3.2). Additionally, we also compare our simulated human behaviors and the real ones for validating the reliability of the human models. However, this is analyzed under the dedicated section for the simulation environment (in Section 6.4). Finally, in *Experiment-2*, we integrate full anticipatory decision-making system on the setup to validate the applicability of, first, our ABPS mechanism from Chapter 5, then our complete framework (in Section 6.3.3). All of the experiments use a within-subject design and we invite different people to each of the experiments.

6.3.1 Experiment-0: Evaluating the Setup and Real-Time Interaction*

We conduct a case study to validate whether our setup can induce dynamic human behaviors, such as, getting tired, making errors, losing attention, and motivation. Also, during this experiment we use the feedback from the participants to fine-tune our human-in-the-loop system. This feedback is both towards the environment and the task design, i.e. whether we could create a realistic, challenging, and still motivating environment for the people, and toward real-time interaction including the collaboration frequency, human observations, and the robot's reliability in its responses. In this experiment, we analyze the following:

*The content of this section mostly appears in our previous study (Kargruber, 2019).

- Robot expressiveness: robot motion design to better explain a robot's decision to human
- The degree of cognitive load on humans for different task types
- Collected human observations to fine tune the robot decision-models and the observation update frequency

6.3.1.1 Evaluating Robot Expressiveness

Our goal in this experiment is to support the hypothesis below:

Hypothesis 3. *Our robot effectively communicates through its gestures (non-verbally), expressing its intentions (internal states and decisions) so that they can be correctly understood by humans.*

In evaluating this hypothesis, this experiment investigates the effect of different robot motion types on the expressiveness and understandability of robot gestures by humans. The inspected robot actions are idling, grasping, canceling all actions, planning for a grasping trajectory, and pointing to remind (see in Section 6.2.5). Especially the *planning* and *pointing to remind* actions are needed to effectively communicate the robot's intention as the planning is a precursor for the human to be aware that the robot may take over the task soon and the pointing action is to remind the human about the task. Since they are hard to interpret due to the robot arm's limited motion capability, these actions have been implemented in two different versions. The other gestures are only implemented in one way and the participants evaluate all of the actions.

Experiment protocol: We have invited 12 people for this experiment. First, we instruct them about the collaboration setup, the task, and the purpose of this experiment. Without any further instructions, the participants are shown 7 gestures, one at a time, in the following order: 1) Grasping: The robot is grasping an object and placing it; 2) Planning-1: The robot is planning for grasping by idling in a position right above the object; 3) Cancelling: The robot has an object, the participant warns the robot, it stops

while moving and puts the object back on the conveyor; 4) Pointing-1: The robot rears up to a middle-high position and goes down again and repeats it twice to take the attention of the person; 5) Planning-2: The robot moves a little towards the conveyor and idles there for a while; 6) Idle: The robot remains at its home position; 7) Pointing-2: The robot rears up to the highest position and goes down towards the cube, repeating this twice. After each gesture, the participants are asked to pick one of the following statements that best explains the intention behind the gesture they just saw.

The robot ...

- took my attention on itself
- reminded me about my task
- took over the task
- canceled its action
- planned for a grasping action
- was idle/inactive

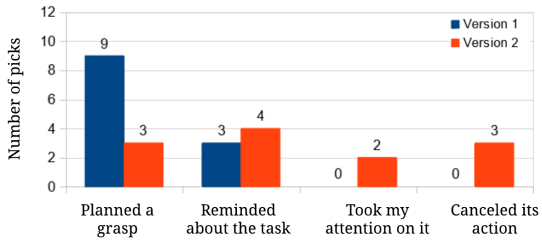
Results:

The replies of the 12 participants are summarized in the histogram graphs in Figure 6.11. For the planning gesture, Figure 6.11a shows that the planning version 1 was correctly recognized by 75% of the participants, while the others thought the robot was reminding about the task. Planning version 2, on the contrary, was only understood by 25% of the participants. However, we note that Planning-1 gesture penetrates a lot into the workspace of the human to communicate its intention more explicitly. Our early interaction experiments have pointed out that such intervening gestures intrusively lead humans to leave the placement task to the robot and thereby hinder effective collaboration. Therefore, we have decided to use both of the planning approaches, where planning-1 is executed by the robot right before it grasps the object (i.e., only if it is certainly taking over) and planning-2 is executed if the robot may take soon (i.e., the robot

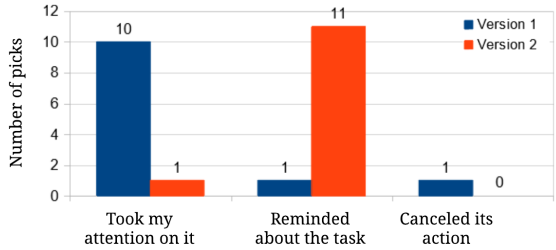
is in anticipation stage-1 in Figure 6.9). We also conclude that besides designing for natural expressions, a good instruction still needs to be given to the workers about the robot moves and in which cases they can still safely operate or they should be precautious.

For the pointing gesture, Figure 6.11b indicates that around 92% of the participants could correctly recognize the pointing-2 action as a reminder about the task. Contrarily, the most participants interpreted the pointing-1 action as a waving gesture only to attract the participant's attention on the robot to start a communication, leaving the actual intention, i.e., reminding, unanticipated. Hence, we pick the pointing-2 gesture as the reminder action. We find that redirecting the attention of a person is more successful if, at first, shared attention is ensured. Both of the two pointing versions have the rearing up and going down actions that took the attention of the human on the robot at first. After shared attention is achieved, then the robot's action to point toward the task is clearly understood as a reminder. However, the robot does not have a mechanism to check if shared attention is ensured. Therefore, sometimes the repetition of going up and down motion (twice in our case) could be misunderstood by a person if during the action the attention is already taken. For future implementations, such a feedback loop is suggested to avoid unnecessary repetitions of gestures that might be interpreted wrongly.

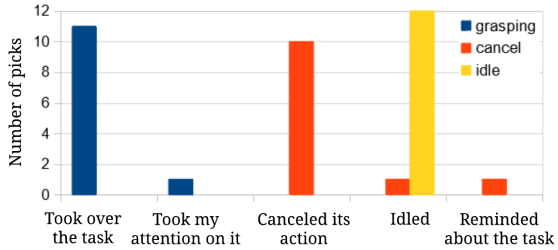
Finally, in Figure 6.11c the responses for the other gestures are summarized. The idle action was correctly recognized by all of the participants. The grasping was interpreted correctly in 92% of the cases and for the canceling action, this rate was 83%. It can, therefore, be stated that the implemented gestures effectively express the robot's intentions and are understandable to the majority of the participants. The gestures are able to effectively communicate the robot's intention even though the participants have never interacted with a robot before. Hence, our findings support **Hypothesis 3**. As a conclusion, the experiments have highlighted the importance of practicing the collaboration with the robot, before a participant start the actual experiments, to evaluate our decision-making approaches more effectively.



(a) The intention behind the planning gestures.



(b) The intention behind the pointing gestures.



(c) The intentions behind the grasping, canceling and idling gestures

Figure 6.11: The intentions behind the robot gestures as understood by the participants.

6.3.1.2 Evaluating Cognitive Load of the Tasks

In this experiment, our goal is to support the hypothesis below:

Hypothesis 4. *The collaboration setup puts a cognitive load on the humans and the degree of cognitive load differs between the task types. As a result of the cognitive load and/or long working hours, unanticipated human behaviors are invoked and observed during a task.*

In order to test this hypothesis, the participants and the robot collaborate over an extended period executing our tasks designed for the experiments (see in Section 6.2.2). Each participant is expected to work on a total of 9 placement tasks with changing types (see the different types in Figure 6.4). Our goal is to compare the types and pick the most suitable one for our following experiments. Our criteria are: the task is cognitively demanding enough to observe unanticipated human behaviors and the task is easy and motivating enough to keep the human engaged in the collaboration.

Objective measures: Below are the objective measures we have taken into account during this experiment:

- **Rewards gathered**: The robot receives positive rewards for each correct object placement (regardless of who achieves it) and is punished for wrong placements and everytime the human warns it stop.
- **Number of robot interference**: How many times the robot has taken over the task.
- **Number of warnings**: The amount of warning gesture a participant made during a task.

Subjective measures: 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. Additionally, the NASA Task Load Index (NASA-TLX) measures are used to evaluate the task load of the different task types (Hart and Staveland, 1988). It measures the task load of a task in 6 dimensions, namely *mental demand*, *physical demand*, *temporal demand*, *performance*, *effort* and *frustration*. The NASA-TLX statements are rated on a 20 point scale. In Table 6.6 we show the questionnaire statements, along with our categorization of them. The statements marked with the category **General** target the experiment in general, and so they are asked only once after the whole experiment is completed (i.e., after all of the tasks are completed). The other statements are task-specific and they are repeatedly asked after each task.

Experiment protocol: We have done some initial tests and realized that task type-3 (in Figure 6.4c) is very difficult and cognitively demanding for humans, which has often led the participants to leave the tasks to the robot. As we could run an experiment only with 9 tasks due to the availability of the participants, which already takes approximately 1.5 hours, we could run each task type only 2 or 3 times depending on the total amount of types compared. This exposure was not long enough for the participants for some

Table 6.6: Subjective statements asked during *Experiment-0*.

Questionnaire Statements	Category	Type
"The tasks were tedious (boring and monotonous)."	Overall cognitive load	Task specific
"The task was challenging for me."		Task specific
"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."		Task specific
"It became harder to remember the rules over the course of the task."		Task specific
"I lost my attention during the task."		Task specific
"The robot helped me remembering the rules"		Task specific
"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	Task specific
"How physically demanding was the task?"		Task specific
"How hurried or rushed was the pace of the task?"		Task specific
"How successful were you in accomplishing what you were asked to do?"		Task specific
"How hard did you have to work to accomplish your level of performance?"		Task specific
"How insecure, discouraged, irritated, stressed, and annoyed were you?"		Task specific

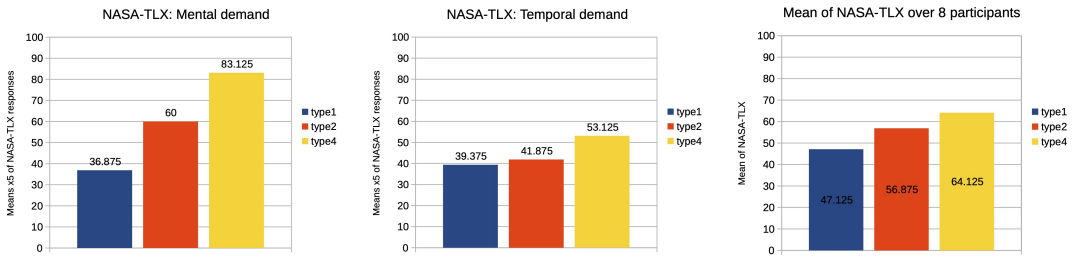
types to understand how a task functions, e.g. in type-5 with the Stroop effect ((Stroop, 1935)) as in Figure 6.4e. Therefore, we drop it from our analysis in this experiment and in Experiment-1 in Section 6.3.2 as it produces heterogeneous effects for different participants introducing more uncertainty into the evaluations.

The experiment uses a within-subject design, in which each of the participants executes each of the 3 task types we would like to examine (type-1, type-2 and type-4 as given in Figure 6.4) 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 and/or the practice-effect the participants may gain throughout the experiment. After the participants arrive, they are first asked to complete a survey about their demographics, a short questionnaire asking their age, background, previous interactions with robots, and their genders. Afterward, we introduce the industrial scenario of a human and the robot collaborating on an assembly line to inspect and pick and place the products on the relevant containers. We also add that the tasks are initially assigned to the participants, and the robot's job is

to assist whenever needed. In the end, both of their goals is to let the human do the sorting correctly for a higher quality of the job and the robot should take over only if it is deemed part of the work, mostly to avoid a wrong sorting.

Then, we run a training phase at the beginning of the experiment to ensure that the participants fully understand all regulations of the experiment and can operate correctly. Additionally it lowers the practice-effect, as all participants have had the same amount of training on the setup before the experiment starts. We demonstrate the participants how to grasp and place the objects and how to interact with the robot. We also remind them that whenever an object is detected in one of the containers, the participants should wait for the audible feedback indicating that the placement is processed. Additionally, an example rule is displayed on the rule monitor and we explain that the rules are only visible for a limited amount of time. Finally, the robot gestures are shown once to each participant and they are described with the robot's intention behind each gesture. In total 12 participants interacted with the robot using the same decision model, out of which 4 had to be excluded due to the lack of sufficient English to fill in the surveys or due to the technical problems during the experiment that led to corrupted data.

Results and Discussions:



(a) mental demand (b) temporal demand (c) TLX score over 6 dimensions

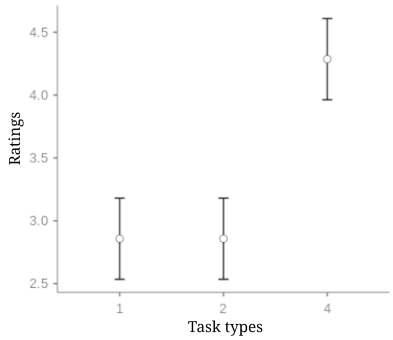
Figure 6.12: NASA-TLX scores

Starting with the analysis of cognitive load induced on the participants, we measure a global scale, i.e., NASA-TLX, to reach a general conclusion of how much workload the participants felt during each task. It should be noted that the NASA-TLX rates a task

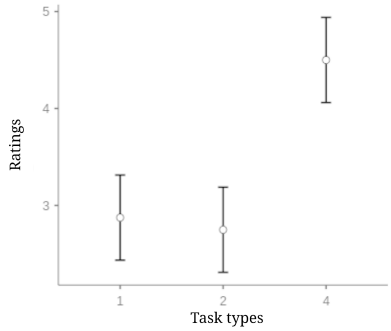
load in six dimensions, and here we discard the physical load in our analysis as our tasks do not require any, which is also stated by the participants. Since in our experiments the tasks have been designed to be cognitively demanding, our focus is on the mental demand dimension that is depicted in Figure 6.12a. Task type-4 has been reported to require the highest mental workload with a mean score of 83.125 out of 100, while task type-2 demanded a score of 60, and for type-1, it is 36.875. Figure 6.12b indicates that the participants have not felt any temporal demand during the task, as the time has not been used as a performance measure and we have not encouraged competition between the robot and the human. The average scores for all 6 dimensions in NASA-TLX are given in Figure 6.12c.

Figure 6.13 shows the survey responses for the statements 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 6.13a). Additionally, an η^2 of 0.42 indicates a large effect size. The post-hoc Tukey-HSD reveals that task types-1 and -2 are perceived as significantly less challenging than type-4 ($P_{tukey} = 0.015$). The mean difficulty for task type-4 is 4.29 out of 5, which is in line with the NASA-TLX test. Similar results are also obtained for the exhaustion and the attention lost during the tasks (see Figure 6.13b and Figure 6.13c, respectively). Task type-4 was significantly more exhaustive and caused more distraction for the participants. As the task rules are perceived to be difficult, the participants needed to look at the task monitor (see Figure 6.2) several times during a task to track its current state. This is also recognized by the robot as distracted since the attention is removed from the work environment.

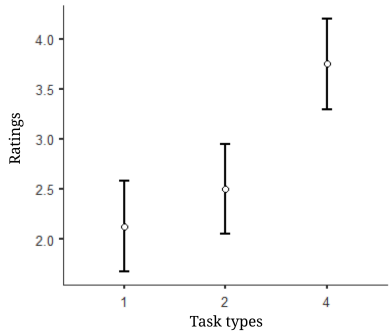
We also ask some general statements to track the change of cognitive load in time. The participants agree that they became more and more exhausted during the experiment for all task types ($mean = 3.875$). This indicates that a task with the same difficulty is perceived to be more demanding in time (in our case, approximately 1 hour). From Figure 6.13d one can conclude that task type-4 provokes the perception of an increase in the task difficulty over time with a strong agreement from the participants (mean =



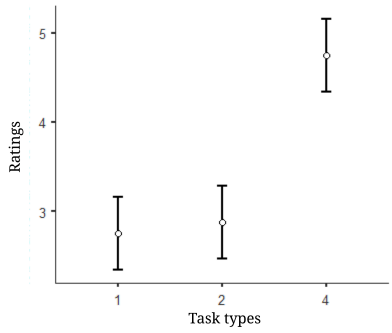
(a) "The task was challenging for me"



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



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

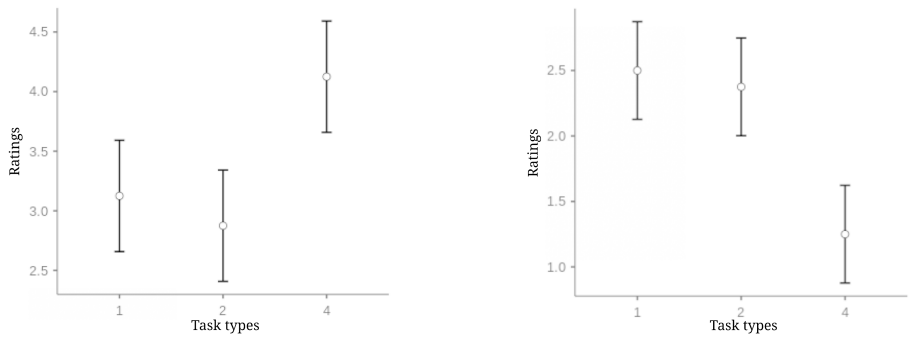


(d) "It became harder to remember the rules over the course of the task."

Figure 6.13: Mean plot of the participant ratings in Likert scale to the statements for the task difficulty and the cognitive load

4.75). On the contrary, the participants do not agree or disagree that task types-1 and -2 are becoming more difficult over time ($mean_{type1} = 2.75, mean_{type2} = 2.88$). The ANOVA confirms the significant difference between the task types ($p = 0.003$) with a large effect size ($\eta^2 = 0.419$) and the post-hoc analysis reveals that type-4 significantly differs from type-1 ($P_{tukey} = 0.006$) and type-2. This shows that task type-4 is best suited for the cases in which the perceived cognitive load on the participants should increase over time. From the analysis above, we can deduce that type-4 is significantly challenging and induces a cognitive load over the participants.

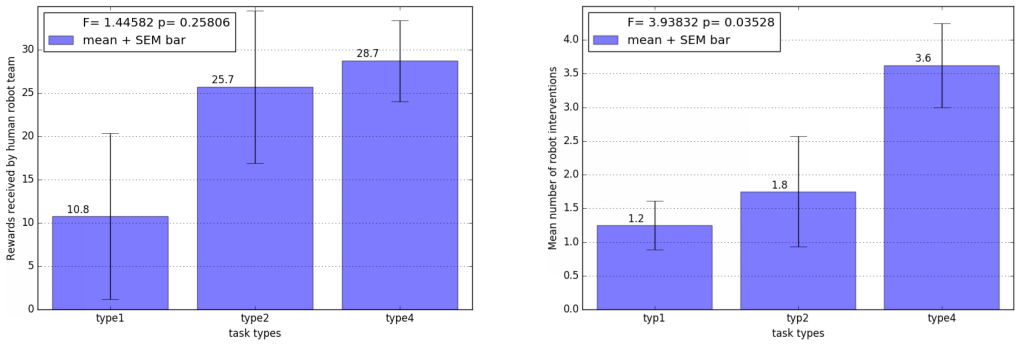
Since we want collaboration on a task, it should not be too overwhelming for the participants and demotivating a lot to leave the task completely to the robot. For this



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

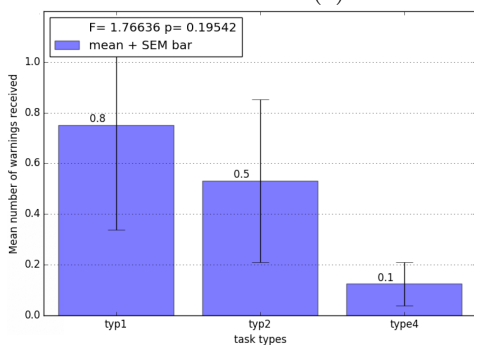
(b) "I would have scored better without the robot"

Figure 6.14: Mean plot of the participant ratings in Likert scale to the statements related to how they perceive the robot's collaboration.



(a) Collected rewards

(b) The amount of robot interference



(c) The amount of human warnings

Figure 6.15: Objective measures averaged over the participants for each task type

purpose, we also conduct analysis checking the success rates, robot's interference, and the participants' opinions about the robot and the reliability of its reactions in a task. First of all, Figure 6.15a compares the received rewards ($mean_{type1} = 10.8, mean_{type2} = 25.7, mean_{type4} = 28.7$) with no significant difference ($p = 0.258$) (see the reward calculation in Section 6.2.4.1). This indicates that even though type-4 is stated to be the most difficult, the mean reward and the success rate are still the highest for this task. This is mainly due to significantly large contributions of the robot in this task, as shown in Figure 6.15b. A robot interference describes a successful take over of the robot, which is a successful placement of an object without receiving a warning from a participant. In other words, the participants also approves the robot's assistance offer. Most robot interferences occurred during task type-4 ($mean = 3.6$), which is significantly larger than during type-2 ($mean = 1.8$) and type-1 ($mean = 1.2$) with $p = 0.035$. This is also supported by the warning amounts given in Figure 6.15c.

Finally, the participants also agree that the robot, in general, helped them to remember the rules in task type-4 ($mean = 4.13$), whereas on the others this is very low (see Figure 6.14a). They also state that in the tasks of type-4 they would have scored significantly worse without the robot's help than in the tasks running type-1, $p = 0.027$ and type-2, $p = 0.045$, (as shown in Figure 6.14b). As a result, it can be stated that the robot could effectively support its human collaborator and a higher number of robot interference has led to higher success rates during the collaboration on task type-4, which is stated to be the most challenging with higher rates of cognitive load. We also observe that the setup is effective in invoking unanticipated human behaviors. This is clear in task type-4, where the robot has interfered the task much more than the others, often without a human warning. Also, the statements under Figure 6.13 are a direct indication of a participant's misplacement, her failure to remember after long waiting times or her lost attention. These analyses prove that the experiment setup is able to put a cognitive load on the humans, the degree of which differs between different task types, and the higher cognitive load leads to more unanticipated human behaviors followed by more human errors, supporting **Hypothesis 4**. As it has led to a fluent collaboration with a

significant amount of invoked unanticipated behaviors, we pick task type-4 as the main collaboration task to test our robot's adaptation capabilities in *Experiment-1*.

6.3.1.3 Feedback for Our Robot Decision-Making System

In this section, our goal is to list the necessary updates needed on the robot's decision-making moving from the findings on the robot's collaboration capability during *Experiment-0* in Section 6.3.1.2. These findings are then fed back in the design process of the robot decision models in Section 6.2.4.1, and in the observation agent in Sections 6.2.3.1 and 6.2.3.2. The A-POMDP models have been designed as high-level decision-making on simulation, where the states and actions can fit into 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; hence, affecting also the robot's decision updates and the model's state transition probabilities. For this purpose, most common interaction cases are identified and examined in detail to assess the robot's reliability in its reactions. We list below some of these common interaction scenarios at different execution speeds and intervals, which are repeatedly analyzed to generate our findings for the robot's decision-making in real-time.

- 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: After a long time of idling, the robot may take over depending on the interacted human's preferences and the decision strategies of a robot model).
- Human warns the robot and the robot should receive and process the warning on time no matter what the robot 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 robot should estimate that the human may lost her attention and take an action.

Real-time observation update frequency: For the observation agent's update frequency (in Figure 6.5), we analyze the average duration of human actions. As Table 6.7 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 found out to be the longest human action on average. We, therefore, set the same action timeout (i.e., a duration when the robot ignores new human observations as long as they are not changing, as shown in Figure 6.5) to be 3 seconds, which is the average time needed for an action to be completed. This also makes sure that during the longest action, i.e., grasping, 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, e.g., the human has grasped an object but idled long without placing it, it is processed to generate a response by the robot decision-making agent. If the robot reacts to each observation (i.e., a decision-making with $1Hz$), even if the human is still progressing with the same action, it will unnecessarily keep the decision-making block busy and possibly cause asynchronous responses. 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 6.7: Average durations of human actions while interacting with the robot

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

Robot decision-making models: First of all, to be able to track a human's status throughout a task consisting of several subtasks and multiple action steps, we add another partially observable state called "human is doing fine" to our A-POMDP model design (see in Section 6.4.1.1). This state lies in the first anticipation stage layer of our model design, allowing the robot to estimate if a human is doing fine or the person is having problems with a task. Then, our goal is to fine-tune the model with the collected observations from

the real experiments. The most difficult step is the tuning of the state transition and observation probabilities. We are aware that there cannot be one robot model that has these probabilities optimized for all types of people and interactions. For this purpose, as we discuss in Chapter 5, our initial goal is again to implement a base A-POMDP model that works reasonably in many interaction scenarios, which is then used to generate several other models for ABPS to select the most suitable one during a specific interaction. For this purpose, we first manually update a model that has shown to perform well with our initial findings from the human experiments (see Figure 6.9). Then, we fine-tune the model using several recorded observations collected throughout *Experiment-0* and the common interaction scenarios mentioned at the beginning of Section .

The final base model is used as the "proactive model" in our experiments to compare our A-POMDP with a reactive model (*Experiment-1* in Section 6.3.2). The reactive model used in Chapter 4 is also updated using the same human observations collected in *Experiment-0* to be fair in our comparisons. For *Experiment-2*, we repeat the steps in the simulation experiments of ABPS in Section 5.3 to generate a policy library. That is, 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 (following the steps under Algorithm 1 in Section 5.2.2). 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 in Section 6.3.3). Finally, we also update the human type estimation of ABPS also incorporating the real human observations collected. The labeling of the human types is done from the objective measures collected from the participants, such as, their success rates, the warning amounts, etc (see the details in Section 6.2.4.2). The updated estimation 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 *Experiment-0* thanks to the simulated data.

6.3.2 Experiment-1: Validating Chapter 4, A-POMDP

In this experiment, our goal is to support the following hypotheses:

Hypothesis 5. *A robot's fluent collaboration with a human contributes to increased performance in a cognitively challenging task when compared to a human working alone.*

Hypothesis 6. *Our A-POMDP model covering and adapting to a human's unanticipated behaviors (extended short-term adaptation) contributes to a more efficient collaboration when compared to a robot model that does not handle such behaviors (detailed in Section 6.3.2.1).*

Hypothesis 7. *Our A-POMDP model covering and adapting to a human's unanticipated behaviors contributes to a more natural collaboration when compared to a robot model that does not handle such behaviors (detailed in Section 6.3.2.1).*

Hypothesis 8. *Our A-POMDP model covering and adapting to a human's unanticipated behaviors has a higher perceived collaboration, trust, and positive teammate traits than a robot model that does not handle such behaviors (detailed in Section 6.3.2.2).*

Hypothesis 9. *Our A-POMDP model covering a human's unanticipated behaviors is able to show a better adaptation than a robot model that does not handle such behaviors.*

Our goal is to repeat the experiments done in simulation in Chapter 4 and demonstrate that anticipating and taking into account the unanticipated human behaviors increases the overall efficiency and the naturalness of an HRC. Similar to the simulation experiments, we let the participants interact autonomously with two different robot planners. The first one is called the *proactive robot* that runs our A-POMDP model in Figure 6.9 and also handles a human's unanticipated behaviors (extended short-term adaptation). This robot first anticipates a human's, e.g., lost attention, incapability, tiredness, then, following these estimations, it estimates if the human needs/wants assistance or not. On the contrary, the other robot, called the *reactive robot*, does not handle the unanticipated behaviors of a human (see Figure 4.2). It treats a human's need for help as a directly

observable (deterministic) state and directly takes over a task no matter what the human's actual internal state is. 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. Hence, through this comparison, our intention is to show the importance of handling the unanticipated human behaviors (i.e., stochastic interpretation of the human states in the anticipation stage-1 in Figure 6.9) for an improved short-term adaptation.

We have conducted within-subject experiments with 14 people from various backgrounds and let them interact with the two robot models.

6.3.2.1 Objective Measures

Our objective is to measure the naturalness and the efficiency of the collaboration. Overall efficiency is mainly explained as the success rate on a task. However, in a work environment, success is also defined by the quality of the work. When we consider the cases where a work division in a task is made according to a known skillset of the collaborating partners, it is expected that each work is completed by its assigned member for higher quality. For example, in a pick and place task, a changing fragility of the products conveyed on the belt may require a work division, where a human should handle the fragile objects while the robot should work on bulky and heavy ones. Besides meeting the skillsets, a work division is also crucial regarding the efficient use of resources. 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 robot's human anticipation skills and its correct interpretation of an assistance need rather than how it executes a task. In reality, our designed tasks are achievable by both the robot and a human due to the limitations of the setup. However, we motivate the participants to handle a subtask by informing them about our scoring system that favors a human achieving a task (see in Section 6.2.2). This information is also known to the robots by

initializing A-POMDP models from the state of "task is assigned to human". As a result, higher efficiency is expected when a task is successfully accomplished by a human. To measure the efficiency, we first calculate a participant's successful contribution to the task as given:

$$C_{human} = \frac{n_{s_{human}}}{n_{total}} \tag{6.3}$$

where $n_{s_{human}}$ is the number of successful placements (successful subtasks) achieved by a participant during a task and n_{total} is the total amount of subtasks. The division gives us a human's successful contribution in a task, C_{human}. The overall success rate, S_{task}, is calculated as:

$$S_{task} = \frac{n_s}{n_{total}} \tag{6.4}$$

where n_s is the total number of successfully achieved subtasks in a task. Following our discussions above, we calculate the efficiency, η_{task} that bares the effects of both the overall success rate and how many of them are achieved by a human collaborator:

$$\eta_{task} = S_{task} \cdot C_{human} \tag{6.5}$$

In addition, for further analysis we also calculate a human's success rate in:

$$S_{human} = \frac{n_{s_{human}}}{n_{s_{human}} + n_{f_{human}}} \tag{6.6}$$

where $n_{s_{human}}$ and $n_{f_{human}}$ are the amount of successful and failed subtasks by humans in a task, respectively. A human success rate, S_{human}, can be defined as the amount of the successful placements of a human collaborator out of all of her attempts.

The term naturalness defines a natural interaction between the robot 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 (Hoffman, 2019b). A good indicator of the naturalness of collaboration is

the level of intrusive behaviors from the robot, which leads to frustration in a human partner. A robot should reliably understand the collaborating human's needs and preferences to adapt to a situation and to avoid intrusive behaviors. In our setup, a warning gesture is devised for a human collaborator so that she communicates her displeasure with a robot behavior. Therefore, the amount of warnings hints about the naturalness of a collaboration. To conclude, our objective measures are the following:

- **Rewards gathered:** We use the same robot reward mechanism as in Section 6.3.1.2 for both of the reactive and the proactive robot, i.e., the same reward function for the robot Markov models (see Section 6.2.4.1 for the reward function).
- **Number of warnings:** The amount of warning gestures a participant made during a task, which hints about the naturalness of a collaboration.
- **Task success rate (S_{task}):** The overall success rate of a task as in Equation 6.4.
- **Human success rate (S_{human}):** The success rate of a human as in Equation 6.6.
- **Human contribution in success (C_{human}):** The human's overall successful contribution in a task, as in Equation 6.3.
- **Task efficiency (η_{task}):** As calculated in Equation 6.5.

6.3.2.2 Subjective Measures

Our goal in the subjective analysis is to subjectively evaluate the fluency of the collaboration from the perspective of the participants. Similar to the ones in *Experiment-0*, the objective measures are collected by means of questionnaire responses that participants complete either after each placement task or at the end of the experiment. The participants use a 5-step Likert scale to describe their agreement with a statement. The statements are given in Table 6.8. The category of questions concerning the robot's effect on the cognitive load is to see if the robots are perceived as a negative or positive influence on such a challenging task. The remaining of the statements are for measuring

Table 6.8: Subjective statements asked during *Experiment-1.*

Questionnaire Statements	Category	Type
"The robot helped me remembering the rules"	Robot's effect on the human cognitive load	Task specific
"The robot's interference distracted me"		Task specific
"Without the robot, I could have achieved a better score."		Task specific
"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	Task specific
"The robot acted as I expected."		Task specific
"The robot took over at the right times when I needed assistance."		Task specific
"I lost my attention during the task."		Task specific
"I had to warn the robot a lot."		Task specific
"I became more and more exhausted during the experiment."		Task specific
"The robot reacts repetitive rather than responding to my behaviors."		Task specific
"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	Task specific
"I trust the robot with the task."		Task specific
"The robot seemed competitive in taking over the task."		Task specific
"The robot could have assisted me more."		Task specific
"I became more comfortable interacting with the robot over the course of the experimer		General
"I liked collaborating with the robot."		General
"I would work with such a robot on such kind of tasks in a working environment."		General

the metrics of *perceived naturalness, trust, positive teammate traits and perceived collaboration with the robots* that are mostly inspired from the relevant HRC research, e.g., in (Hoffman, 2019b; Koppula, Jain, and Saxena, 2016; Nikolaidis and Shah, 2013; Nikolaidis et al., 2017), and complemented with additional statements. Also, we repeat the same statements from *Experiment-0* regarding the cognitive load on humans. Even though we categorize the statements for simplicity, they are interchangeably evaluated under relevant hypotheses. There are also similar statements that are shuffled in the questionnaires. This is to check the consistency of a participant's ratings. The statements categorized as task-specific are asked after each task completion, whereas the general ones are rated 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 with a participant to let them compare the two robots.

6.3.2.3 Experiment Protocol

We have invited 14 people from different backgrounds. Three of them are from computer science, the others are from social sciences, molecular biology, law, electrical engineering,

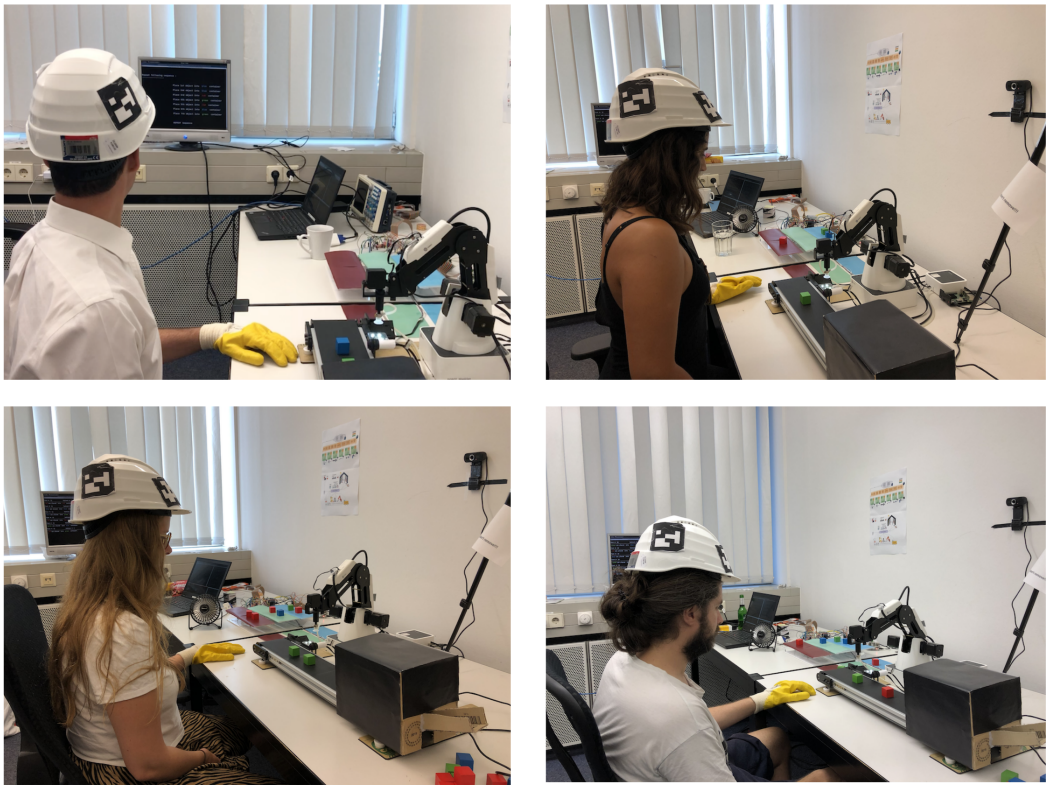


Figure 6.16: Participants collaborating with the robot during *Experiment-1*.

chemical engineering, tourism, public affairs, medical school, business school and high school students. We note that none of the participants have had any interaction with a robot before. The ages range from 17 to 38, and they are from different ethnic groups.

The experiment protocol is mostly the same as the protocol of *Experiment-0* (see in Section 6.3.1.2). A participant takes a seat and the experiment starts (see Figure 6.16). We make the same introduction to the setup and the definition of a task as in *Experiment-0*. The operator describes how to complete a task successfully and motivates the participant that the tasks are always assigned to the human and for each subtask the human achieves the team receives a higher score than the robot completing it (see the scoring system in Section 6.2.2). The operator also mentions that there is no time limitation. Finally, each participant is trained on how to warn the robot to stop, as well as how to grasp

and place the objects. One final remark is made for the sound feedback the participants need to wait after each placement, i.e., the approval signal from the system.

We designed a within-subject experiment to compare the two robots better, i.e., the proactive and the reactive robot. Thus, each of the participants interacted with both of the robot types. We note that the participants knew that there were two types of robots 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 anonimty. As mentioned, we have picked task type-4 as the most suitable task for collaboration in Section 6.3.1.2. 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 6.17. First, we do a training round with each participant on a simpler task. Then, the experiment starts. In general, there are three different types of interaction: 1) The participant completes one task alone, i.e., without any robot interaction, 2) The participant collaborates with the reactive robot for 3 tasks, 3) The participant collaborates with the proactive robot for 3 tasks. In total, each participant completes 7 tasks. In order to remove any practice-effect over different robot types, randomly selected 7 participants interacted first with the reactive robot and the other 7 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, a participant needs to also fill out general survey questions (see in Table 6.8). In total, an experiment with a participant lasts approximately 1,5 hours.

6.3.2.4 Results and Discussions

In this section, our goal is to support the hypotheses defined for *Experiment-1*. Both of the robot models receive the same observation vector from the environment (see in Section 6.2.3.2) and have the same rewarding mechanism. As mentioned, the reactive robot deterministically decides to take over a task if the human stays "idle" and/or "is looking around" for 15 seconds and if the human misplaces a cube (a failed subtask). We observe that this provides significantly faster reaction time to the reactive robot

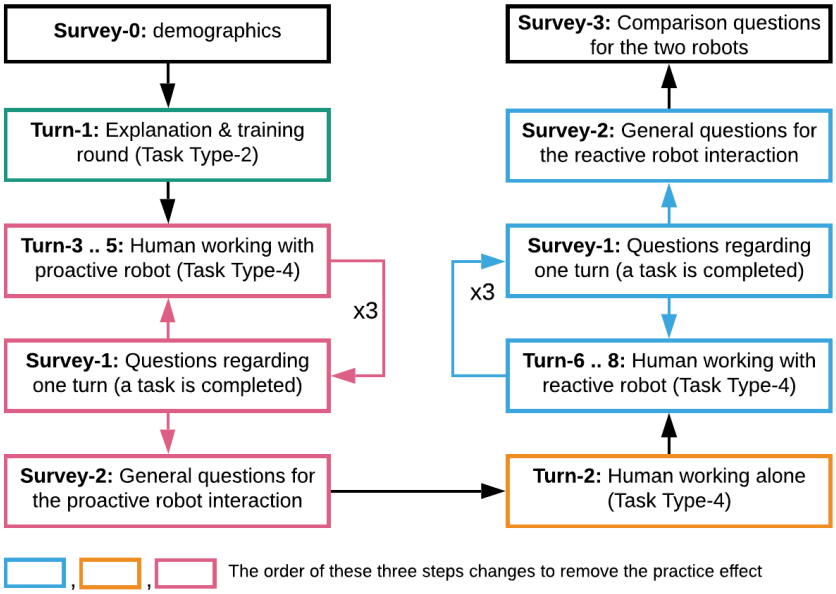


Figure 6.17: The procedure of *Experiment-1* for each participant

when compared to the proactive one that first assesses and makes sure the human really needs/wants the help. For example, most of the participants failed in a subtask but they still remembered the remaining of the task rules. If the proactive robot is able to estimate that correctly then it will wait in the state of "human is doing ok" (in Figure 6.9) to observe more and let the human continue whereas in the same case the reactive robot will take over immediately. This waiting time (i.e., iteration on the same state) becomes longer (i.e., an increased belief in this state) in the proactive robot if the human has achieved more subtasks before starting to fail. However, it gives the advantage of assessing a human's need for assistance more accurately.

Hypothesis 5 *In this hypothesis, we state that a robot's fluent collaboration with a human contributes to increased efficiency on a cognitively challenging task, when compared to a human working alone.*

In evaluating this hypothesis, we use the following measures:

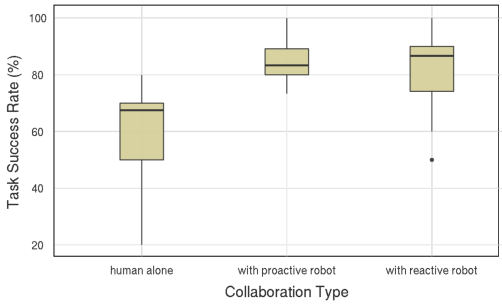
- Objective measures: Overall task success rates, human success rates, task efficiency

- Subjective statements: "The robot helped me remembering the rules", "a robot collaboration is beneficial in such tasks", "I could have achieved more without the robot".

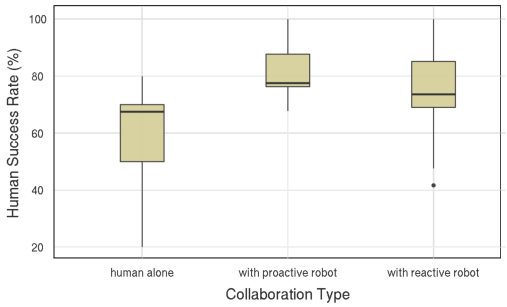
In Figure 6.18, we give the box and whisker plots of the success and efficiency analysis. As seen in Figure 6.18a, the success rate of the participants working without a robot is worse than a collaboration with any of the robot models. In Figure 6.18e, we further analyze that human alone cases reach a mean success rate of 60.36%, while in the proactive robot and the reactive collaborations are significantly higher with the means of 84.78% and 81.19%, respectively ($p < 0.05$). Collaboration with a proactive robot contributes positively to a task success by 41% and with the reactive one, it is increased by 35%. Similarly, a human's success rate has also significantly increased when the human collaborated with either of the robots (see Figure 6.18b), whereas in proactive one this increase is slightly more.

According to our observations, the robots helped the participants remember the task rules if they trust the robot, who has a very high success rate alone (above 95%). This is also in line with the participants' statements that the robots helped them remember the task rules (see the high mean ratings in 5-step Likert scale in Figure 6.19a) and that a robot collaboration is beneficial in such tasks (see Figure 6.19d). Finally, the measured task efficiencies are shown in Figure 6.18d and in Figure 6.18e. Both, the reactive and the proactive robot have contributed significantly positively to the task efficiency, compared to a human working alone (to an increase of 56% for the proactive robot and 21% for the reactive one to the overall task success rates). Thereby, we underscore the importance of such cobots collaborating with humans in challenging tasks. The success rate analysis, the efficiency results, and the subjective ratings of the participants support **Hypothesis 5**.

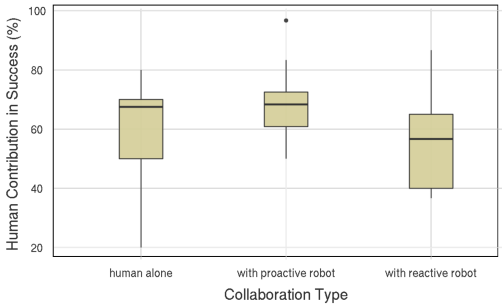
Hypothesis 6 *In this hypothesis, we state that our A-POMDP model covering and adapting to a human's unanticipated behaviors, i.e., the proactive robot with extended short-term adaptation, contributes to a more efficient collaboration when compared to a robot model that does not handle such behaviors, i.e., the reactive robot.*



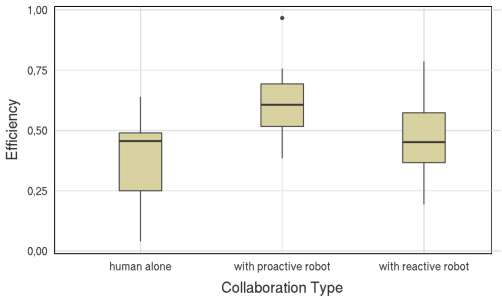
(a) Overall task success rates



(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.

Figure 6.18: (a)-(d) Box and whisker plots of the overall task success rates, a human's success rate, a human's successful contribution to the overall task, and the efficiency values all averaged over 14 participants during their alone performance 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 robots.

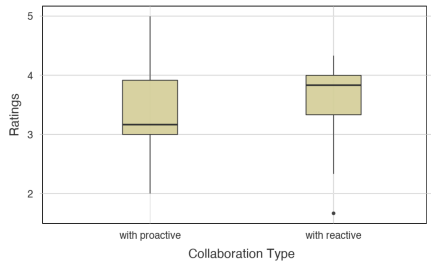
In evaluating this hypothesis, we use the following measures:

- Objective measures: Overall task success rates, human success rates in a task, a human's successful contribution to a task, task efficiency, and the rewards gathered by the robots.

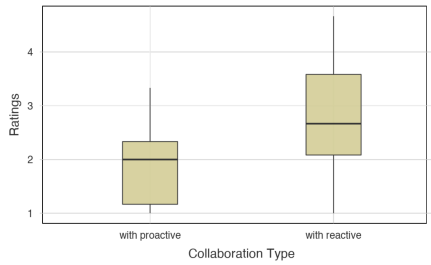
- Subjective statements: "Robot collaboration is beneficial in such tasks".

Even though the proactive robot on average has achieved more success than the reactive one, in Figure 6.18e we see that the ANOVA results on overall task success show no significant difference between the cases with two robots. However, the success rate distribution in the reactive robot case has a higher variance whereas in the proactive case it is distributed between $75\% - 100\%$ (see Figure 6.18a). This suggests that the proactive robot provides a more stable and a higher likelihood of more success than the reactive one. Our decision models only decide to the level of taking over a subtask, after which the performance of placing an object is the same for both robots (with the success rate of over 95%). Therefore, a significant number of the participants have preferred to leave the task to the robots, regardless of their type, once they realize the robot's high performance. This leads to higher success rates in cases where mostly the robot achieves the tasks; hence, we did not obtain a significant difference between the robots on the overall task success.

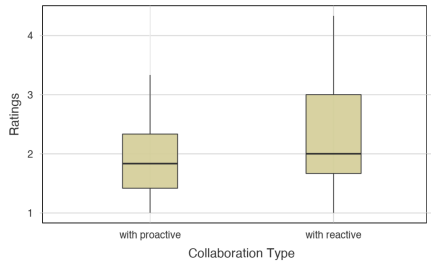
In terms of a human's contribution to the task, our analyses show that the proactive robot ensures the involvement and contribution of its human partner who is the main assignee of every task. In Figure 6.18c, we show that the proactive robot has significantly increased a human's contribution to success when compared to the reactive robot. The reactive robot also respects the initial assignment; however, it favors for taking over a task when, for example, a human idles too long or they fail in a previous subtask, regardless of the dynamic human traits and preferences. This has led to a decrease in a human's contribution to success during her collaboration with the reactive robot when compared to her performance alone (as shown in Figure 6.18e). On the other hand, the proactive robot encourages its human partner to fulfill a task. The difference between the human



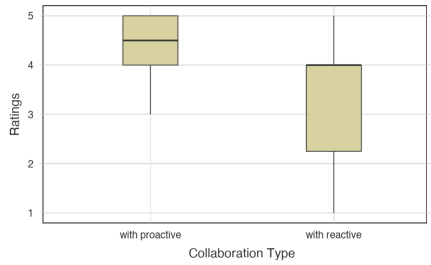
(a) "Robot helped me remembering the rules."



(b) "Robot interference distracted me."



(c) "I could have achieved more without the robot."



(d) "A robot collaboration is beneficial in such tasks."

Statement	Mean rating with proactive robot	Mean rating with reactive robot	ANOVA
"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**

(e) Mean values and the ANOVA results.

Figure 6.19: (a)-(d) The subjective statements that asked to the participants for analyzing a robot's impact on such a challenging task. We use a 5-step Likert scale for the ratings and the results are averaged over 14 participants. (e) The ANOVA results of the subjective statements on the robot impacts for both of the robots.

contribution in success rates of the two robots is significantly in favor of the proactive robot ($p = 0.038$).

As we mentioned before, higher efficiency is achieved when a task is successfully accomplished by its assigned collaborator. The efficiency results are shown in Figure 6.18d. It is clear that the collaboration with the proactive robot has mostly guaranteed an efficiency above 50%. Figure 6.18e shows that the proactive robot has achieved a mean efficiency of 0.61 out of 1.0, significantly increasing the efficiency of a human working

alone by 57% ($p = 0.0023$), also ruling out the efficiency reached with the reactive robot ($p = 0.0104$). Finally, the participants think that a collaboration with the proactive robot is significantly more beneficial in such challenging tasks than a collaboration with the reactive one ($p = 0.040$ in Figure 6.19d). From the objective and subjective analyses on the efficiency, we conclude that a collaboration with our A-POMDP robot model, i.e., the proactive robot, reaches to a more efficient collaboration when compared to a collaboration with a collaborative robot that also anticipates human behaviors but does not handle a human's unanticipated behaviors. This supports **Hypothesis 6**.

Hypothesis 7 *In this hypothesis, we state that our A-POMDP model covering and adapting to a human's unanticipated behaviors (extended short-term adaptation) contributes to a more natural collaboration when compared to a robot model that does not handle such behaviors.*

In evaluating this hypothesis, we use the following measures:

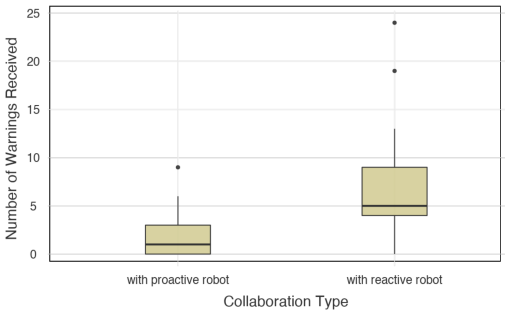
- Objective measures: The number of warnings received by a participant.
- Subjective statements: "Robot interference distracted me", "I had to warn the robot a lot", "the robot was able to understand (anticipate) my assistance needs and behaviors during the task", "the robot acted as I expected", "the 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".

As discussed in Section 6.3.2.1, the naturalness reflects a fluent communication between a human and a robot where they understand their behaviors and respond reliably. The handovers and turn-taking need to be interpreted correctly, even non-verbally, by both of the collaborators. In our setup, a robot receiving fewer warnings (i.e. stop commands) and positive subjective statements after a task from a human collaborator mean that it is able to provide a more fluent and so a more natural collaboration. In Figure 6.20a, we show the distribution of the number of warnings each robot has received during the experiments. The distribution suggests that the proactive robot could keep the warnings

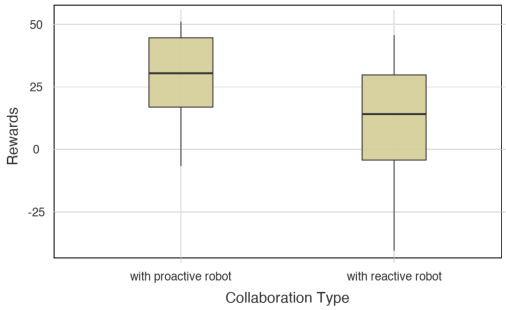
close to zero, meaning that it mostly estimated a human's unanticipated behaviors and the need for help correctly. In Figure 6.20c, we reflect the ANOVA results that point out a significant difference between the warning levels of the two robots ($p < 0.0001$), where the mean value of the reactive robot is 3.6 times of the proactive robot's.

The participants have also evaluated if a collaboration with a robot has felt comparatively natural to them, i.e., more human-like. A natural collaboration should not hinder the performance of any of the collaborating partners. However, the participants stated that the reactive robot's interference distracted them significantly more than the proactive case (with $p = 0.023$ as shown in Figure 6.19e). This is majorly due to the unexpected interferences of the reactive robot, which is also supported by the participant ratings for the statement of "the robot acted as expected" being significantly higher for the proactive robot than the reactive one (with $p = 0.012$ as shown in Figure 6.21g). 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 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 (Bauer, Wollherr, and Buss, 2008; Hoffman, 2019b; Thomaz, Hoffman, and Cakmak, 2016). The reactive robot discards unanticipated and dynamic human behaviors, for example, it mostly took over when a participant was evaluating (trying to remember the rules) before acting on a subtask. This poor anticipation of a human's assistance needs and behaviors is also stated by the participants as they rated the reactive robot significantly lower in its ability to anticipate than the proactive robot (with $p = 0.006$ in Figure 6.19e).

Total discounted rewards the robots have received after each task are shown in Figure 6.20b as a distribution. Since it is the combination of task success and the number of warnings the robot received in a task, a higher reward informs about fewer number of warnings received and more subtasks succeeded (i.e., more cubes are placed correctly). In Figure 6.20c, we analyze that the proactive robot has received significantly more rewards ($p < 0.0001$) than the reactive robot, multiplying the mean reward of the reactive one by 2.6. This directly implies a significantly more efficient and more natural collaboration,



(a) Number of warnings received



(b) Rewards gathered by the robots

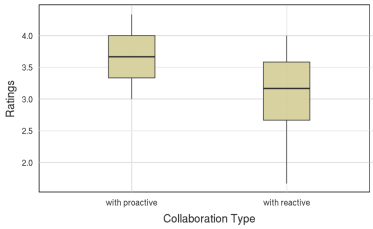
	Comparisons	Mean values with proactive robot	Mean values with reactive robot	ANOVA
Number of warnings	with proactive robot vs with reactive robot	1.88	6.79	*F(1, 82) = 37.68, p < 0.0001* **Significant**
Rewards	with proactive robot vs with reactive robot	28.95	10.88	*F(1, 82) = 17.00, p < 0.0001* **Significant**

(c) Mean values and the ANOVA results

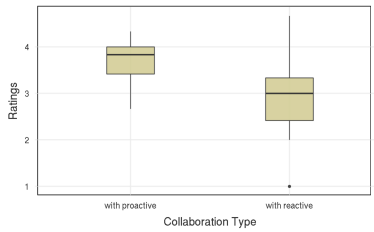
Figure 6.20: (a), (b) Box and whisker plots of the number of warnings the proactive robot and the reactive robot have received from a human and the rewards they gather during a task, averaged over 14 participants worked on 42 tasks for each of the robots in total. (c) Mean values of the two robots and the ANOVA results for the comparison of their performance.

as another support of **Hypothesis 6** and for **Hypothesis 7**. Finally, the participants also stated that the proactive robot was able to adapt significantly better to their assistance needs (with $p = 0.004$) whereas the reactive robot behaves repetitive rather than responding to their changing behaviors ($p < 0.001$, see the rating distributions and analysis in Figure 6.21). The participant ratings show great consistency, suggesting that the unexpected robot interference occurs mostly due to the wrong anticipation of and adaptation to the current behavior of a person. This then results in more warnings to stop the robot interference as also validated by the participants with their higher ratings for the reactive robot ($p < 0.001$ in Figure 6.21d). All these objective measurements and subjective statements of the participants on naturalness supports **Hypothesis 7**.

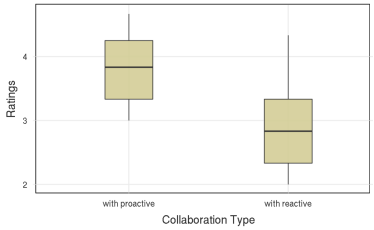
Hypothesis 8 *In this hypothesis, we state that our A-POMDP model covering and adapting to a human's unanticipated behaviors (extended short-term adaptation) has a higher*



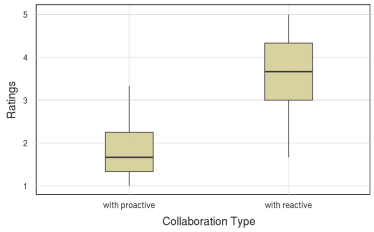
(a) "The robot was able to anticipate my assistance needs and behaviors during the task."



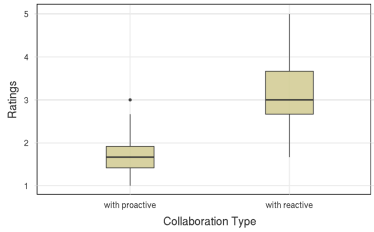
(b) "The robot acted as I expected."



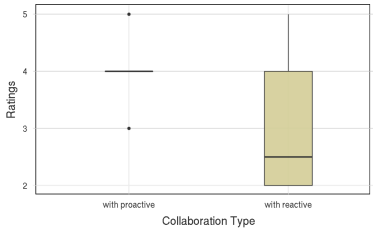
(c) "The robot took over at the right times when I needed assistance."



(d) "I had to warn the robot a lot."



(e) "The robot reacts repetitive rather than responding to my behaviors."



(f) "The robot was able to adapt to my assistance needs over the course of the experiment."

Statement	Mean rating with proactive robot	Mean rating with reactive robot	ANOVA
"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**

(g) Mean values and the ANOVA results on the subjective statements.

Figure 6.21: (a)-(f) Box and whisker plots of the subjective statements that are asked to the participants for analyzing the robot's reliability and adaptation ability, and the naturalness of the collaboration. We use a 5-step Likert scale for the ratings and the results are averaged over 14 participants. (g) The statements are also compared and analyzed for the proactive and the reactive robot.

perceived collaboration, trust, and positive teammate traits than a robot model that does not handle such behaviors.

In evaluating this hypothesis, we use the following measures:

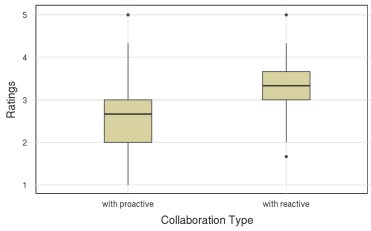
- Objective measures: Number of warnings received by a participant.
- Subjective statements: "The robot took over at the right times when I needed assistance", "the robot acted as I expected", "the robot was able to adapt to my assistance needs over the course of the experiment", "I did not want the robot to take over the task", "the robot could have assisted me more", "I trust the robot with the task", "The robot seemed competitive in taking over the task", "I liked collaborating with the robot", "I would work with such a robot on such kind of tasks in a work environment".

This hypothesis is evaluated through the subjective statements of the participants. First of all, we recall that the participants are only informed about an overall success rate and a score value (also success related as mentioned in Section 6.2.2) to inform how well they did during a task. That said, their ratings to the statements can only be affected by this information and their own observations during the task. As mentioned before, both of the robots are equally capable of achieving a task with almost 95% success rates. Since the collaboration with both of the robots has achieved very high and almost similar success rates, the participants rated their trust to both of the robots very high (the proactive robot with the mean rating of 4.43, the reactive robot with 4.05 out of 5.00 as shown in Figure 6.22b). In fact, we previously speculated that a robot's success rate would be the highest influence in the trust evaluations of the participants so we were not expecting a significant difference between the trust levels of the robots. On the contrary, the participants still thought that the proactive robot is significantly more trusted than the reactive one (with $p = 0.041$ as in Figure 6.22g).

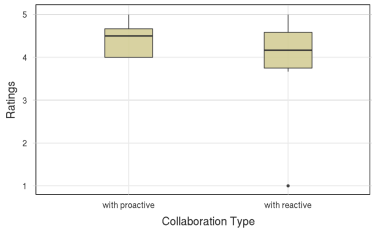
The higher trust evaluation for the proactive robot is in line with and can be explained through the other subjective statements. We asked the participants to evaluate the anticipation skills, the timing, and the general responses of the robots. They think that

the proactive robot took over the task at the right times when they needed assistance with significantly more accurate timing than the reactive robot (with $p = 0.002$), stressing the proactive robot's more reliable responses. 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 (the former with $p = 0.012$ and the latter with $p = 0.004$ from Figure 6.21g). 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 higher for the reactive robot that took over more frequently ($p = 0.030$ in Figure 6.22g). The number of warnings received by the robots also support this statement (see Figure 6.20a), showing the participants' higher dissatisfaction with the reactive robot in general. One rating, nevertheless, shows a slight conflict with these statements. The participants rated "the robot could have assisted me more" more for the proactive robot (see Figure 6.22d). This, however, does not have a compelling effect as the rated difference is not significant ($p = 0.543$ as shown in Figure 6.22g), and the mean values of both are very low ($\mu_{proactive} = 2.29, \mu_{reactive} = 2.17$). That means the participants believe that both of the robots have assisted them enough or above their expectations.

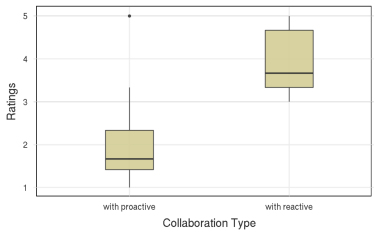
In general, better anticipation of a person's assistance needs, right timing in assisting, and more trust suggest a higher acceptance for the proactive robot. We also showed that the proactive robot has provided better timing for the assistance, respected its partner's needs and preferences, contributed to an increased performance of its partner, and led to a more efficient task. As a consequence, we derive 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 such kind of demanding tasks significantly more than the reactive robot, even though both of the robots are rated very high ($\mu_{proactive} = 4.50$ and $\mu_{reactive} = 3.43$ out of 5.00, with $p = 0.004$ as shown in Figure 6.22g). More positive teammate traits and a higher trust may already indicate a higher perceived collaboration for the participants. To support this, we also asked the participants more direct statements. They think that



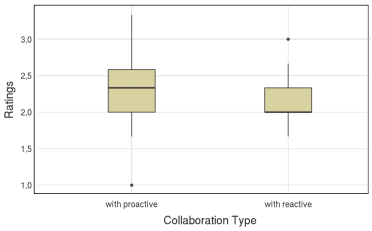
(a) "I did not want the robot to take over the task."



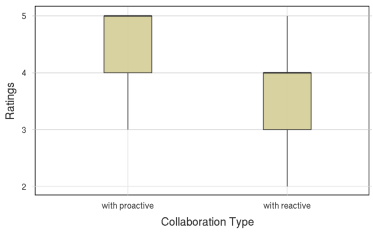
(b) "I trust the robot with the task."



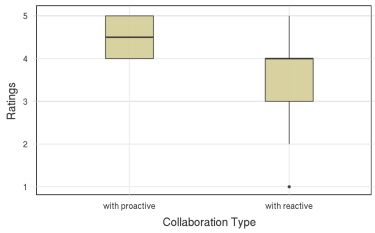
(c) "The robot seemed competitive in taking over the task."



(d) "The robot could have assisted me more."



(e) "I liked collaborating with the robot."



(f) "I would work with such a robot on such kind of tasks in a work environment."

Statement	Mean rating with proactive robot	Mean rating with reactive robot	ANOVA
"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**

(g) "Mean values and the ANOVA results of the statements."

Figure 6.22: (a)-(f) Box and whisker plots of the subjective statements that are asked to the participants for analyzing their trust in the robots, their perceived collaboration, and teammate traits. We use a 5-step Likert scale for the ratings and the results are averaged over 14 participants. (g) The statements are also compared and analyzed for the proactive and the reactive robot.

the reactive robot was competitive and aggressive in its behaviors, whereas the ratings for the proactive robot were below average as seen in Figure 6.22c ($\mu_{proactive} = 2.10$ and $\mu_{reactive} = 3.90$ out of 5.00 with $p < 0.001$). In a collaboration task where the partners share a mutual goal, a competition between the partners is not encouraged for the 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 ($p = 0.011$). This shows a higher perceived collaboration for the proactive robot and supports **Hypothesis 8**.

Hypothesis 9 *In this hypothesis, we state that our A-POMDP model covering a human's unanticipated behaviors is able to show a better adaptation than a robot model that does not handle such behaviors.*

In evaluating this hypothesis, we use the following measures:

- Subjective statements: "The robot was able to understand (anticipate) my assistance needs and behaviors during the task", "the robot took over at the right times when I needed assistance", "the robot could have assisted me more", "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 comparison questions: "Which robot was more reactive (following preset rules) in its takeovers rather than anticipating and responding to your behaviors?", "which robot was learning and adapting better to your assistance needs?"

As a matter of fact, the previous hypotheses have already manifested that the proactive robot has more natural behaviors (i.e., less intrusive and more reliable), is more trusted, is a better teammate and so possesses higher collaboration skills. Moreover, it contributes more to its human partner's success and to the efficiency of a task. We can already deduce that all of these are possible as a result of better adaptation skills of a cobot to its human partner. However, we still analyze the adaptation of the robot separately in this hypothesis for the sake of its importance in this thesis. Adaptation is a concept that is

hard to directly observe and evaluate in HRI in general, especially when the environment is not significantly changing to induce observable changes in human behaviors. In our case, the task difficulties and the environmental conditions remain the same for both of the robots throughout the whole experiment. Thus, the robots mostly observe similar responses from humans. In addition, the adaptation in our context requires reasoning on the hidden human states, which is hard to explicitly evaluate. The ground truth information, whether a robot has correctly anticipated and adapted to a human's needs and preferences, is completely hidden and only known to the interacted human. For that reason, we mostly make use of the subjective analysis in evaluating this hypothesis.

Table 6.9: Frequencies of general comparison questions asked to each of 14 participants at the end of the experiment. Robot type-1 is the proactive robot and Robot type-2 is the reactive robot, both of which are renamed for anonymity.

Levels	Counts	% of Total	Cumulative %
Robot type-1	4	28.6 %	28.6 %
Robot type-2	10	71.4 %	100.0 %

(a) "Which robot was more reactive (following preset rules) in its takeovers rather than being proactive and anticipate and respond to your behaviors?"

Levels	Counts	% of Total	Cumulative %
Robot type-1	11	78.6 %	78.6 %
Robot type-2	3	21.4 %	100.0 %

(b) "Which robot was learning and adapting better to your assistance needs?"

The participants stated that the proactive robot was able to adapt to their assistance needs better than the reactive one (with $p = 0.004$ as shown in Figure 6.21g). As shown in the same figure, they added that the reactive robot was repetitive in its behaviors rather than showing adaptive responses to their changing behaviors ($\mu_{reactive} = 3.07$ and $\mu_{proactive} = 1.74$ with $p < 0.001$). They were able to assess that the proactive robot had been able to understand their assistance needs correctly and significantly, better than the reactive one, despite both robots were rated relatively high for this ($\mu_{reactive} = 3.00$ and $\mu_{proactive} = 3.69$ with $p = 0.006$). Finally, at the end of the experiments we asked the participants two technical comparison questions even though they are not from robotics

or any related background. During the experiments, we renamed the proactive robot "robot type-1" and the reactive robot "robot type-2" to keep them anonymous. For the first question, the vast majority of the participants assessed that the reactive robot was following preset rules instead of responding to their changing needs and preferences (71.4% of the participants picked the robot type-2 as shown in Table 6.10a). The second question examined which robot was learning and adapting better to the participant's assistance needs. The subjects voted for the proactive robot significantly more than the reactive one, once again showing a great consistency with the previous statements (78.6% voted for the proactive robot in Table 6.10b). That said, the results support **Hypothesis 9**, which favors the adaptation skills of our proactive robot.

Concluding Remarks: We show in **Hypothesis 4** that during a task with a high load and/or after long working hours, a human may exert some unanticipated behaviors, e.g., in our case, getting harder to remember the task (placement) rules, losing attention during the task and getting exhausted (see Figure 6.13). These behaviors have also been observed as a result of the changing collaborativeness of humans. For instance, despite their bad performances, some participants still did not want the robot to take over the task in many cases (see Figure 6.22a). Such dynamic human behaviors have affected the participants' performance, needs, and preferences during a collaboration. Hence, we prove that a robot should effectively anticipate and adapt to these conditions (recall the unanticipated human conditions in Section 1.2.1). The challenge here is that such information is not always directly observable to anticipate; however, we could still show their importance through our novel experiment setup.

During the experiments, the reactive robot assumed that a human always accepts a robot's assistance offer and a human is always motivated and careful, discarding the unanticipated behaviors. This has led to reactive takeovers in many cases, e.g., after a human fails in a subtask. The proactive robot, on the other hand, observed the human longer before jumping into a conclusion, and stochastically estimated their need for help carefully assessing the current situation. The estimation of such unanticipated behaviors

has helped the robot to reach to more fine-grained conclusions about a human, such as: "the human may still be assessing the subtask (after a very long wait), as they have mostly succeeded so far", "the human may have forgotten the rules (e.g. after some failed placements); however, they may not want my help as they have rejected before", "the human may need my help as they have failed several times in a row and on this subtask they have idled very long". As our robot decision model is a POMDP, it is not learning from the history of interaction. Such conclusions are reached due to the probabilistic distributions over our state machine design with multiple anticipation stages (see Figure 6.9). For example, a person who has succeeded in 4 subtasks in a row would more likely to be estimated as "Human is doing ok", even if the human fails in the next iteration. This gives more time and room for the human to try to succeed again in the following subtask instead of the robot reactively taking over. If, for instance, the human in fact forgot the rest of the rules, her longer waiting times or her further failures led our robot to estimate her need for assistance eventually.

In conclusion, in a task where a contribution from a human collaborator is also expected and crucial, a robot collaborator should be aware of such unanticipated human behaviors and adapt to them. This is a very important but missing adaptation skill in most of the current HRC studies. Even though both robots contributed significantly to the overall success rate of a task when compared to a human working alone, handling such behaviors has also led to a more human contribution to the success of a task. We also show that this extended adaptation has contributed positively to the efficiency and the naturalness of a collaboration (through **Hypothesis 6** and **Hypothesis 7**). The proactive robot acted more naturally by showing better and wider anticipation and adaptation, as also perceived by the participants. Finally, we prove that our anticipatory robot model is more trustworthy, has more positive teammate traits, and has better collaboration skills (under **Hypothesis 8**). All in all, such positive traits are shown to be a result of a wider, more reliable, and more natural anticipation and adaptation provided by our A-POMDP robot model design (through **Hypothesis 9**). With this experiment and analysis, we validate our implementations in Chapter 4, i.e., our robot's extended short-term adaptation skills.

6.3.3 Experiment-2: Validating Chapter 5, ABPS In Collaboration

In this experiment, our goal is to support the hypothesis below:

Hypothesis 10. *Our integrated system with A-POMDP models and the ABPS mechanism provide a 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.*

This experiment involves the performance tests of our complete framework, FABRIC, also integrating ABPS. The process needs a policy library including many A-POMDP models. For that, we randomize models moving from our base model to obtain more reliable policies, as done in Chapter 5 in the simulation experiments (see in Section 5.2.2.1). The base model is the one used in *Experiment-1*, which has proven to be natural, reliable, and efficiently managing short-term human adaptation. For the training phase, we run the policy library against several human models in the simulation (as in Section 5.3.1.2). The policy generated after solving for the base model turns out to be the best policy of the library in hindsight. In *Experiment-1*, this model alone has provided a sufficient performance to prove its effectiveness of covering unanticipated human behaviors in a challenging work environment. However, we have observed very dynamic characteristics and preferences in the participants. Selecting different policies (i.e., strategies) would provide even better adaptation than a single policy with a limited adaptation (i.e. only during a task, which we call short-term).

In this experiment, our first goal is to show that the ABPS mechanism is able to provide this adaptation fast and reliably, improving a robot's collaboration performance. Also, we want to prove the effectiveness and applicability of our complete system in a real-world scenario. We evaluate the long-term adaptation capability of the system by highlighting the long-term differences in human behaviors and the robot's ability to detect and respond to such changes. However, we leave the analysis on the contribution of ABPS on a collaboration performance in real world, when compared to the best policy in hindsight running alone (as in Section 5.3), as future work. This is because the number

of experiments we could conduct are limited and testing both of the scenarios would require a longer-term and a bigger scale study[†]. Hence, we do not have significant data for such a comparison. Yet, our results still hint at improvements on the performance with ABPS since it leads to an increase in the task efficiency over time in all of the cases.

6.3.3.1 Objective Measures

We use some of the objective measures that were used in *Experiment-1* (see in Section 6.3.2.1). These 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 (from Equation 6.3) and the task efficiency (from Equation 6.5). To prove the adaptation capability of the robot and a human's changing behaviors in the long-term (i.e., over the course of a collaboration), we also analyze their changes over task assignments in time.

6.3.3.2 Subjective Measures

The subjective measures are obtained from the questionnaires that are asked to the participants. We use the same survey in *Experiment-1* (see in Section 6.3.2.2), only with minor changes to evaluate the robot's long-term adaptation capability, i.e., the adaptation between the tasks. We ask the participants to evaluate their cognitive load and the robot's effect on it during the tasks to analyze their dynamics over the course of the experiment. This gives more insights on the dynamics of the participants' changing stamina, motivation, and perceived difficulty of the tasks. In addition, the participants are asked to evaluate the naturalness of the collaboration and the reliability of the robot. Finally, we also evaluate the participants' trust in the robot and their perceived collaboration and teammate traits of the robot. Table 6.10 demonstrates the statements that are asked after each task completion. As always, they are categorized according to our intended evaluations, i.e., the ones with the "task-specific" label are asked after each task, and the ones with "general" are evaluated only once at the end of the experiment.

[†]The experiments in this section were interrupted by the COVID-19 pandemic outbreak

Table 6.10: Subjective statements asked during *Experiment-2*.

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

6.3.3.3 Experiment Protocol

We have invited 5 people who had no interaction with a robot before and are from different backgrounds and demographics (ages between 22 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 *Experiment-1* (see in Section 6.3.2.3). The only difference is that we do not run different robot models as in the previous experiment but only the complete framework. However, the ABPS mechanism may select a different policy for a task; therefore, the participants compare the robot performance between the tasks (i.e., the long-term adaptation as we call), without knowing its strategy may change.

Each participant works on 8 tasks in a row and fills out a survey evaluating the performance of a task right after it is completed. Then, at the of the experiment they answer another survey for the general statements (in Table 6.10). We select to work on task type-5 (as shown in Figure 6.4e) that implements the Stroop effect (Stroop, 1935). This task has been perceived to be the most difficult one before. The reason for this selection is to be able to invoke changing human characteristics like expertise during our experiments, which normally require more practices. The participants did not know the

tasks have this effect, and it took 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.

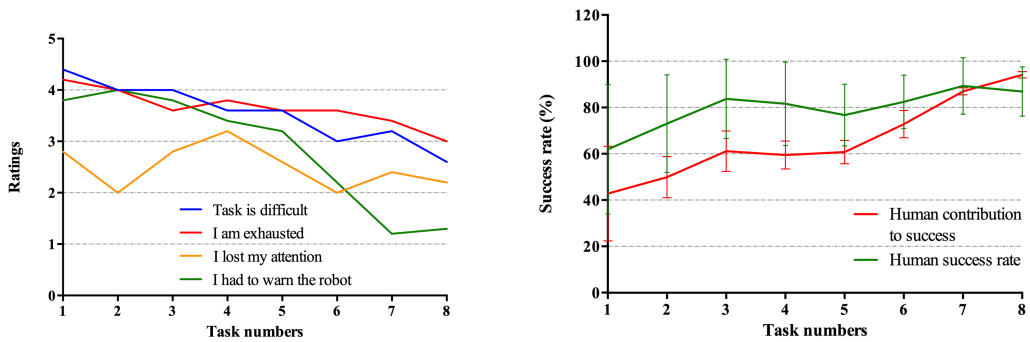
6.3.3.4 Results

Our goal is to support **Hypothesis 10**. To begin with, we analyze how and if a participant's perceived cognitive load, expertise, and trust change over time during the experiments (i.e, as we call the long-term behaviors). Figure 6.23a shows the subjective evaluations of the participants on their perceived task difficulty, attention, exhaustion, collaborativeness, and their trust in the robots on average. 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 the course of an experiment. This can be explained from Figure 6.23b visualizing the change in human success rates and the contribution of a human to the overall success in a task (see its definition 6.3). We deduce that the participants gained more expertise and got used to the task and the environment in time, which points out the practice-effect and so explains the decrease in their perceived difficulty of and the exhaustion in a task (i.e., the cognitive load). A human, on average, has contributed more to a task success in time, achieving more than 90% of all successful placements at the last task (see Figure 6.23b). Additionally, their success rates in each placement also increase, which is in line with their increasing successful contribution, all pointing to their increased expertise. Hence, they have taken over the tasks more often over time.

The collaborativeness of the participants has also changed during the experiments. That is, Figure 6.23a demonstrates that the participants warned the robot less in time whereas Figure 6.25a shows that their trust in the robot increased. That said, their expertise (i.e., through their success rates), collaborativeness (i.e., through their trust in the robot and the number of robot warnings), stamina (i.e., through their feeling of exhaustion), and motivation (i.e., through their handling of more subtasks) do change in time and that a robot should adapt to them. That said, we conclude that our assumption

on the change of human characteristics is valid and that we could invoke and observe this during our experiments. For example, the participants were more motivated to complete a task themselves due to their increased expertise toward the end of the experiments. In that case, the robot should give more time and a chance for the human to finalize a task than at the beginning of the collaboration. As opposed to that, some of the participants built up more trust in the robot resulting in leaving a task more often to the robot. Such drastic and stochastic behavioral changes are very difficult to model in a single decision-making strategy. Hence, the ABPS mechanism picked different strategies (i.e., policies in our case) to be able to handle such changes.

ABPS always selects the best policy in the policy library, which has collected the most rewards during the training phase. This, as mentioned, turned out to be a policy generated from the base A-POMDP model that is used throughout *Experiment-1*. In Figure 6.24, we demonstrate how ABPS has responded to such changes in human characteristics by selecting different policies. Figure 6.24a gives the total discounted rewards the robot has collected in 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 third task starts, ABPS has picked another policy



(a) Human's perceived cognitive load and trust.

(b) Human's task success rates and contributions.

Figure 6.23: 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.

in 60% of the experiments, i.e., for 3 participants out of 5 in total, according to its current estimate of that human type. With that, at the end of the third task the collected rewards on average have been almost doubled compared to the second task. For this particular case, if we examine a human's average contribution in Figure 6.23b, we see that it has increased only less than 20% on this task, meaning that the rest of the positive effect on the robot's reward, which is the combination of subtask success and the number of received warnings, has resulted from this policy change by ABPS. As such, the number of warnings received has also dropped around 33% from the second task to the third one (in Figure 6.24d). This shows the significant positive effect of the ABPS mechanism.

Looking at all plots in Figure 6.24, at 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 robot 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 6.25b and their contribution to success in Figure 6.23b after the third task. This adaptation is mutual, where both the human and the robot have finally achieved higher-levels of coordination after the third task. We can see the same effect also on the task efficiency in Figure 6.24e, the warning counts, and the time a task took in Figure 6.24f, all of which have also almost reached to a steady-state after the fourth task. We note that ABPS has changed policies also toward the end of the experiments (see at the seventh and the eight tasks with 20% changing rates in Figure 6.24a). This was a necessary adaptation by the robot in response to the overall increasing expertise of the participants (see the increasing human contributions after the fifth task in Figure 6.23b). Even though ABPS has only responded after the sixth task to this change, it could select strategies where the robot has mostly left the tasks to the participants and take over the tasks only if the human failed constantly or idled for longer times. The latter was often observed after the participants discovered how the robot takes over and so they were intentionally waiting for the robot to pick up,

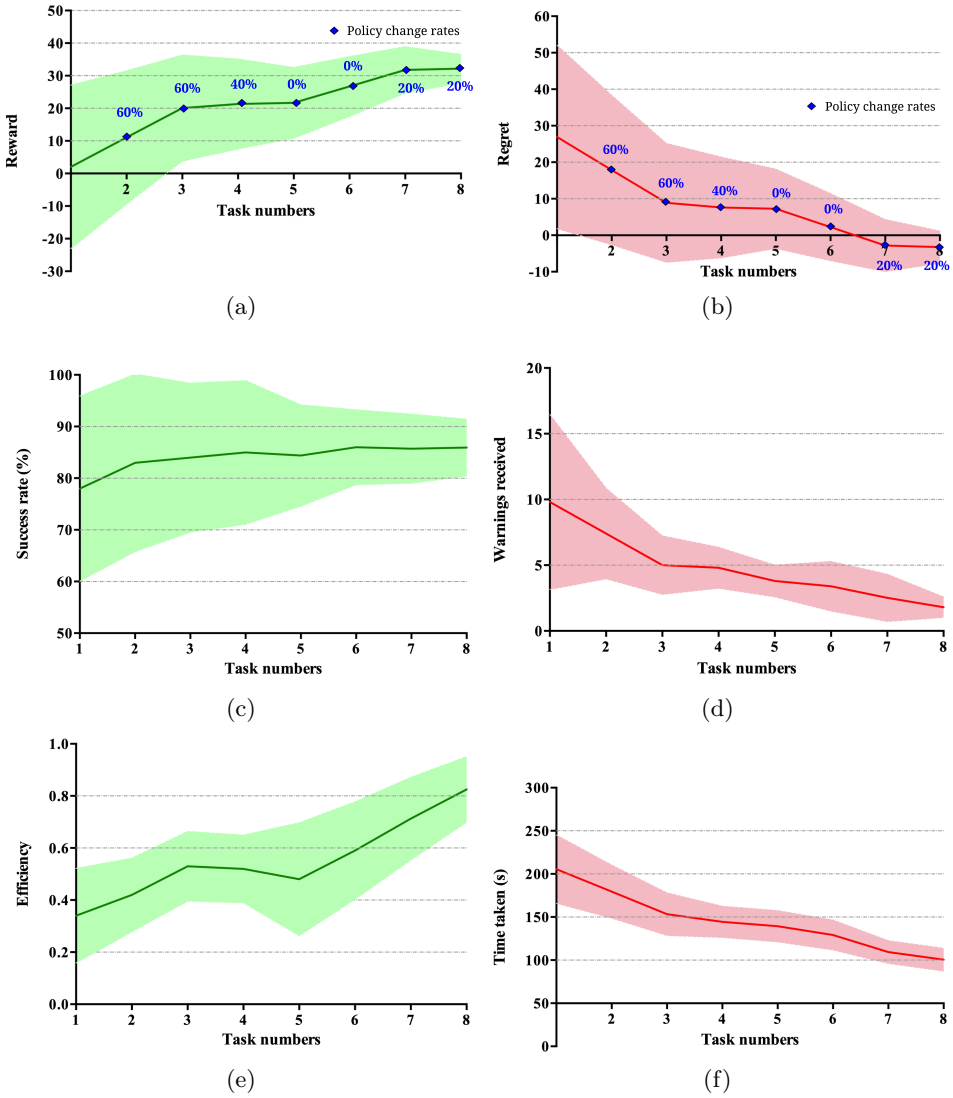
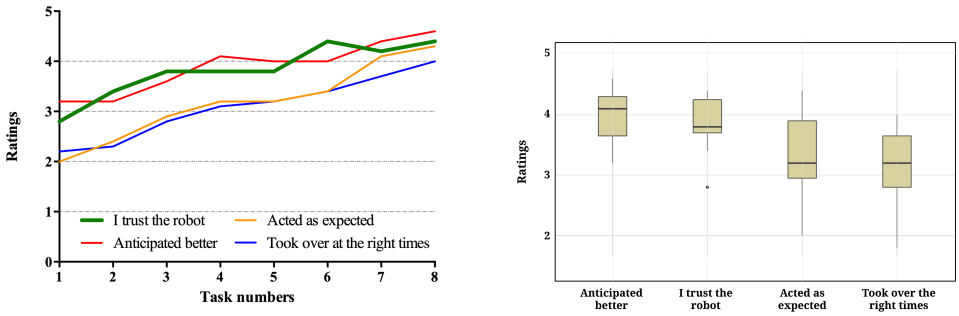


Figure 6.24: The first two plots give the performance of ABPS: a) moving average total discounted reward over task assignments collected by the robot; b) moving average regret (as in Equation 5.5) over task assignments with error bars denoting the standard deviation, collected by the robot. The policy change rates on these two figures denote the average rate of the robot 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 robot averaged over the task assignments: c) moving average of the task success rate; d) number of warnings received by the robot; e) moving average of the task efficiency as calculated in Equation 6.5; f) moving average of the task durations in seconds.

which is another example of the mutual adaptation. All these favored constant drops in the warning amounts, at the end almost reaching to the level of 0.



(a) (b)

Figure 6.25: The subjective evaluations of the participants in 5-step Likert scale 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.

Finally, we show the performance of ABPS also through the change of the regret values in time (Figure 6.24b). The regret is calculated from Equation 5.5, which denotes the distance of a total discounted reward collected in a task from the maximum utility that can be obtained by the best policy for the current human type. However, since the actual human types are unknown to the robot and to us, for this calculation we approximate the regret by taking the average of the maximum utilities collected by the policies in the policy library during the training phase run on the simulation environment (see in Section 6.2.4.2). Hence, we observe negative values of regret in Figure 6.24b since the rewards collected at these tasks were more than this assumed maximum utility value. Overall, as with the reward changes, the moving average regret has reached an almost steady-state after the third task and the robot was able to keep it at the lower levels, without any further increase. Then, thanks to the increasing team success with the human expertise, the robot's successful strategy change, and so the constant decrease in the amounts of warnings from the participants, the regret values have further decreased after the fifth task. All in all, drastic human characteristic changes have been observed

between the 1st and the 3rd tasks and between the 5th and the 7th tasks, and we prove that ABPS has gradually adapted to them despite their constant change throughout the experiments. We add that the response of ABPS is considerably fast but it could be faster with a more accurate human type estimation. This is also possible by allowing the policy selection even during the task, which requires a difficult process of configuring a stochastic decision-making system from any state with the inherited parameters from the previous policy. Also, we prove 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 6.25.

Table 6.11: Average ratings of the participant answers to the general statements asked at the end of each experiment.

	General Questions	Ratings
Mean (likert)	Robot showed positive teammate traits	4.6
	Coordination with robot got worse	1.8
	I became more comfortable	4.6
	Robot was able to adapt	4.6
	I would work with such a robot	4.4
True/False	Robot was learning and adapting	Agreed: 100 %
	Robot showed a high-level coordination	Agreed: 80 %

In addition to the performance of ABPS, we also evaluate the collaboration skills and teammate traits of our robot perceived by the participants, and their trust in the robot. In general, the better anticipation of a person's assistance needs, the right timings in assisting, the expected and reliable robot responses, and an increasing trust suggest a higher acceptance for a robot over the course of the experiments, which are also the case in Figure 6.25b. We see in this figure that, on average, the participants were satisfied with the performance of the robot throughout the experiment with the constantly increasing ratings of these given statements, reaching up to the levels of 65% to 75% of the maximum rating (in Figure 6.25a). In particular, the trust level has reached almost 90% of the maximum rating at the end of the experiments. Our robot could provide positive

teammate traits through anticipating and respecting its partner's needs and preferences, contributing to an increased performance of its partner and leading to a more efficient task. Supporting this, the participants evaluated that the robot showed positive teammate traits, the robot was able to adapt to the changing human needs and preferences, they became more comfortable with the robot (with the mean ratings of 4.6 for these three statements) and they would work with such a robot (with the mean rating of 4.4), as given in Table 6.11. In addition, all of the participants agreed that the robot was learning and adapting, and 80% of them confirmed that the robot has high coordination skills. In conclusion, all these objective measures and the positive subjective statements of the participants support **Hypothesis 10**.

Finally, even though the comparative analysis on our framework with and without ABPS to prove the necessity of such a policy selection running on top of intention-aware models for long-term adaptation are left as a future work, we could still show the positive effect of ABPS during the experiments. Given that ABPS always starts the collaboration with the proactive model (i.e., A-POMDP base model) from *Experiment-1*, the increasing efficiency and naturalness of collaboration after selecting different policies hints at the positive contribution of such a mechanism running on top of the human intention-aware models.

6.4 HRC Simulation in a Factory Environment

6.4.1 Methodology

Similar to the real setup described in Section 6.2, here we detail the dynamics and the architecture of the simulation environment, i.e., a 3D factory environment, and the simulated human models, which are used for training our robot models and for our early evaluations in Chapter 4 and Chapter 5. The simulation consists of existing human and robot models (PR2 is selected but we are indifferent to the robot hardware), a conveyor system, 3D printer, produced packages, two containers for processed products, a container for unprocessed products and a restroom (see Figures 6.26. As we mention in

Figure 6.26: Simulation of an HRC at a conveyor belt for the task of product inspection and storing.

the previous chapters, all of our scenarios consist of several sequential task assignments to simulate long-term collaboration. A task in our case is a product inspection and storing job. It starts with a user-defined task assignment, e.g., based on the product's weight and fragility. 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, or else it runs and the product falls into the uninspected-product container leading to a task failure.

We implement our robotic framework in Figure 3.1 in ROS and use our simulation environment developed under the MORSE environment. The deployed Multi-Agent System (MAS) architecture for simulations is given in Figure 6.27, which also resembles our real setup architecture in Figure 6.1. Both the robot and the human are controlled by their decision models whereas the generated action decisions are executed in MORSE. For the robot, the decision-making agent runs our A-POMDP models and the ABPS mechanism (see Algorithm 1), but it is possible to replace it with a different strategy. The observations the robot receives are the 3D human body joints that are always avail-

able 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 existing studies, e.g., (Roitberg et al., 2014), has been implemented to recognize the constrained and distinct simulated human gestures from the body joints available.

We have designed certain human actions, such as, looking around, grasping and placing the package, warning the robot, etc. that are required for our scenario (see the human actions in the experiments in Section 4.3.1). However, this list can be extended accordingly. Based on the available list of actions, we have designed human models that simulate a variety of human decisions. Just like the robot, the decisions are executed in MORSE, and the human decision model receives observations for the current state of the environment as feedback. In addition, we develop a remote control interface to allow the control of the simulated human manually, replacing the automated decision models. Our initial idea was to conduct user studies through this interface; however, we aim for a real collaboration environment in order not to constrain any unanticipated human behaviors. In summary, this environment allows for a fully automated long-term HRC to train and test robot decision-making solutions under various conditions.

6.4.1.1 Human Simulation: Human Decision Models

Simulating humans allow us to scale the experiments to emulate many different combinations of human behaviors, including unanticipated ones, to rigorously test our robot decision-making solutions (see our motivation in Section 1.1.3). 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, getting tired fast, being distracted easily, building trust or distrust in the robot. A representative proof of concept human decision model is built using an MDP as shown in Figure 6.28. We note that only the transitions with non-negligible probabilities are shown in the figure.

In our model design we are inspired from the available studies analyzing human workers operating on repeated tedious tasks in a workplace (Gombolay et al., 2017; Ji, Lan, and

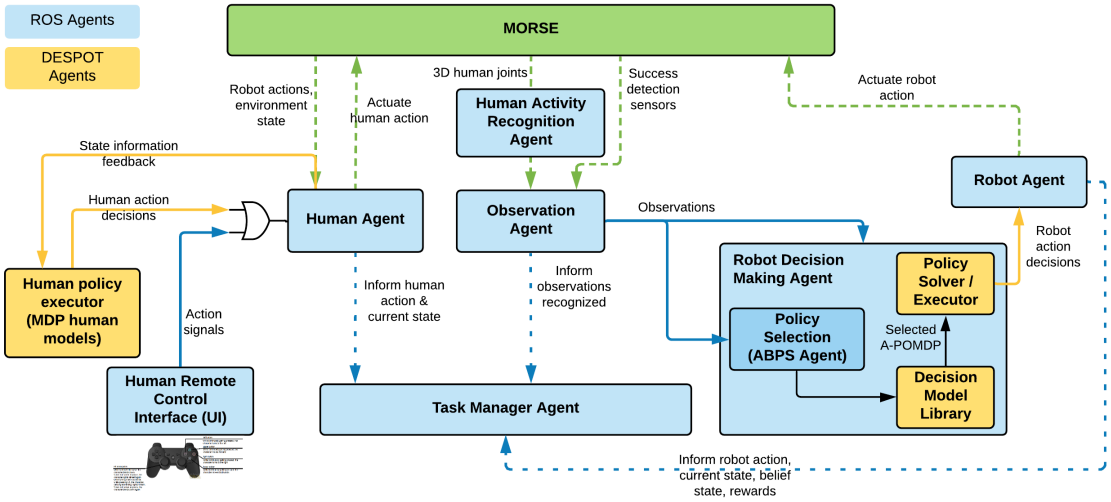


Figure 6.27: Our HRC simulation MAS architecture

Looney, 2006; McGuire et al., 2018). 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 towards their goal may be imperfect (Hiatt et al., 2017). Simulating such a human has been shown to be accurate using an MDP to generate a policy for a human agent (Bandyopadhyay et al., 2013).

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. Another example would be that a beginner and non-collaborative human model receives positive rewards when the human cannot handle the task and each time the human warns the robot when the robot interferes with the task. The model is inspired by our expectations that a human chooses an action based on the collaborating 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

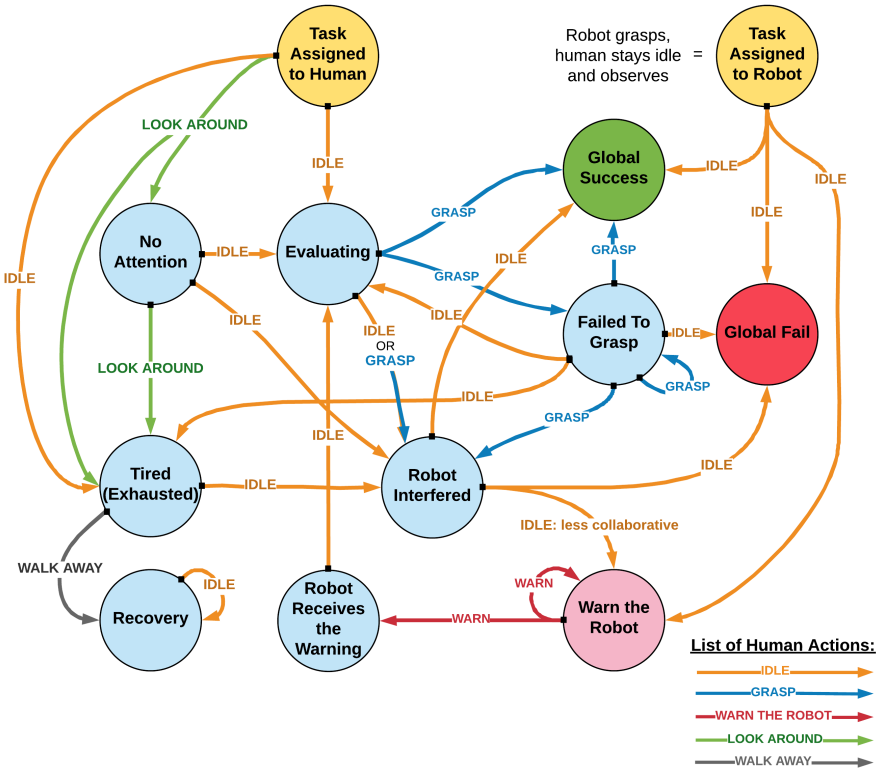


Figure 6.28: Human MDP models in detail, showing state-action connections for the most probable transitions.

$s, s' \in S$, $a \in A$, 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, i.e., short-term changing behaviors. 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), e.g., *a beginner, tired and collaborative human*, a *mid-expert, high-stamina, and less-collaborative human.* This modeling scheme is then used to automate the testing of our robot reasoning under various hard-to-predict conditions. We run analysis on the reliability of our modeling scheme and the diversity of behaviors toward large scale training and tests in Section 6.4.2.

6.4.2 Validating the Simulated Human Models

In this section, we validate the reliability of the simulated human models used throughout the experiments in Chapter 4 and Chapter 5 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 (see in Section 6.4.1.1). 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 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 6.28). 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 *Experiment-1* (i.e., from 14 participants). As we discuss 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 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 to 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 (see in Section 6.2.3.2). 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 6.2.3.1), 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

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)

Figure 6.29: The likelihoods of: a) the simulated observations generated from the same simulated models; b) the participant observations generated from the simulated models.

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 6.29a, whereas in Figure 6.29b, 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 values that can actually be obtained given the randomness and all possible state transitions in the simulated humans.

Figure 6.29a shows the distribution of these values over each model. The biggest 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 the random walk we implement is not that excessive as the generated actions are able to show the intrinsic characteristics of their models. Yet the values are comparatively 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 6.29b, 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 amount of the human observations are very less compared to the simulated ones. Finally, as it is visualized in Figure 6.29b, 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 cobots. Our simulated models, specifically, Model-4,-5 and -8 in our case, contribute greatly to that diversity.

6.5 Conclusion

Within the scope of this chapter, we prepare an evaluation environment that validates our anticipatory cobot more realistically by introducing a rather unconstrained environment that invokes unanticipated human behaviors that are mostly overlooked in the

existing HRC solutions. For this purpose, we follow the conventional method of evaluation in robotic solutions by first designing a novel simulated human model and a 3D factory environment that samples such dynamic human behaviors, including the ones causing human errors. The simulation environment has already been used in Chapter 4 and Chapter 5 to train on large scale data and to be rigorously tested facing relatively difficult to anticipate human conditions, as we call the unanticipated behaviors. We then transfer the simulation results into the real world through our novel experiment setup and the collaboration task that together induce cognitive load on humans, in the end invoking such unanticipated behaviors. Our results show that we could observe a variety of human responses and collaboration preferences during the experiments. The user studies successfully validate our findings in Chapter 4, our A-POMDP with its short-term adaptation skills. The studies also show that our simulated human models are capable of providing realistic human behaviors with even more diversity than the user studies thanks to their scalable behavior sampling. Hence, we prove the reliability of our simulation environment and its necessity in training more accurate decision-making models and their rigorous tests before their real human collaborations. Finally, we integrate our complete framework, including the ABPS mechanism from Chapter 5, as a human-in-the-loop autonomous system. We show that our framework can operate in such a dynamic environment, it is applicable and it is able to provide a long- and short-term adaptation to various humans.

7 Concluding Remarks

7.1 Summary

In this thesis, we focus on an extended human adaptation of collaborative robots (cobots) during a human-robot collaboration (HRC). An efficient HRC is only possible when robots anticipate a human partner's state of mind, including their plans and goals (intentions). Such anticipated knowledge should then be used in a robot's autonomous decision-making to adapt to reach high-level coordination, i.e., a fluent collaboration. Current research on anticipatory robots has to limit a human's intention space due to model complexities, in the end, overlooking many possible human contingencies including their erroneous behaviors. We call the collection of such behaviors "unanticipated" human behaviors. Our premise is to investigate the extent of human adaptation a robot can/should reach toward their fluent collaboration. In doing so, we formulate the entire thesis around three main topics: *short-term adaptation, long-term adaptation,* and *evaluation of adaptation.*

We ask the following research questions toward the three topics: *1)* "How can a robot anticipate and adapt to unanticipated human behaviors, including a human's short-term changing intent (motivation), availability, capability and willingness to collaborate?"; *2)* "How can a robot handle a diversity of long-term human characteristics, which may eventually affect a human's short-term behaviors?"; *3)* "How can we develop reliable human models that simulate a great diversity of human behaviors including their unanticipated behaviors in work environments?"; *4)* "How can we design a real collaboration setup and a task without making assumptions on human intentions and, as a result, this invokes unanticipated human behaviors in order to properly evaluate our robot's short-

and long-term adaptation goals?". With the purpose of obtaining an integrated system, in Chapter 3 we devise our novel lightweight autonomous framework, called FABRIC, that hierarchically integrates our approaches to each question above. In the cognitive level, the robot adapts to a human's short-term changes (e.g. a tough day at work for the human) and at the meta-cognitive level, a human's long-term characteristics are handled.

In Chapter 4, we start with the short-term adaptation and we answer the Research Question 1. We examine the effects of a robot's anticipation and response to such unanticipated human behaviors, which may be observed in a collaboration with repeated and tedious tasks. Our approach is our novel anticipatory partially observable Markov decision process (A-POMDP), that handles the adaptation to the short-term changing human state of mind in two stages. In the first stage, the planner incorporates a human's changing *intent (motivation), availability, and capability*, e.g., lost attention, getting tired. Then, in the second stage, through these first anticipated states, it tries to estimate if the human needs help and whether the robot should intervene. While doing so, A-POMDP goal is to increase the efficiency of the collaboration, ensuring the safety and the autonomy of the human partner. In this chapter, as a proof of concept of our A-POMDP, we use simulated human models in the experiments, which provide a more robust setup to train and test such various uncommon and unanticipated conditions. Many features are assumed random in the simulation, such as real humans' trust in the robot may build up over the course of the collaboration, thereby creating a more difficult experimental case. In a collaboration task of picking and placing various products running on a conveyor belt, we have shown that anticipating and taking into account such human variability increases the overall efficiency (increased success rate over a shorter time) and the naturalness (fewer warnings received from the human, hence less intrusive robot behaviors) of an HRC. The promising results have encouraged us to move toward validating them with real human experiments.

The next chapter, Chapter 5, focuses on long-term adaptation and answers the Research Question 2. We introduce our novel anticipatory Bayesian policy selection (ABPS) as a complementary solution to the existing intention-aware robot decision-making mod-

els, e.g., our A-POMDP models. Such models provide a limited adaptation since they get computationally expensive and less accurate as a wider variety of human behaviors are modeled, such as in long-term interactions involving multiple people with various characteristics and preferences. We examine the effects of our ABPS on a collaborative robot's adaptation to unknown human types and their long-term workplace characteristics, in our case, their level of *expertise, stamina (or fatigue), attention and collaborativeness.* Our results have shown that ABPS is a fast and reliable policy selection mechanism for HRC scenarios. Having such a mechanism on top of a robot's intention-aware decision-making has contributed positively to the efficiency and naturalness of the collaboration by providing a better adaptation, i.e., extended with long-term adaptation, to the collaborating human, when compared to the state-of-the-art robot decision-making models running alone. In the experiments, we have integrated the complete architecture we describe in Chapter 3, capable of running ABPS on top of our decision-making models. As a proof of concept for ABPS, we run the experiments on the same simulation setup with extended human variety. These early validations have encouraged us to deploy the full system in the real world for real human feedback.

Chapter 6 provides solutions for the Research Questions 3 and 4 toward the final evaluation of our robot's adaptation skills. In HRI evaluations in general, the diversity of human behaviors the robots face is often limited due to constrained environments and/or rather short-term interactions. Toward evaluating our ideas and solutions, we need to observe a greater diversity of human behaviors, including the unanticipated ones possibly leading to human errors or inefficient collaboration if not handled by the robot. For this purpose, we follow the conventional way of evaluation in robotics, i.e., training and running rigorous tests in simulation, then deploying the solution in the real world. We first design a novel simulated human model and a 3D factory environment that samples such dynamic human behaviors. The simulation environment has already been used in Chapter 4 and Chapter 5 to train on a large scale data and to test our solutions under relatively harsh human conditions. We are aware of the possible biases in the experiments which could be introduced by the simulated humans. However, the

abstracted states in our design are shown to be observed in real humans through the case studies in this chapter. On the other hand, the simulation provides even more diversity than the user studies thanks to its scalable behavior sampling. This and the case studies 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 consider an HRC scenario at a conveyor belt for the task of sorting products according to their colors and placing them on different colored containers. The cognitive load is induced through the confusing color matchings and the long experiment hours. Our case studies show that we could observe a variety of human responses and collaboration preferences during the experiments, including human behaviors that lead to failures and mistakes. After that, we conduct our first user study that successfully validates the results in Chapter 4; hence, we prove our A-POMDP model's effectiveness and naturality in providing a better short-term adaptation. Through this experiment, we also show the reliability of our human simulations and their necessity in training more accurate anticipatory decision-making models. Finally, we integrate our complete framework, including the ABPS module from Chapter 5, as a human-in-the-loop autonomous system. We improve the estimation and decision models by training with real human data along with the ones obtained from the simulation. Through another round of user studies, we show that our complete framework with ABPS is able to operate in such a dynamic environment and it is applicable and able to provide an extended long- and short-term adaptation to various humans, and so validating our findings in Chapter 5. In summary, through our integrated solution, FABRIC, we provide significant improvements in a collaborative robot's human adaptation toward a more reliable, natural, and efficient collaboration, contributing to the long-term use of collaborative robots.

7.2 Limitations and Future Works

We discuss, in general, the lack of human diversity considered in the existing anticipatory collaborative robots and their applications in HRC. For that, we provide solutions for

extended human adaptation. Even though we use simulated humans to generate more diversity of human behaviors, in addition to the real human observations collected, our human type estimation model was not accurate for certain cases. For instance, from the observations of a human idling longer, our system cannot accurately estimate whether the human is tired or she is a beginner and cannot handle the task. This is mainly because we assume a limited human type and action space that defines any human as a distribution of some known workplace characteristics and behaviors. Our simulated humans are also modeled with the same assumption; hence, some simulation runs have generated repeating patterns in the long-term. In this study, since our goal was to highlight the importance of handling unanticipated human behaviors, not being able to differentiate a tired human from a beginner one has not produced a significant impact as the robot has responded similarly to both cases through offering more assistance. However, during the user studies, we found that an accurate estimation and more tailored handling of such unanticipated human behaviors would result in even more efficient collaboration. In general, we present ways of how to generate and observe various dynamics of human behaviors through our novel experiment setup and human simulations. In the future, we believe that more data should be collected from bigger scale long-term user studies to better model the long-term traits of humans and to generate more accurate human simulations. Such data and simulations will help the HRC community and the other applications involving human-aware planning to build more reliable interactive robots toward their long-term acceptance.

Furthermore, during the user studies, some of the participants gained enough expertise over time that they were perfectly capable of achieving the tasks alone. In such cases, they stated that they could do better without the robot as it was distracting them. In the short-term experiments, we showed our robot's positive contribution to such challenging tasks when compared to a human working without the robot. However, long-term human alone experiments are also needed on the same task so that when we compare it with our robot's collaboration, we can better analyze the robot's contribution when human characteristics, e.g., expertise, change in the long-term. Two participants also added that

calling for robot assistance only when they want to could be more efficient when compared to our robot autonomously deciding it. For that, we need experiments on such a command and control system to compare its efficiency with our system in the long-term. After that, we can also analyze the additional cognitive load such a command system may put on the operators. Finally, we can further improve our system by developing a policy change mechanism even during a task, which requires interrupting and running Markov decision processes other than their start and terminate states with the inherited parameters from previous strategies. This will decrease the adaptation times of the system contributing to fast responses. 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, the improved system can be deployed on a larger scale industrial setup to prove its applicability with the real human workers for a broader impact.

Bibliography

Albrecht, Stefano V. and Peter Stone (2018). "Autonomous agents modelling other agents: A comprehensive survey and open problems". In: *Artificial Intelligence* 258, pp. 66–95. ISSN: 0004-3702.

Andriella, Antonio, Carme Torras, and Guillem Alenyà (2019). "Short-Term Human–Robot Interaction Adaptability in Real-World Environments". In: *International Journal of Social Robotics*. ISSN: 1875-4805.

Askarpour, Mehrnoosh, Dino Mandrioli, Matteo Rossi, and Federico Vicentini (2017). "Modeling Operator Behavior in the Safety Analysis of Collaborative Robotic Applications". In: *Computer Safety, Reliability, and Security*. Ed. by Stefano Tonetta, Erwin Schoitsch, and Friedemann Bitsch. Cham: Springer International Publishing, pp. 89–104. ISBN: 978-3-319-66266-4.

– (2019). "Formal model of human erroneous behavior for safety analysis in collaborative robotics". In: *Robotics and Computer-Integrated Manufacturing* 57, pp. 465–476.

Baker, Chris L. and Joshua B. Tenenbaum (2014). "Modeling Human Plan Recognition using Bayesian Theory of Mind". In: *Plan, Activity, and Intent Recognition: Theory and Practice*, pp. 177–204.

Baker, Simon, Iain Matthews, Jing Xiao, Ralph Gross, Takahiro Ishikawa, and Takeo Kanade (2004). *Real-Time Non-Rigid Driver Head Tracking for Driver Mental State Estimation*. Tech. rep. CMU-RI-TR-04-10. Pittsburgh, PA: Carnegie Mellon University.

Bandyopadhyay, Tirthankar, Kok Sung Won, Emilio Frazzoli, David Hsu, Wee Sun Lee, and Daniela Rus (2013). "Intention-Aware Motion Planning". In: *Algorithmic Foundations of Robotics X*. Ed. by Emilio Frazzoli, Tomas Lozano-Perez, Nicholas Roy, and Daniela Rus. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 475–491.

Bauer, Andrea, Dirk Wollherr, and Martin Buss (2008). "Human-robot collaboration: A survey". In: *International Journal of Humanoid Robotics* 05.01, pp. 47–66.

Baxter, Paul (2016). "Memory-Centred Cognitive Architectures for Robots Interacting Socially with Humans". In: *ArXiv* abs/1602.05638.

Berlin, Matt, Jesse Gray, Andrea L. Thomaz, and Cynthia Breazeal (2006). "Perspective Taking: An Organizing Principle for Learning in Human-Robot Interaction". In: *Proceedings of the 21st National Conference On Artificial Intelligence (AAAI)*, pp. 1444–1450.

Bodden, Christopher, Daniel Rakita, Bilge Mutlu, and Michael Gleicher (2016). "Evaluating intent-expressive robot arm motion". In: *2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. New York, NY, USA: IEEE, pp. 658–663.

Bodenhausen, Galen V. and Kurt Hugenberg (2011). "Attention, perception, and social cognition". In: *Social Cognition*. Taylor and Francis, pp. 1–22.

Breazeal, C., C. D. Kidd, A. L. Thomaz, G. Hoffman, and M. Berlin (2005). "Effects of nonverbal communication on efficiency and robustness in human-robot teamwork". In: *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 708–713.

Broz, Frank, Illah Nourbakhsh, and Reid Simmons (2011). "Designing POMDP models of socially situated tasks". In: *International Symposium on Robot and Human Interactive Communication, 2011 RO-MAN*, pp. 39–46.

– (2013). "Planning for Human-Robot Interaction in Socially Situated Tasks: The Impact of Representing Time and Intention". In: *International Journal of Social Robotics* 5.2, pp. 193–214.

Calisgan, E., 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: *2012 IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication*. Paris, France: IEEE, pp. 418–423.

Chang, M. Lee, R. A. Gutierrez, P. Khante, E. Schaertl Short, and A. Lockerd Thomaz (2018). "Effects of Integrated Intent Recognition and Communication on Human-Robot Collaboration". In: *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 3381–3386.

Chao, Crystal and Andrea L. Thomaz (2012). "Timing in Multimodal Turn-Taking Interactions: Control and Analysis Using Timed Petri Nets". In: *J. Hum.-Robot Interact.* 1.1, pp. 4–25.

Chen, Min, 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. Chicago, IL, USA: ACM, pp. 307–315.

Cheng, Shinko Y., Sangho Park, and Mohan M. Trivedi (2007). "Multi-Spectral and Multi-Perspective Video Arrays for Driver Body Tracking and Activity Analysis". In: *Comput. Vis. Image Underst.* 106.2–3, pp. 245–257. ISSN: 1077-3142.

Cherubini, Andrea, Robin Passama, André Crosnier, Antoine Lasnier, and Philippe Fraisse (2016). "Collaborative manufacturing with physical human–robot interaction". In: *Robotics and Computer-Integrated Manufacturing* 40, pp. 1–13. ISSN: 07365845.

Cohen, Philip R. and Hector Levesque (1991). "Teamwork". In: *Nous* 24.4, pp. 485–512.

Dautenhahn, Kerstin (2007). "Socially intelligent robots: dimensions of human-robot interaction". In: *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 362.1480, pp. 679–704.

Devin, Sandra 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)*, pp. 319–326.

Dragan, A. D., S. Bauman, J. Forlizzi, and S. S. Srinivasa (2015). "Effects of Robot Motion on Human-Robot Collaboration". In: *2015 10th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp. 51–58.

Dragan, Anca and Siddhartha Srinivasa (2013). "Generating Legible Motion". In: *Robotics: Science and Systems IX*. Robotics: Science and Systems Foundation.

Dragan, Anca D., Kenton C.T. Lee, and Siddhartha S. Srinivasa (2013). "Legibility and predictability of robot motion". In: *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. Tokyo, Japan: IEEE, pp. 301–308.

Garrido-Jurado, S., R. Muñoz-Salinas, F.J. Madrid-Cuevas, and M.J. Marín-Jiménez (2014). "Automatic generation and detection of highly reliable fiducial markers under occlusion". In: *Pattern Recognition* 47.6, pp. 2280–2292. ISSN: 00313203.

Gielniak, Michael J and Andrea L Thomaz (2011). "Anticipation in Robot Motion". In: *2011 RO-MAN*. IEEE, p. 7.

Gleeson, Brian, Karon MacLean, Amir Haddadi, Elizabeth Croft, and Javier Alcazar (2013). "Gestures for industry: Intuitive human-robot communication from human observation". In: *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. Tokyo, Japan: IEEE, pp. 349–356.

Gombolay, Matthew, Anna Bair, Cindy Huang, and Julie Shah (2017). "Computational Design of Mixed-initiative Human-robot Teaming That Considers Human Factors". In: *Int. J. Rob. Res.* 36.5-7, pp. 597–617. ISSN: 0278-3649.

Görür, O. Can and Şahin Albayrak (2016). "A Cognitive Architecture Incorporating Theory of Mind in Social Robots towards Their Personal Assistance at Home". In: *Workshop on Bio-inspired Social Robot Learning in Home Scenarios at IEEE/RSJ International Conference on Intelligent Robots and Systems 2016 (IROS'16)*. Daejeon, Korea.

Görür, O. Can and Aydan M. Erkmen (2014). "Elastic networks in reshaping human intentions by proactive social robot moves". In: *The 23rd IEEE International Symposium on Robot and Human Interactive Communication*, pp. 1012–1017.

– (2015). "Intention and Body-Mood Engineering via Proactive Robot Moves in HRI". In: *Handbook of Research on Synthesizing Human Emotion in Intelligent Systems and Robotics*. Ed. by Jordi Vallverdú. IGI Global, pp. 256–284.

Görür, O. Can, 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. Montreal QC, Canada: International Foundation for Autonomous Agents and Multiagent Systems, pp. 77–85.

Görür, O. Can, Benjamin Rosman, Guy Hoffman, and Şahin Albayrak (2017). "Toward Integrating Theory of Mind into Adaptive Decision-Making of Social Robots to Understand Human Intention". In: *Workshop on Intentions in HRI at ACM/IEEE International Conference on Human-Robot Interaction (HRI'17)*.

Görür, O. Can, 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. Chicago, IL, USA, pp. 398–406.

Green, Scott A., Mark Billinghurst, XiaoQi Chen, and J. Geoffrey Chase (2008). "Human-Robot Collaboration: A Literature Review and Augmented Reality Approach in Design". In: *International Journal of Advanced Robotic Systems* 5.1, p. 1.

Hart, Sandra G. and Lowell E. Staveland (1988). "Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research". In: *Advances in Psychology.* Vol. 52. Elsevier, pp. 139–183. ISBN: 978-0-444-70388-0.

Hiatt, Laura M, Cody Narber, Esube Bekele, Sangeet S Khemlani, and J Gregory Trafton (2017). "Human modeling for human-robot collaboration". In: *The International Journal of Robotics Research* 36.5-7, pp. 580–596.

Hiatt, Laura M., 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)*, pp. 2066–2071.

Hoffman, Guy (2019a). *Anki, Jibo, and Kuri: What We Can Learn from Social Robots That Didn't Make It.*

– (2019b). "Evaluating Fluency in Human-Robot Collaboration". In: *IEEE Transactions on Human-Machine Systems* 49.3, pp. 209–218.

Hoffman, Guy, Gurit E. Birnbaum, Keinan Vanunu, Omri Sass, and Harry T. Reis (2014). "Robot Responsiveness to Human Disclosure Affects Social Impression and Appeal". In: *Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interaction.* HRI '14. Bielefeld, Germany: Association for Computing Machinery, pp. 1–8. ISBN: 9781450326582.

Hoffman, Guy and Cynthia Breazeal (2004). "Collaboration in Human-Robot Teams". In: *AIAA 1st Intelligent Systems Technical Conference.* Chicago, IL, USA.

– (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, pp. 1–8.

– (2010). "Effects of anticipatory perceptual simulation on practiced human-robot tasks". In: *Autonomous Robots* 28.4, pp. 403–423. ISSN: 1573-7527.

Holladay, Rachel M., Anca D. Dragan, and Siddhartha S. Srinivasa (2014). "Legible robot pointing". In: *The 23rd IEEE International Symposium on Robot and Human Interactive Communication.* Edinburgh, UK: IEEE, pp. 217–223.

Holtzen, Steven, Yibiao Zhao, Tao Gao, and Song-chun Zhu (2016). "Inferring Human Intent from Video by Sampling Hierarchical Plans". In: *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'16)*, pp. 1489–1496.

Honig, Shanee and Tal Oron-Gilad (2018). "Understanding and Resolving Failures in Human-Robot Interaction: Literature Review and Model Development". In: *Frontiers in Psychology* 9, p. 861. ISSN: 1664-1078.

Horvath, Adam O. and Leslie S. Greenberg (1989). "Development and validation of the Working Alliance Inventory". In: *Journal of Counseling Psychology* 36.2, pp. 223–233.

Huang, Chien Ming 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)*, pp. 83–90.

Imai, M., T. Ono, and H. Ishiguro (2003). "Physical relation and expression: joint attention for human-robot interaction". In: *IEEE Transactions on Industrial Electronics* 50.4, pp. 636–643. ISSN: 0278-0046.

Irfan, B., A. Ramachandran, S. Spaulding, D. F. Glas, I. Leite, and K. L. Koay (2019). "Personalization in Long-Term Human-Robot Interaction". In: *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp. 685–686.

Ji, Qiang, Peilin Lan, and Carl Looney (2006). "A probabilistic framework for modeling and real-time monitoring human fatigue". In: *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans* 36, pp. 862–875.

Kamar, Ece, Ya'akov Gal, and Barbara J. Grosz (2009). "Incorporating Helpful Behavior into Collaborative Planning". In: *Proceedings of The 8th International Conference on Autonomous Agents and Multiagent Systems - Volume 2*. AAMAS '09. Budapest, Hungary, pp. 875–882. ISBN: 978-0-9817381-7-8.

Kargruber, Elia Jona (2019). "Toward User Studies in Real-World Settings: Application of an Autonomous Anticipatory Robotic System for Evaluating Human-Robot Collaboration on an Assembly Line". MA thesis. Berlin, Germany: Technische Universität Berlin.

Kidd, C. D. and C. Breazeal (2008). "Robots at home: Understanding long-term human-robot interaction". In: *2008 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 3230–3235.

Kirsch, Alexandra, Thibault Kruse, E. Akin Sisbot, Rachid Alami, Martin Lawitzky, Dražen Brščić, Sandra Hirche, Patrizia Basili, and Stefan Glasauer (2010). "Plan-Based Control of Joint Human-Robot Activities". In: *KI - Künstliche Intelligenz* 24.3, pp. 223–231.

Koppenborg, Markus, Peter Nickel, Birgit Naber, Andy Lungfiel, and Michael Huelke (2017). "Effects of movement speed and predictability in human-robot collaboration". In: *Human Factors and Ergonomics in Manufacturing & Service Industries* 27.4, pp. 197–209. ISSN: 10908471.

Koppula, Hema S., Ashesh Jain, and Ashutosh Saxena (2016). "Anticipatory Planning for Human-Robot Teams". In: *Experimental Robotics*. Ed. by M. Ani Hsieh, Oussama Khatib, and Vijay Kumar. Cham: Springer International Publishing, pp. 453–470.

Kuka (2016). *Hello Industrie 4.0 Glossary*. Tech. rep. Kuka AG.

Kurup, Unmesh and Christian Lebiere (2012). "What can cognitive architectures do for robotics?" In: *Biologically Inspired Cognitive Architectures* 2, pp. 88–99.

Lasota, Przemyslaw A., Terrence Fong, and Julie A. Shah (2017). "A Survey of Methods for Safe Human-Robot Interaction". In: *Foundations and Trends in Robotics* 5.4, pp. 261–349.

Lasota, Przemyslaw A. and Julie A. Shah (2015). "Analyzing the Effects of Human-Aware Motion Planning on Close-Proximity Human–Robot Collaboration". In: *Human Factors: The Journal of the Human Factors and Ergonomics Society* 57.1, pp. 21–33. ISSN: 0018-7208, 1547-8181.

Leite, Iolanda, Ginevra Castellano, André Pereira, Carlos Martinho, and Ana Paiva (2014). "Empathic Robots for Long-term Interaction". In: *International Journal of Social Robotics* 6.3. ISSN: 1875-4805.

Leite, Iolanda, Carlos Martinho, and Ana Paiva (2013). "Social Robots for Long-Term Interaction: A Survey". In: *International Journal of Social Robotics* 5.2, pp. 291–308.

Li, J., T. Lei, and F. Zhang (2018). "An Gaussian-Mixture Hidden Markov Models for Action Recognition Based on Key Frame". In: *2018 11th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, pp. 1–5.

Lichtenthäler, Christina and Alexandra Kirsch (2016). "Legibility of Robot Behavior : A Literature Review". working paper or preprint.

Liu, Hongyi and Lihui Wang (2018). "Gesture recognition for human-robot collaboration: A review". In: *International Journal of Industrial Ergonomics* 68, pp. 355–367. ISSN: 0169-8141.

Mainprice, Jim, Rafi Hayne, and Dmitry Berenson (2015). "Predicting human reaching motion in collaborative tasks using Inverse Optimal Control and iterative re-planning". In: *2015 IEEE International Conference on Robotics and Automation (ICRA)*. Seattle, WA, USA: IEEE, pp. 885–892. ISBN: 978-1-4799-6923-4.

Mavridis, Nikolaos (2015a). "A review of verbal and non-verbal human-robot interactive communication". In: *Robotics and Autonomous Systems* 63, pp. 22–35.

– (2015b). "A review of verbal and non-verbal human–robot interactive communication". In: *Robotics and Autonomous Systems* 63, pp. 22–35. ISSN: 0921-8890.

McGuire, S., 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". In: *IEEE Robotics and Automation Letters* 3.3, pp. 1639–1646.

Milliez, Grégoire, 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)*, pp. 43–50.

Mitsunaga, N., C. Smith, T. Kanda, H. Ishiguro, and N. Hagita (2008). "Adapting Robot Behavior for Human–Robot Interaction". In: *IEEE Transactions on Robotics* 24.4, pp. 911–916.

Nguyen, Truong-Huy Dinh, David Hsu, Wee-Sun Lee, Tze-Yun Leong, Leslie P Kaelbling, Tomas Lozano-Perez, and Andrew H Grant (2011). "CAPIR: Collaborative Action Planning with Intention Recognition". In: *Seventh Artificial Intelligence and Interactive Digital Entertainment Conference*, p. 6.

Nikolaidis, S., P. Lasota, G. Rossano, C. Martinez, T. Fuhlbrigge, and J. Shah (2013). "Human-robot collaboration in manufacturing: Quantitative evaluation of predictable, convergent joint action". In: *IEEE ISR 2013*, pp. 1–6.

Nikolaidis, Stefanos, 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)*, pp. 189–196.

Nikolaidis, Stefanos and Julie Shah (2013). "Human-robot cross-training: Computational formulation, modeling and evaluation of a human team training strategy". In: *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. Tokyo, Japan: IEEE, pp. 33–40.

Nikolaidis, Stefanos, 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, pp. 294–302.

OECD (2017). *The Next Production Revolution*, p. 440.

Penkov, Svetlin, Alejandro Bordallo, and Subramanian Ramamoorthy (2016). "Inverse Eye Tracking for Intention Inference and Symbol Grounding in Human-Robot Collaboration". In: *Robotics: Science and Systems (RSS), Workshop on Planning for Human-Robot Interaction*.

Pfeiffer, Sabine (2016). "Robots, Industry 4.0 and Humans, or Why Assembly Work Is More than Routine Work". In: *Societies* 6.2, p. 16. ISSN: 2075-4698.

Pöppel, Jan and Stefan Kopp (2018). "Satisficing Models of Bayesian Theory of Mind for Explaining Behavior of Differently Uncertain Agents". In: *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems*. AAMAS '18. Stockholm, Sweden, pp. 470–478.

Ramakrishnan, Ramya, Chongjie Zhang, and Julie Shah (2017). "Perturbation Training for Human-robot Teams". In: *Journal of Artificial Intelligence Research* 59.1, pp. 495–541.

Ramirez-Amaro, Karinne, Michael Beetz, and Gordon Cheng (2015). "Understanding the intention of human activities through semantic perception: observation, understanding and execution on a humanoid robot". In: *Advanced Robotics* 29.5, pp. 345–362. ISSN: 0169-1864, 1568-5535.

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

Rojko, Andreja (2017). "Industry 4.0 Concept: Background and Overview". In: *International Journal of Interactive Mobile Technologies (iJIM)* 11.5, p. 77.

Rosman, Benjamin, Majd Hawasly, and Subramanian Ramamoorthy (2016). "Bayesian policy reuse". In: *Machine Learning* 104.1, pp. 99–127.

Sanelli, Valerio, Michael Cashmore, Daniele Magazzeni, and Luca Iocchi (2017). *Short-Term Human-Robot Interaction through Conditional Planning and Execution*.

Satoglu, Sule, Alp Ustundag, Emre Cevikcan, and Mehmet Bulent Durmusoglu (2018). "Lean Transformation Integrated with Industry 4.0 Implementation Methodology". In: *Industrial Engineering in the Industry 4.0 Era*. Cham: Springer.

Scassellati, Brian (2002). "Theory of mind for a humanoid robot". In: *Autonomous Robots* 12.1, pp. 13–24.

Stroop, J Ridley (1935). "Studies of interference in serial verbal reactions". In: *Journal of experimental psychology* 18.6, p. 643.

Sun, Ron, Xi Zhang, and Robert Mathews (2006). "Modeling meta-cognition in a cognitive architecture". In: *Cognitive Systems Research* 7.4, pp. 327–338. ISSN: 1389-0417.

Suzuki, Satoshi (1985). "Topological Structural Analysis of Digitized Binary Images by Border Following". In: *Computer vision, graphics, and image processing* 30.1, pp. 32–46.

Takayama, L., D. Dooley, and W. Ju (2011). "Expressing thought: Improving robot readability with animation principles". In: *2011 6th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp. 69–76.

Tan, J. T. C., F. Duan, Y. Zhang, K. Watanabe, R. Kato, and T. Arai (2009). "Human-robot collaboration in cellular manufacturing: Design and development". In: *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 29–34.

Tapus, Adriana, Maja J. Mataric, and Brian Scassellati (2007). "The Grand Challenges in Socially Assistive Robotics". In: *IEEE Robotics and Automation Magazine* 14.1, pp. 35–42.

Thiede, Keith W., Mary C. M. Anderson, and David Therriault (2003). "Accuracy of metacognitive monitoring affects learning of texts". In: *Journal of Educational Psychology*, pp. 66–73.

Thomaz, Andrea, Guy Hoffman, and Maya Cakmak (2016). "Computational Human-Robot Interaction". In: *Foundations and Trends in Robotics* 4.2-3, pp. 105–223.

Trafton, Greg, Laura Hiatt, Anthony Harrison, Frank Tamborello, Sangeet Khemlani, and Alan Schultz (2013). "ACT-R/E: An Embodied Cognitive Architecture for Human-Robot Interaction". In: *Journal of Human-Robot Interaction* 2.1, pp. 30–55.

Tulli, Silvia, 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: *BNAIC/BENELEARN*.

Vi, Minh Nghiem (2019). "A Human-Robot Collaboration Setup: Detecting Human Actions in a Task Using Semantic Perception". Bachelor's Thesis. Berlin, Germany: Technische Universität Berlin.

Villani, Valeria, Fabio Pini, Francesco Leali, and Cristian Secchi (2018). "Survey on human–robot collaboration in industrial settings: Safety, intuitive interfaces and applications". In: *Mechatronics* 55, pp. 248–266. ISSN: 09574158.

Wilcox, R., S. Nikolaidis, and J. Shah (2012). "Optimization of Temporal Dynamics for Adaptive Human-Robot Interaction in Assembly Manufacturing". In: *Robotics: Science and Systems*.

Yamato, J., J. Ohya, and K. Ishii (1992). "Recognizing human action in time-sequential images using hidden Markov model". In: *Proceedings 1992 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 379–385.

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