Unmanned Systems, Vol. 0, No. 0 (2013) 1–17

Co-adaptive Human-Robot Cooperation: Summary and Challenges

Sofie Ahlberg[a], Agnes Axelsson[a], Pian Yu[a], Wenceslao Shaw Cortez[a], Yuan Gao[b,c] *, Ali Ghadirzadeh[a], Ginevra Castellano[b], Danica Kragic[a], Gabriel Skantze[a], and Dimos V. Dimarogonas[a]

[a] *Department of Intelligent Systems, KTH Royal Institute of Technology, Stockholm, Sweden*
E-mail: {sofa,agnaxe,piany,wencsc,algh,dani,dimos}@kth.se,gabriel@speech.kth.se

[b] *Department of Information Technology, Uppsala University, Uppsala, Sweden*
E-mail: ginevra.castellano@it.uu.se

[c] *Center for Intelligent Robots, Shenzhen Institute of Artificial Intelligence and Robotics for Society, Shenzhen, China*
E-mail: gaoyuankidult@gmail.com

The work presented here is a culmination of developments within the Swedish project *COIN*: Co-adaptive human-robot interactive systems, funded by the Swedish Foundation for Strategic Research (SSF), which addresses a unified framework for co-adaptive methodologies in human-robot co-existence. We investigate co-adaptation in the context of safe planning/control, trust, and multi-modal human-robot interactions, and present novel methods that allow humans and robots to adapt to one another and discuss directions for future work.

Keywords: Co-adaptive systems; Human-in-the-loop systems; Human-robot interaction.

1. Introduction

An overarching goal of robotics research is for robots to co-exist with humans in the real world. Significant applications of human-robot co-existence include domestic[1] and industrial environments.[2] Examples of human-robot co-existence include robotic manipulators safely cooperating with humans in a manipulation task[3] as well as cooperative manned/unmanned flight formation control.[4] There also exist frameworks that address relations in human-robot collaboration/co-existence.[5] A key component of co-existence is the ability of both humans and robots to *adapt* to one another. Adaptation can occur bi-directionally, that is the human adapts their behaviour to the robot as the robot adapts to the human. This bi-directional adaptation is also coined **co-adaptation**.[6] Recent work in co-adaptation is focused in developing trust between human and robot.[6,7] However, it is arguable that the concept of co-adaptation is much broader than trust.

Human-robot co-existence is indeed an interdisciplinary field encompassing robot planning/control, trust, multi-modal human-robot interaction, natural language processing, and machine learning. Robot planning and control investigates how to implement autonomous decision making and execute actions in a safe manner to accomplish a task with/for a human. Multi-modal human-robot interaction (HRI) addresses multi-modal communication between humans and robots, including speech/non-speech related communication as well as physical interactions/safety between the human and robot. In order to cooperate, humans and robots must be able to communicate. Natural language processing (NLP) handles the use of natural language for communication between human and machine. This establishes a basis of communication between the two. Finally, machine learning addresses the ability of the robot to update its knowledge of the environment, task, and human.

We claim that such an interdisciplinary topic as human-robot co-existence can greatly benefit from co-adaptation. Here we focus on three main concepts in co-existence: a) safe planning/control b) trust and c) multi-modal HRI with respect to NLP and machine learning. Each of these concepts can be viewed by a feedforward/feedback mechanism that describes the exchange of information and actions between humans and robots (see Figure 1). The "Planner", "Control Law", and "Agent Action" boxes of Figure 1 represent the hierarchical levels of autonomy in the robot system. The "Planner" stage determines what actions need to be taken, the "Control Law" determines how to achieve the desired action, and the "Agent Action" is the resulting response of the robot agent. Feedforward and feedback exchanges are represented by blue

*Most of the work was done at Uppsala University.

and red arrows, respectively. The interactions between humans and the various levels of autonomy of the robot system are dependent on the task. Figure 1 is meant as an abstract template to depict the exchange of information between a human and robot agent. To better explain this template, consider for example a standard navigation task in which a human is commanding a robot to reach a certain goal location. First, the human must specify the location of the goal to the robot, and this information exchange is represented by the arrow between "Human" and "Planner". Second, the "Planner" discretizes the entire robot workspace and an algorithm chooses the sequence of cells of the workspace that will lead the robot to the goal. The sequence of cell locations the robot must pass through is sent from "Planner" to "Control Law". The "Control Law" block then computes the necessary motion command (steering angle and velocity) based on the robot's current position to reach each individual cell, and ultimately reach the goal. The steering angle and velocity commands are then sent to "Agent Action", whereas the position of the robot with respect to the cell is fed back to the "Planner" so that the planning algorithm is aware of the current cell the robot occupies and can update the plan accordingly. Then, the agent executes the steering angle/velocity command, and feeds back its current position to the "Control Law". To complete the loop, the arrow between "Agent Action" and "Human" represents human perception, i.e., the human sees that the robot is indeed achieving its goal. As the human sees the robot's motion, they may update the plan online in which case the process is repeated for the new goal to be reached. The last remaining arrow not yet addressed in this example is from "Human" to "Control Law" in which the human may provide joystick commands to directly control the robot steering angle/velocity so that the human may intervene in the "Control Law" block.

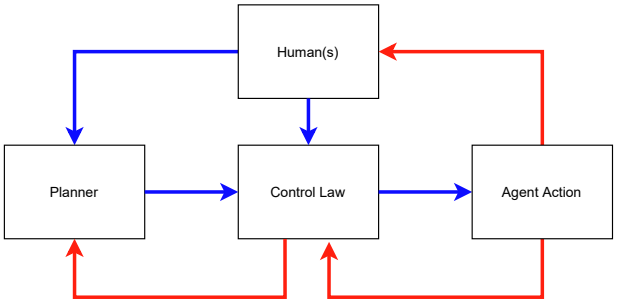

Fig. 1. Co-adaptation Template

The work presented here is a culmination of developments within the Swedish project *COIN*: Co-adaptive human-robot interactive systems, funded by the Swedish Foundation for Strategic Research (SSF), which addresses a unified framework for co-adaptive methodologies in human-robot co-existence. We investigate co-adaptation in the context of safe planning/control, trust, and multi-modal human-robot interactions. This is done in the context of Figure 1 further confined to each of the three main co-existence concepts and investigating the synergies thereof. We present novel methods that allow humans and robots to adapt to one another and discuss directions for future work.

The rest of the paper is organized as follows: In Section 2, we summarize the related literature in the context of co-adaptation for human-robot systems. In Section 3, we summarize co-adaptation in the context of safe planning and control. In Section 4, new developments in co-adaptation and trust are presented. Section 5 addresses multi-modal interaction in the context of co-adaptation. Section 6 presents challenges in existing methods and directions for future work. Finally, Section 7 concludes the paper.

2. Related work

2.1. *Safe Planning and Control of Robot Systems*

One objective in robotics research is to synthesise controllers for autonomous robotic systems that are given complex task and safety specifications. During the past few years, the design of algorithms to generate such controllers has found a wide support in formal verification methods[8] that combine two major advantages: temporal logics provide a rigorous and highly expressive specification means with some resemblance to the natural language.[9] Formal verification-based algorithms can then be employed to synthesize a robot controller guaranteeing the accomplishment of the given temporal logic specification in the given environment.[10,11] Some recent papers[12,13] use barrier functions to satisfy specifications expressed by linear temporal logic (LTL). The control policies in these works are correct-by-design but does not address uncertainties, unknowns, dynamical environments or how to progress when the given specification is not feasible.

Several recent studies in temporal logic-based control have focused on dealing with general sources of uncertainty,[14] or inherent system uncertainty.[15] In these works, an offline synthesis procedure yields a correct-by-design non-adaptive feedback controller that is not suitable for deployment in co-adaptive systems. Some studied topics include addressing efficient reconfiguration of robot controllers based on learning of a partially unknown, static environment,[16,17] and a subclass of dynamic environments.[18,19] Therein, the synthesized controllers adapt to the system over time, but the representation of the uncertain elements and the learning procedure are too simplified to cover the aspects brought by human presence. Non-feasible specifications have been addressed by revising the formula but keeping it as close as possible to the original task[20] and by minimizing introduced metrics which measure the violation of the task.[21] This is an important part of human-robot co-adaption since the likeliness of specification feasibility decreases as the robot adapts for the human.

To our best knowledge, none of the existing literature in temporal logic-based planning and control has aimed specifically at co-adaptive human-in-the-loop robotic systems. Temporal logic specifications have been combined with a variety of additional criteria, such as robot's energy consumption, deadlines etc.[22] However, none of them considers a systematic human-in-the-loop approach to permit *co-adaptive* planning and control.

2.2. *Co-adaptation and Trust*

The ability to interact with humans in a socially acceptable way and adapt to their needs, preferences, interests, and emotions in a contingent manner is of paramount importance for robots to become achieve trust. Simulating the tremendous social adaptation abilities that characterize human interactions requires the establishment of bidirectional processes in which humans and robots synchronize and adapt to each other in real-time to achieve mutual co-adaptation.[23–25]

In social robotics, while there is a large body of literature on robots learning skills from human teachers in a social context, endowing robots with the ability to adapt to humans via incremental learning is still heavily underexplored. A strand of approaches leverages advances in automatic human behavior analysis, affect recognition, and reinforcement-learning (RL) to incrementally adapt the robot's behaviors to maintain the user in a positive affective state.[23] These methods use RL to find an optimal policy of robot behavior with respect to rewards calculated as a function of the automatically detected user's affective state. Despite the promising results, to date, traditional RL methods using affect-based rewards are not sensitive to context and have relied solely on the often noisy outputs of automatic affect detectors programmed in the robot.[24] However, an open problem is how to leverage human feedback towards the development of robust approaches for affect co-adaptation in social HRI. Similarly, little previous work has addressed the development of context-sensitive methods for co-adapting problems with emotion-based reward mechanisms that can handle multiple users.[23,26]

Few researchers have investigated RL for social human-robot co-adaptation over the last few decades. Different RL algorithms have been applied to implement adaptive behavior selection in different fields, such as education. For example, a Q learning-based effective model was considered to determine the verbal and non-verbal behaviors of social robots in an educational game to facilitate effective personalization.[27] In addition, contextual bandit algorithms were applied to adaptively control the pace of interaction based on user performance and effective feedback.[28] RL was also used to personalize the robot according to each individual's learning difficulty level.[29] An RL framework was proposed to enable robots to select supportive behaviors in game-based learning scenarios to maximize task performance.[24] Furthermore, we observe a growing trend of applying different learning-based mechanisms in the HRI domain.[30–34] Simultaneously, various RL models have been applied to give information to robot tutors.[35] The results of these studies influenced individuals' positive attitudes and contributed to improved job performance or learning ability.

2.3. *Multi-Modal Interaction*

Traditional computer systems expect their users to interact with the system in specific ways. In the 1960s and 1970s, the main form of interaction was syntactically precise command line interfaces, and while user design has since moved on to interfaces that more closely align with how we have learned to interact with computers, the interaction is still static, and something the user has to learn.

When humans communicate with humans, the parties of the interaction adapt how they present their utterances and reactions to the other, ending up in a situation where both believe the interaction to be proceeding well enough for whatever the goal is.[36,37] The difference between this and the traditional computer interface is that the computer interface does not adapt back to the user – even though the user does in fact adapt to what the system expects them to do. Multi-modally interactive systems, and multi-modally interactive agents, are computer systems that use some part of how humans are used to adapting to a co-adapting partner to enhance the interaction between the human user and the robotic, or computerised, system. Such a system can adapt its style of speaking, including both verbal (choice of words, pitch, speech rhythm, etc.) and non-verbal expressions (facial gestures). This kind of entrainment can be expected to increase the level of rapport between the user and agent,[38] which in turn can increase for example the learning efficiency.[39] Second, such a system can also adapt the interactional dynamics, which include for example turn-taking and the pace of information delivery.[40] Just like humans do, the system can package information into appropriately sized chunks and then monitor (verbal and non-verbal) feedback from the user in order ensure common ground. Repeated positive feedback might mean that the pace can be increased, whereas repeated negative feedback means that the instructions must be explained in more detail. The system can also monitor the attention of the users and adapts its speaking to that.

In situated multi-party interaction, it is not certain that the users always are attending the system. For example, they could shift their attention towards the task at hand, or towards each other. Since the attention of humans is limited, the system must be aware of this when interacting with the users.[41] While theory about how humans ground information between each other is well-established and has good models and abstractions of how those processes work, as described briefly above, few interactive human-to-robot systems actually use those models or abstractions as a basis for interacting with their users. To be able to fully use models of grounding like that of Clark,[36,42] interactive systems must structure their information and information about what their users know and think about

that information in some way that relates to the grounding state; we address this in Section 5.1.

We also consider non-verbal interaction between human and robot partners in which multi-modal feedback can help the robot to better understand the state, action and intention of the human partner, e.g., through gaze and gesture recognition,[43] body movement prediction and recognition,[44–47] force/torque measurements in physical interactions,[48] and modeling bio-signals.[49] We describe our method to implement pro-active robot behavior based on multi-modal feedback in non-verbal settings in Section 5.2.

3. Human-in-the-loop Plan and Control Synthesis with Safety Guarantees

In this section we discuss the role of co-adaptation in the context of robot planning and control. We consider a novel form of robot planning using formal methods and temporal logic specifications.[8] Temporal logic offers ways to express complex tasks and planning problems as formulas that can be adapted to human language, and can be used for example with speech-type commands.[50,51]

We address co-adaptation in robot planning and control using the block diagram in Figure 2. As shown in the figure, the human(s) can provide several forms of inputs to the autonomous system. First, the human can express a desired plan as a high-level temporal logic formula to be executed. The planner then develops a sequence of actions implemented by a control law to realize the humans' plan. Second, the human can also provide non-speech, joystick inputs into the system during the low-level control of the robot. The human receives feedback of the resulting plan by visual perception of the robot's actions, and can then adapt the plan/low-level control accordingly. The robot is able to adapt to the resulting human actions by using sensor measurements and changing the control law/plan as required to satisfy the original high-level task.

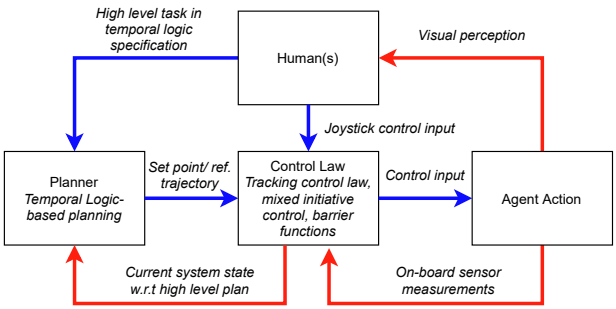

Fig. 2. Co-adaptation for Planning and Control Synthesis

3.1. *Language-based Plan Synthesis*

There exist several forms of temporal logic specifications, including Linear Temporal Logic (LTL) and Metric Interval Temporal Logic (MITL).[8] Both LTL and MITL formulas are composed of temporal operators (*always* $\Box$, *eventually* $\Diamond$, *until* $\mathcal{U}$, and *next* $\bigcirc$), logic connectives (*conjunction* $\wedge$, *disjunction* $\vee$, *negation* $\neg$, *implication* $\implies$) and atomic propositions. Simple examples of tasks which can be expressed are: *eventually a must be true* ($\Diamond a$), and *b and c must always be true* ($\Box(b \wedge c)$). By combining operators, the tasks can become more complicated such as: *a must be true on a regular basis* ($\Box\Diamond a$), or *eventually b must be avoided forever* ($\Diamond\Box\neg b$). In MITL each temporal operator is associated with a time interval I which limits the impact of the specific part of the formula to a fixed time interval. The time intervals can include any non-negative values. If $I = [0, \infty)$ the implication is the same as for LTL. In our work we have limited the expressiveness to time intervals on the form $I = [0, x > 0]$, which is equivalent to putting deadlines on the operators. By using LTL or MITL as our specifications we allow the human who is giving the tasks to express them in an easy-to-understand, high-level manner. Considering a potentially large environment where the properties the human is interested in may be satisfied in multiple non-connected areas as well as overlap with each other, it can become very complicated and limiting when described in a conventional point-to-point navigation task. As a helpful example one can consider a robot charged with delivering coffee to workers in an open office space. In the office there are multiple tables and chairs at different locations as well as two coffee machines. The robot must move to one of the coffee machines, get the coffee and take it to the worker, without colliding with any of the furniture. While it would be difficult to express this task as a low-level specification, and impossible to do so without limiting the robot to using a specific coffee machine, with LTL it is as simple as: $\Diamond(\textit{at coffee machine}) \wedge (\textit{at coffee machine} \implies \textit{pour coffee}) \wedge (\textit{pour coffee} \implies \Diamond \textit{at worker}) \wedge \Box(\neg \textit{at furniture})$.

3.2. *Human-in-the-Loop Planning and Control*

Here, we present methods for control and plan synthesis (for LTL and MITL) with the consideration of humans in-the-loop. In both cases the overall framework follows the steps: i) a human supplies the system with a temporal logic task, ii) the temporal logic formula is translated into an automaton which is later intersected with a transition system, iii) a graph search algorithm is applied to find a nominal plan, iv) human control input is merged with the nominal plan through a mixed-initiative controller and, v) the resulting control plan is used by the system to learn some user-preferred parameter used in the planning. Returning to Fig. 2, we note that step i is represented by the blue arrow *'High level task in temporal logic specification'*, step ii-iii and v all take place within the box 'Planner *Temporal Logic-based planning*', and step iv occurs in the box *'Control Law Tracking control law...'*.

For LTL specifications, we suggest a framework where the human can assign hard, soft, and temporary tasks (sim-

ple motion tasks which are performed once) to a robot, as well as give joystick control input.[52] The system applies a classical automata approach to find an initial plan which satisfies the hard constraints while balancing the violation of the soft constraint against the limit on the control input. During the execution of the plan the human can give input as temporary tasks and joystick commands. A path patching approach is used to create small deviations from the original plan to satisfy the temporary tasks and a mixed-initiative controller is applied to ensure safety when the human gives joystick commands. The path patching considers the planned path and searches for the point at which the goal region of the temporary task can be reached in the shortest time, while having the least impact on the cost associated with the non-temporary tasks. This is solved by an optimization problem on which details can be found in.[52] The mixed-initiative controller is an additive control function which measures the distance to unsafe regions d_t and determines how much of the human control input is used depending on this distance:

$$u = u_r + \kappa(d_t)u_h$$

where $\kappa(d_t) \in [0, 1]$ is designed such that $\kappa = 1$ if $d_t \geq d_s + \epsilon$, $\kappa \in (0, 1)$ is decreasing with d_t if $d_s + \epsilon > d_t > d_s$, and $\kappa = 0$ if $d_t < d_s$ for some safety design parameters d_s and ϵ. This is satisfied by the function we suggested in:[52]

$$\kappa = \frac{\rho(d_t - d_s)}{\rho(d_t - d_s) + \rho(\epsilon + d_s - d_t)}$$
$$\rho(s) = \begin{cases} e^{-1/s}, & s > 0 \\ 0, & s \leq 0. \end{cases}$$

The unsafe regions are in turn determined based on the criterion that the hard task can not be satisfied whence an unsafe region is entered. As an example, consider a robot tasked with repeatedly picking up and delivering packages at fixed locations (hard task). The robot should avoid an area where humans sometimes pass through (soft task). However, this area is in the middle of the shortest path, and to avoid it the robot must use more resources (exceeding the control limit). Initially the robot plans to completely avoid the area, the human co-pilot then uses his/her joystick to steer the robot through the edge of it indicating that a small violation of the soft task is preferred over using more resources. The mixed-initiative allows the human interference since the hard task is not affected. The system updates its knowledge of the human preference based on the real-time trajectory, and uses it in future planning.

For MITL we suggest a framework for least-violating control. By least-violating control, we mean that the framework will satisfy the high level task if it is feasible. We note that the use of timing in the MITL setup may be difficult to ensure in the event of humans or obstacles that may obstruct the robot. In this sense, least-violating control allows the system to satisfy the high level task with respect to a quantitative metric coined the *hybrid distance*. The hybrid distance measures the violation of a MITL specification.[53] The classical automata approach is applied with the hybrid distance as a cost function to find the least-violating plan. The hybrid distance consists of two quantitative measurements of the violation and a design parameter which determines a priority balance between the violation types. The role of the human in this framework is to supply the system with a task and a value on the design parameter. This design parameter value can be given directly or via human control input to the low level controller. The latter is further discussed below in the context of co-adaption and learning. The human control input is applied to the system through a mixed-initiative controller. The mixed-initiative controller used for MITL follows the same concept as that for LTL, albeit in the MITL case the unsafe regions are dependent on time.

In both the LTL and MITL frameworks, mixed-initiative control is employed to allow a human to control the system subject to satisfying the hard tasks. Now, let's consider how the autonomous system can learn and update its plan for human preference. This allows the system to co-adapt with the human in a safe manner. We suggest two approaches for co-adaption. Both treat the hard task as fixed and use low-level human control input to change elements of the soft task to closer reflect the indicated preference.

First, the human control input to the low level controller can be used in an inverse reinforcement learning (IRL) algorithm.[52,53] This approach can be applied for both LTL and MITL. In the first case the algorithm is used to learn the value of a weight parameter which determines the priority between satisfaction of the soft constraints and the limit of the control input. It is assumed that the human is helping the system and the trajectory resulting from the human's influence is used to determine how the weight should be improved.[52] For example, if the human steers the robot in a trajectory that violates the soft constraint to a greater extent than the system's plan, this indicates that the human prefers the limitation on the control input to be of more importance. By changing the value of the weight accordingly the system can take this in consideration when finding a new plan. If the assumption that the human is consequently trying to help is true, the system will eventually find the optimal path and the human will no longer be needed as a co-pilot. For MITL the system instead learns the value of the internal design parameter in the hybrid distance. This internal design parameter determines the relative importance of *continuous* and *discrete* violations of the soft task. The continuous violation consists of unmet deadlines, while the discrete violation consists of visits to undesired regions. As for LTL, it is assumed that the human is helping the system. In this case, we maximize the improvement of the violation achieved by following the trajectory the human indicates compared to all previous plans.[53] An example of this is if the human steers the robot towards a target region, using a shorter path than the system's plan indicated, by going through an undesired region. This indicates that the human has a preference towards deadline satisfaction, i.e. minimizing continuous violations. By up-

dating the design parameter to reflect this, a new plan can be synthesized which better suits the human's need. It is important to note here that this only affects the satisfaction of the soft constraint. These parameters have no impact on the planning for satisfaction of the hard constraints. Hence, safety is not affected by the learning, and if combined with the mixed-initiative controller discussed in Section 3 safety is guaranteed independent of the quality of the current estimation of the parameters.

Another direction that has been considered for co-adaption with regards to MITL is to consider a system where the human has his/her own soft task to satisfy. By making some assumptions on the task structure and that the resulting path with human inputs is acceptable for the task, the system can learn this task using a simple algorithm.[54] As for the previously discussed co-adaption approaches, safety is solely reliant on the mixed-initiative controller since the impact of the changes caused by the learning is limited to the satisfaction of the soft constraint.

4. Co-adaptation and Trust

Co-adaptation and trust is vital for human-robot interaction. In this section, we address the issue of trust and co-adaptation between humans and robots in different scenarios. In the next paragraphs we will analyze how our studied scenarios help robots to build social state perceived trust with humans, in addition to their specific implementation that we discuss in the related papers. Combined with the co-adaptation framework, the problem of trust and co-adaptation in different scenarios can be comprehensively analyzed from three aspects, namely the learning framework, the emotion-based co-adaptation approach, and the comprehensive guidelines for human users.

The overall structure for co-adaptation and trust is shown in Figure 3, which illustrates the relationship between our studies and the never-ending circle of co-adaptation. Specifically, in this Figure 3, we first develop a co-adaptation framework based on an affective model as our control law. More concretely, we design an interactive model based on the psychological model with adjustable key parameters using machine learning methods. This model can generate robots' behaviors, such as verbal utterances or gestures, for human-robot interaction in the real world. These verbal utterances or gestures that are fed back to the human will elicit a response from the human. The reinforcement learning model then observes the human response and adjusts the parameters of the previous control model so that the model will produce actions that are more suitable for a smoother co-adaptation process of human-robot interaction.

In our HRI scenarios, the participants interact with a robot implemented with RL. The participants give explicit or implicit feedback to the robot verbally. While the robot processes the feedback by taking psychological dynamics into account, the robot, based on the algorithmic dynamics, also uses different modalities to influence the participants.

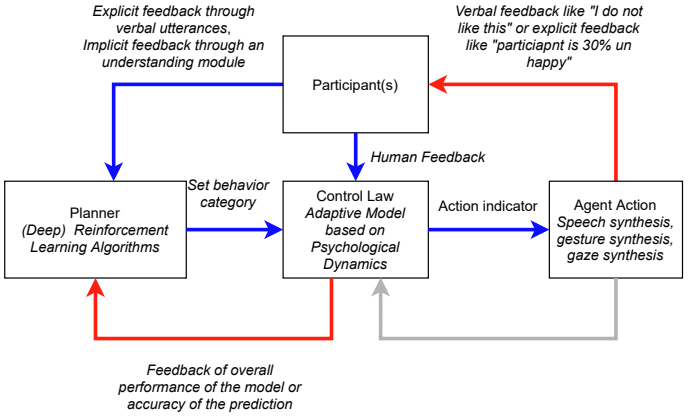

Fig. 3. Co-adaptation for Trusted Human-Robot Interactions

We mainly considered three scenarios. The first is the "tutorial scenario", where a robot acts as a tutor to guide an interactive learning process. We build a framework associated with this scenario to support the connection between different learning algorithms and robotic interfaces. The second scenario is the group-approaching scenario, in which a robot needs to learn the best strategy to approach a group of people while considering the feedback of the agents in the group. The third scenario is the "escape room scenario", where the robot acts as a guide or an adversarial agent. In this scenario, participants are asked to interact with the robot to find a way to escape the room. For the escape room, we built another framework that is integrated with augmented reality (AR) to support the communication between algorithms and the scenario.

In the previous scenarios, affect-based co-adaptation approaches are also built to interact with participants. The scenarios are established and tested with different interactive patterns. For example, in the tutorial scenario,[24] a framework that incorporates the algorithm Exp3[23,24,26,55] allows us to test the overall effect of the algorithm during a human-robot interaction. The escape room scenario is implemented to guide or test player's internal state during their exploration in the escape room scenario. In the following paragraphs, we will use the escape room scenario and the logic represented by Figure 3 as an example to introduce the adaptive process.

4.1. *Adaptation in an Escape Room Scenario*

In socially assistive robotics, an important research area is the development of adaptation techniques and understanding their effect on human-robot interaction. To investigate whether the basic trust could be observed during the human-robot co-adaptation in an escape room scenario, we present a meta-learning based policy gradient method for addressing the problem of adaptation in human-robot interaction and also investigate its role as a mechanism for

trust modelling. By building an escape room scenario in mixed reality with a robot, we test our hypothesis that bi-directional trust can be influenced by different adaptation algorithms. We found that our proposed model increased the perceived trustworthiness of the robot and influenced the dynamics of gaining the human's trust. Additionally, participants evaluated that the robot perceived them as more trustworthy during the interactions with the meta-learning based adaptation compared to the previously studied statistical adaptation model.

In order to model the interaction in our scenario, we loosely follow the assumption that a human-like robot should have the tripartite mental activities, namely conation, cognition, and affection. This assumption is inspired by Hilgard's tripartite classification of mental activities of human personality, and intelligence in modern behaviour psychology.[56] For our particular escape room scenario, each mental functionality instance is modelled and implemented as an adversarial Multi-Armed Bandit (MAB) problem. All of the three instances operate independently throughout the interaction process. We define three different meta-learning processes and for each mention functionality. Each meta-learning strategy contains two processes $\{\zeta_p^c, \zeta_r^c\}$, where c corresponds with a mental functionality, p and r stand for two different processes.

During the interaction, the robot needs to optimize all of its policies $(\pi^c)^*$ to learn the most preferred action for each MAB environment. To keep the generality of the concept of trust, we also assume the observational states of each category $\mathbf{s}^c$ to be fixed for each category c. As stated before, our methods involve two training processes. Firstly, we model the human feedback of each action of MAB as a Gaussian distribution $r^c \sim \mathcal{N}(\mu^c, (\sigma^c)^2)$ for all the auxiliary environments. This modelling is based on the condition that the signal of emotion recognition normally follows Gaussian distributions.[57] Simultaneously, we use ζ_p^c to train randomly initialized π^c to get a meta policy π^c_{meta}. π^c_{meta} learns the inner structure of the problem, which makes the adaptation during the interactive session much faster and data-efficient. Finally, we conduct human experiments and study different subjective measures along with interaction. We used MAML [58] as the meta-learning algorithm M^c and trust region policy optimization (TRPO)[59] as the optimization algorithm for all policies. MAML is a meta-learning algorithm based on gradient optimization. From a high-level perspective, meta-learning algorithms aim to improve the adaptability of the learning model. In theory, this kind of algorithm is able to generate good initialization parameters for the TRPO algorithm for a set of tasks. Good parameter initialization allows TRPO to learn better and perform well even with a small amount of data for training. An escape room scenario was then built to conduct a between-subjects study, in which participants interacted with a Pepper robot[a]. The escape room was created in augmented reality, and participants were required to wear a Mixed Reality headset HoloLens[b] to see the walls of the virtual maze, triggers, keys, and the exit door(see Figure 4). The interaction consists of three parts: an instance of conation, an instance of affection, and an instance of cognition. Each instance is triggered by the participant's position in the virtual maze, recorded from the HoloLens. For each instance, the robot chooses one out of four actions, according to a probability distribution provided by the algorithm. Here, an action is implemented as a verbal question. After the participant answers, the robot updates the probability associated with the previous question and gives feedback accordingly. After completing all the interaction steps, the participant can escape the room, and a run is over.

For the control group, a statistical MAB algorithm Exp3 (**C1**), implemented as in the previous studies, was used. The experimental group interacted under our proposed model, a policy gradient-based solution for the MAB problem, together with meta-learning (**C2**).

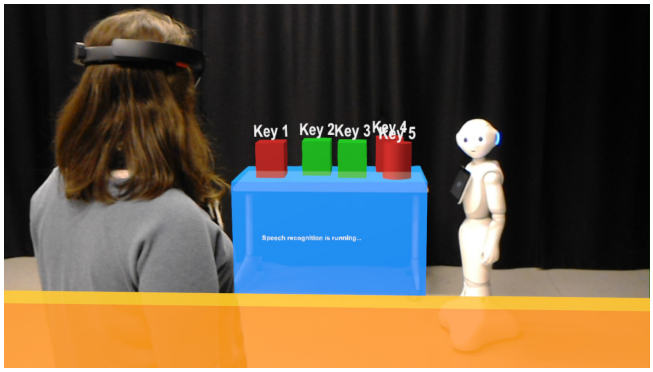

Fig. 4. One of the interactive processes from a third-person perspective. Participants wear a mixed reality headset through which they can observe virtual objects (e.g., the keys, the blue table, and the orange walls of the escape room) being augmented into the real world. From this first interaction, through verbal utterances, the participant starts to understand what the robot wants to know, and the robot also begins to learn the participant's preferences.

Bi-directional perceived trust is measured from the participant's point of view, i.e., how trustworthy they perceive the robot, and how much they think the robot trusts them in return. We evaluate the bi-directional perceived trust by following Salem et al.'s work on trust evaluation in HRI.[60] For this purpose, two hypotheses are made: **H1**: the perceived trust of a participant towards the robot is *higher* in the meta-learning supported group and **H2**: the perceived trust of a robot towards the participant is *higher* in the meta-learning supported group. Among two conditions, a statistically significant effect of the condi-

[a]https://www.softbankrobotics.com/emea/en/pepper
[b]https://www.microsoft.com/en-us/hololens

tion was found on perceived trust towards the participant, $F(1, 22) = 16.44, p < .001, \eta^2 = .428$. Even though the overall robot's trust towards the participant depends on the condition, the differences in dynamics of how the perceived robot's trust towards the participant changes in two conditions are not statistically significant, $F(3, 66) = 1.673, p = .181, \eta^2 = .071$.

In summary, we designed an escape room scenario in mixed reality to evaluate the proposed method and investigate its potential effects on the perceived bi-directional trust. We point out that the meta-learning processes could be understood as a basic trust model for fast adaptation. This allows people who interact with robots to gain a perception of trustworthiness towards the robot. Our results show that not only did the algorithm adopt a higher learning rate after the meta-learning process, but it also increased the participant's perception on how trustworthy the robot perceives them.

4.2. *Summary*

While working on this project, we tried to implement different scenarios to demonstrate that fluid human-machine interaction can be achieved by adapting to human needs. However, this adaptation can happen in different forms, given that it is simple engineering, meta-learning, learning of few, or just pure reward engineering. When the same smooth interaction occurs multiple times in a task, basic trust is generated. Although we have implemented different scenarios to build human-robot interaction scenarios, we use an escape room scenario as an example to illustrate the contribution of our research to socially conscious human-robot co-adaptation.

As a scenario implemented in our project, this scenario is typical of our overall project thinking direction. In this direction, our project adopts different machine learning mechanisms to realize different dynamic processes of socially aware human-robot co-adaption, including human-robot interaction processes in continuous and discrete control scenarios. These projects have collectively helped us understand the trust and co-adaptation problems in human-robot interaction, provided us with a deeper understanding of these problems, and inspired our future research topics.

5. Multi-Modal Interaction

This section addresses how co-adaptation is addressed for multimodal interaction between humans and robots. These types of adaptation happen in verbal or non-verbal modalities, depending on how the system interacts with the user and what the mutual task is. In Section 5.1, we describe our work in using behaviour trees and knowledge graphs to model systems that can adapt to user behaviours through verbal adaptation. Section 5.2 describes our approach for non-verbal adaptation using behaviour trees and reinforcement learning.

5.1. *Modelling verbal adaptation using behaviour trees and knowledge graphs*

This sub-section addresses verbal adaptation from a robot towards a reacting user. An example of the type of adaptation we describe here would be when a conversational system adapts its speech to the user by repeating a misunderstood utterance, or by changing the way it plans to refer to an entity by knowing that the person is more likely to understand a reference A than another reference B. Our scenario for exploring this form of adaptation is one where a robot is presenting a piece of art to a human, similar to a museum setting, as seen in Figure 5. In such a setting, it is important that the presenter (the robot) adapts the presentation to the listener (the human), based on the feedback it receives.

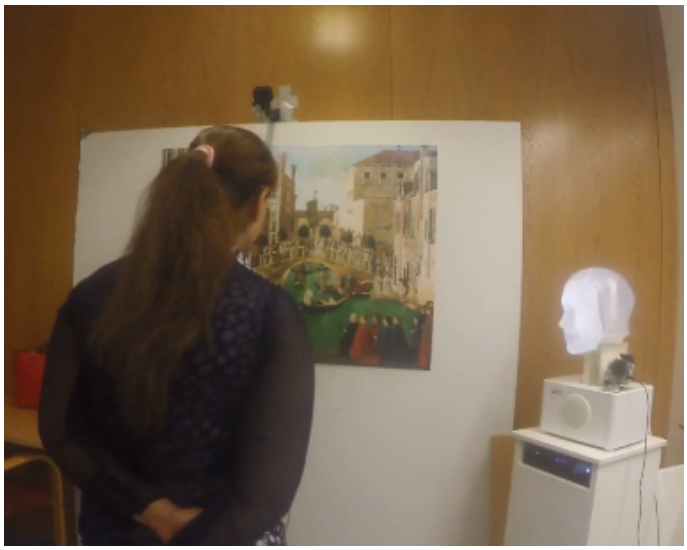

Fig. 5. The robot presenting a piece of art to a human listener.

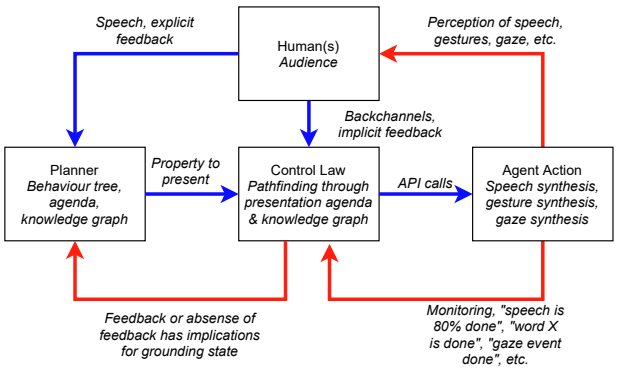

Fig. 6. Co-adaptation for Multi-Modal Interaction (verbal)

Figure 6 shows an adaptation of Figure 1 for the application of multimodal presentation between a human and a

robot. The data that flows from the robot to the user consists of verbal and non-verbal feedback and communicative information, presented as speech, gestures, facial or body expressions, as well as gaze. The robot's task is to interpret such information given by the user to adapt its presentation, whereas the user's task is to understand the information being presented, and provide feedback which indicates the user's level of understanding.

The presentation system can be analysed by mapping individual components to the blocks of Figure 6. From this perspective, the planner component (the leftmost box in the figure) of the system is a behaviour tree communicating with a knowledge graph, which serves as a user model. The shortest-term adaptation happens through a microplanner component which is controlled by the behaviour tree using information extracted from and stored in the knowledge graph. The behaviour tree is shown in Figure 7.

Table 1. An example of the system output from Section 5.1.1. *GC* is grounding criterion, *att* is attention, *und* is understanding and *acc* is acceptance as described by Clark and Brennan[36] and Clark.[42]

Speaker	Speech	Interpretation / action	GC
Robot:	Van Gogh died i-		*Att*
User:	Huh?	Negative hea.	*Und*
Robot:	Vincent van Gogh...	Alternative reference	*Und*
User:	I don't recognise that name.	Negative und.	*Und*
Robot:	the artist...		*Und*
User:	Oh.	Positive und.	*Und*
Robot:	died in 1890.		*Acc*
User:	...	Not meeting GC	*Acc*
Robot:	... 130 years ago.	Alternative reference	*Acc*
User:	...	Not meeting GC	*Acc*
Robot:	Right?	Eliciting acceptance	*Acc*
User:	Yeah, sure, I got it.	Positive acceptance	*Acc*

5.1.1. *System components*

A **behaviour tree** controls the presentation system's behaviour priority. The primary motivation for using a behaviour tree for modelling this type of high-level behaviour is that it allows the system to model the concepts of *upward evidence* and *downward completion*, which are central concepts in the theory of feedback and grounding in human-human communication.[42] According to Clark, positive feedback on a high level also implies positive feedback on all lower levels, since it is impossible to accept or understand an utterance if one has not also heard and attended to it, and the same holds in reverse for negative feedback. This basic correspondence is modelled in the coloured boxes found in Figure 7.

In Figure 7, behaviour tree selectors and sequences are used to prioritise first finding a user (top left), then interacting with that user and giving them the turn if they are trying to speak. If the user does not take the turn, the robot takes it by first reacting to user feedback, if applicable, and then ensuring joint attention, joint hearing, joint understanding and joint acceptance in order, as defined by Clark.[42] The order of the operations emulates Clark's *upward evidence*. After these operations have been allowed to modify the system's planned utterance, it speaks (bottom right).

To serve as a source of structured information for how the behaviour tree retrieves and stores information about the presentation and the users attending to that presentation, a **knowledge graph** extracted from WikiData[63] is used. This structure is used for two purposes. First, is queried to give the system ways to refer to previously grounded information when trying to present new information. Second, it is itself used to store the user's grounding status (i.e., how well the user has understood the presented facts) in terms of those pieces of information. Wikidata and the similar DBPedia are both commonly used to drive content-independent chatbots and dialogue systems.[64–66]

An example of how the state of grounded information changes over the course of a presentation is shown in Table 1. The user gives verbal feedback indicating that they do not know who Vincent van Gogh was. At first, this feedback is unclear, resulting in a minor correction corresponding to the middle box in Figure 6. As the human gives more clear negative feedback indicating that they do not know who van Gogh was, the system must re-start its utterance completely, using the referring expression *the artist* instead. As the user responds positively to the completed line *The artist died in 1890*, the knowledge graph is updated with positive acceptance in regards to the statement that this is true for the user, and the behaviour tree in Figure 7 moves on with the presentation by choosing some other statement to present, if possible and appropriate.

The types of feedback that update the knowledge graph flow through either the blue edge labelled *'speech, explicit feedback'*, or the red edge labelled *'feedback or absence of feedback...'* in Figure 6. The former is used for feedback that directly changes the state of the knowledge graph, like if the user says "I don't know who Vincent van Gogh was." The latter is used if the user's implicit feedback or absence of feedback does or does not meet the grounding criterion set in the microplanner.

In the terminology of Levelt,[67] *microplanning* is the process through which a speaker chooses the way to present a piece of information that takes the audience's perspective into account. The behaviour tree shown in Figure 7 uses a **microplanner** component derived from this definition to store the utterance it has decided to present from the knowledge graph. This corresponds to the middle box labelled *'Control Law'* in Figure 6. We use a basic theme-rheme format for our utterances. The *theme* of an utterance is the thing or person being talked about, and the *rheme* is the statement being presented about the theme. Viewing

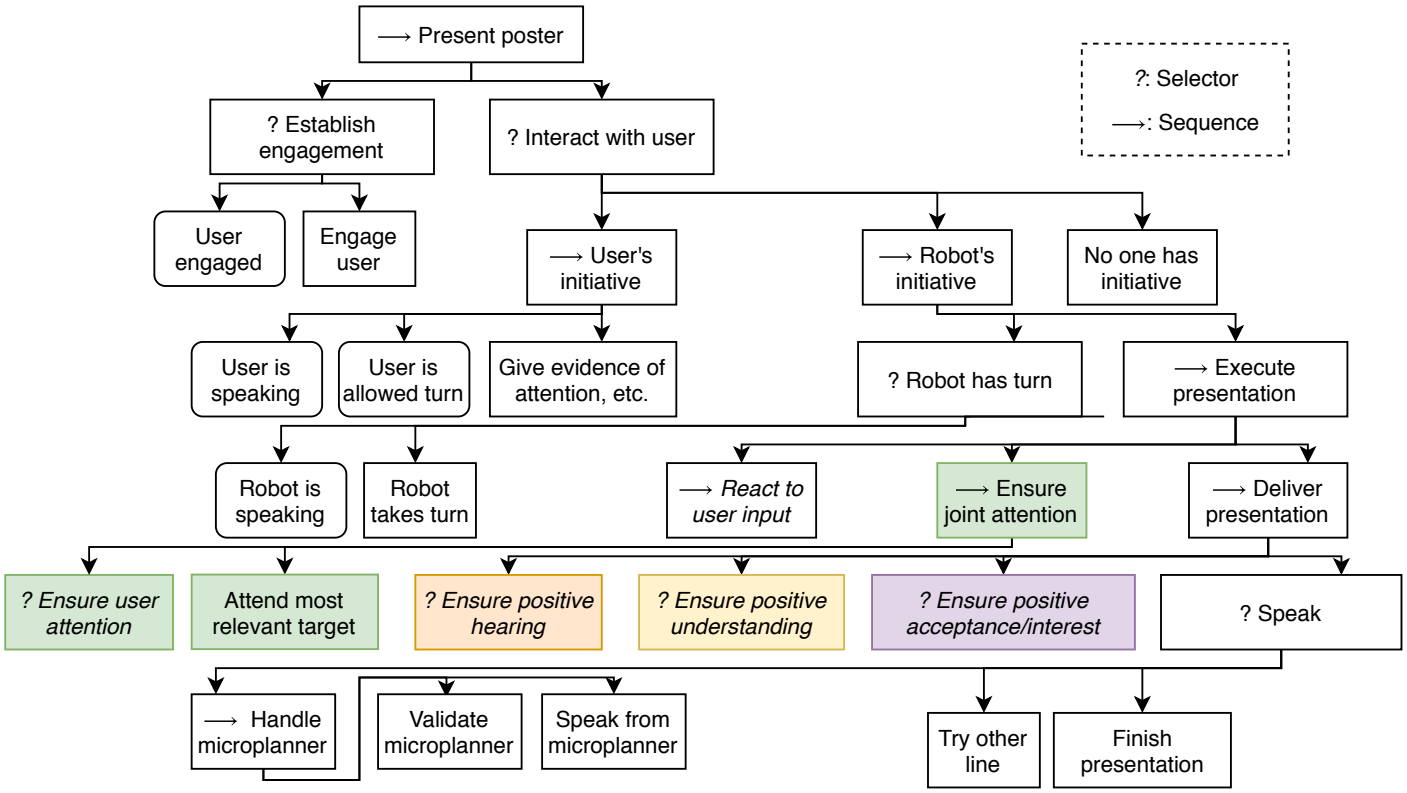

Fig. 7. The behaviour tree driving the interactions described in Section 5.1. For a brief explanation of the tree, see Section 5.1.1. The graphical representation of the tree is adapted from;[61] rounded nodes represent *queries* and rectangular nodes represent *actions*. Nodes with text in italics contain sub-trees that were omitted for space; for their implementation, see.[62]

the dialogue in terms of what is being spoken about and what is being said about the theme maps well to a knowledge graph. The rheme can be broken down into a verb-like string and a reference to a noun phrase through the knowledge graph. As the user gives or does not give feedback in real time (*'backchannels, implicit feedback'* in Figure 6), assumptions about the grounding state may be met or unmet, which changes which references to the knowledge graph are used, making the behaviour tree fill in the slots of the microplanner with alternative edges of the knowledge graph if possible. This serves as the system's primary short-term adaptation.

5.1.2. *Evaluation*

To evaluate the presentation robot, 30 test participants were recruited to interact with it as it presented a painting (as seen in Figure 5). We compared two versions of the system described as follows. In one version, the robot adapted the order of its presentation to the feedback given by the user. This feedback was classified by a Wizard of Oz. In the other version, the robot would always proceed through its presentation in the same way, ignoring any feedback from the user. Each participant interacted with both versions, using a within-group experiment design.

Subjective evaluations, performed using the Godspeed[68] and Social Presence[69] evaluations, showed that users consistently preferred the adaptive system.[70] When users were asked what the differences between the two systems were, they answered in a way that generally indicated that they did not pick up that one system was adapting to them and the other not, which implies that the differences in subjective evaluations were subconscious. We have also shown that users prefer adaptations that relate to the type of feedback that was given by the user.[62]

These results show that the types of feedback that flow over the blue edges labelled *'speech, explicit feedback'* and *'backchannels, implicit feedback'* are useful for creating an adaptive presentation system that adapts in a way that is actually preferred by users.

To specifically evaluate the use of a knowledge graph to adapt the presentation, an indirect study was performed over Amazon's Mechanical Turk crowdworker platform. A simulated user was implemented that interacted with the system in dialogues that followed certain pre-determined scripts. In these dialogues, the simulated user behaved in a predetermined manner where they would react with strong or weak positive feedback to the system's first presented line, and then react with negative hearing or understanding to the system's second line. After the user responded

negatively, the Mechanical Turk evaluators were given four options for how the system should express its repair to the utterance that had failed.

When the simulated user reacted with negative hearing, we found a significant ordering effect where repeating the failed utterance slowly was the most highly-ranked option. When the simulated user responded with negative understanding, an alternative reference that referred back to the first thing the system had said was most highly-ranked. This shows that users preferred different ways of repairing failed utterances depending on both previously grounded information and on the form of the feedback given by the user, in this case the simulated user.[62]

5.2. *Modeling non-verbal adaptation using behavior tree and reinforcement learning*

This section addresses co-adaptation in non-verbal human-robot interaction. Here, we focus learning and adaptation of system's behavior to implement robot proactive behavior as introduced in.[47] As an example, consider a human-robot setting in which the robot's behavior is continuously adapted using reinforcement learning to better coordinate with a human partner. Our scenario for exploring this form of adaptation is a collaborative human-robot packaging scenario in which the robot proactively helps the human by understanding the intention of the human partner by interpreting the body motion data captured by a body motion capture suit.

In this task setting, the human partner has two options to choose from: box type 1 or box type 2. Then, the human picks up the box and moves it to the robot. The box can be arbitrarily placed in two positions in front of the robot based on the human's preference. The robot observes the position and type of the box by scanning it. Depending on the type of box, the robot picks up the right item and puts it inside the box. Then the human can either ask for more items or wrap up the box. If more items are needed, the human puts an air bubble wrap into the box, and the robot puts more items inside the box. Otherwise, the human picks up the wrap tape, and the robot lifts the box to help the human to complete the wrapping. However, this plan adds unnecessary delays to the task execution since the robot behaves re-actively. The robot picks up the items for the right box, after the box is placed and scanned. However, the robot can gather this information well before the box is placed on the table. The learning problem we address here is to train a model that enables the robot to estimate the state of the human in the task by observing raw body motion data. Based on such modeling, the robot can make proactive action decisions. Proactive decisions are made as soon as the potential benefits of timely assistance outweigh the risk of incorrectly acting on incomplete observations which adds extra delays to the task completion to retract the incorrect action.

In such a setting, it is important for the robot to learn a general model of human body motion and adapt the model to each person as different people not only differ in size, but also in behavior and the way they complete a task.

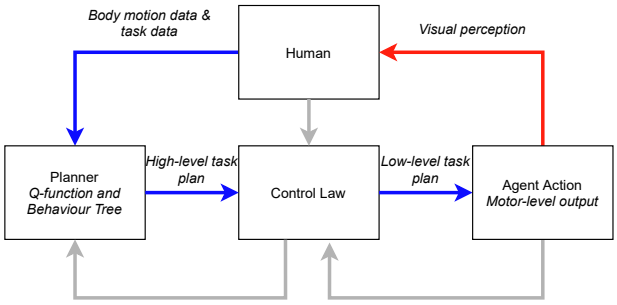

Fig. 8. Co-adaptation for Multi-Modal Interaction (non-verbal)

Figure 8 shows an adaptation of Figure 1 for the application of learning and adaptation of high-level human intention in non-verbal collaborative settings. The robot perceives non-verbal measures of human, in this example, human body motion data, representing the state of the person in the task. Besides, high-level task information is also provided to the robot to correct its behavior in cases the robot fails to be proactive or makes a wrong proactive decision. Such information is also used as the reward signal for the robot to adapt its behavior using reinforcement learning. The human partner and the robot in turn perform their share of the task and wait for the other partner to also complete the part of the task. The joint adaptation is done to enable the robot to perform proactive actions to reduce the total waiting time of the task. The human partner adapts to the robot by behaving such that the robot model can interpret the behavior more accurately.

5.2.1. *System components*

A **behavior tree** is integrated in our learning framework to structure the prior knowledge. The behavior tree is manually designed and it contains a complete plan for the robot to finish the task together with the human partner. However, the behavior tree performs the task completely reactively based on the feedback received from the task and does not include any trainable or adaptable parameter. Therefore, the performance of the system with the behavior tree alone is not optimal.

In our work,[47] we proposed to equip the behavior tree with a **Q-learning** component that receives any temporal measure of human body from different sensing modalities, e.g., body motion data or gaze information that comprise the information required to estimate the state of the person in the collaborative task. The Q-learning node is integrated into the behavior tree to make proactive action decisions. The node optimizes a Q-function that receives an action decision and multi-modal sensory observations, as well as the state of the behavior tree as the input, and recurrently

updates its internal state. It outputs a value to each pair of the internal state and the action decision. The action decisions come from an action set. For every action in the set, a trained Q-function assigns a value proportional to the expected amount of reduced waiting time when taking the proactive action decision at the given time-step compared to not taking any proactive decision. The state of a node of a behavior tree can be either running, success, or failure. The Q-learning node returns failure when it needs more observation to make a proactive decision, i.e., when the action "wait to collect more data" has the highest value. The Q-learning node changes its status to running when an action decision other than "wait to collect more data" has a high positive value given the multi-modal observation until the current time-step. Upon successful completion of the proactive action decision, the node returns success.

5.2.2. *Evaluation*

We evaluate the learning framework on a collaborative packaging task in which a person wears a Rokoko motion capture suit and collaborates with an ABB YuMi robot. In total, 7 students from the lab (4 males and 3 females in the age range 24 to 32 years old) participated in this experiment. We compare our method with a baseline method that trains a supervised classifier. The implementation of the baseline is described in the following.

Supervised learning baseline: Given the motion data as input, we trained a classifier that outputs the right robot action and a measure of uncertainty at every time-step. When the uncertainty is higher than a threshold, the output of the model is discarded and the command "wait to collect more data" is executed. Otherwise, the action given by the output of the model is executed. The entire network is trained end-to-end with supervised learning by providing the correct action labels for every piece of multi-modal sensory observations. The new node replaces the Q-learning node.

We evaluated the efficiency of our proposed RL based adaptation mechanism and compared it to the baseline provided a range of threshold values. Our proposed RL method outperforms the baseline approaches in terms of the average reward. The best performance for the baseline method is achieved by implementing the uncertainty measure using bootstrapping and dropout techniques and setting the threshold to 0.3 (please refer to[47] for complete details). In this case, the average reward is about 2.9 seconds which is considerably lower than the average reward given by the proposed RL framework (3.4 seconds). Besides, please note that the proposed RL framework does not require any supervision, while the baseline methods are trained using annotated motion data.

6. Challenges and Future Work

In this section we address remaining challenges and directions for future investigations to improve upon co-adaptive methods for human-robot co-existence.

6.1. *Human-in-the-loop Planning and Control with Safety Guarantees*

Despite the significant results presented, there are still open questions to be addressed for human-robot co-adaption in the context of planning and control. First, much of the presented work focuses on graph-based methods for developing safe-by-design plans that satisfy the high-level temporal logic specifications. One known shortcoming of these approaches is that they do not scale well with the number of states in the system. If the number of states in the transition system is too large, the development of such safe-by-design plans may be intractable. Future work needs to address how such plans can be computed for larger systems.

Furthermore, the existing methods presented here are dependent on a pre-defined task for the robot agents. In a co-adaptive interaction scenario, this task would be given and possibly modified online via communication with a human. Although natural language processing methods exist to convert speech to text, the temporal logic specifications themselves do not reflect true human speech. Temporal logic offers a rich method of specifying tasks, but the more complicated the task, the less "human-like" the specification. For example, humans do not typically speak in combinations of "always", "eventually", "until", "next", and "never". Future work will investigate methods to convert speech into correct-by-design task specifications to bridge the gap between spoken human commands and temporal logic specifications.

With regards to motion planning, much of the current framework was designed for homogeneous groups of robots with single integrator dynamics. Although some initial progress was made towards heterogeneous multi-agent planning, the method was more heuristic and does not yet extend to a co-adaptive framework for heterogeneous robots. Furthermore, the focus on motion planning for single-integrator dynamics is restrictive. Future work should address heterogeneous collaboration with more complex robot dynamics (e.g., double integrator, Euler-Lagrange) in the context of human-robot co-adaptation to further improve performance.

Finally, the human preference learning considers social acceptability from the perspective of satisfying the hard/soft tasks by letting the human emphasize to what degree the soft task should be satisfied. This is somewhat restrictive, and could be improved by also considering *how* a human would prefer a task to be satisfied. In many cases, there are multiple ways of satisfying high level tasks. The proposed algorithm seeks the optimal plan based on predefined weights on the edges of the transition system (e.g. reflecting transition time or distance). However this optimal plan may not be the human preference. Future work should address how the weights of the system should be updated online to consider human-preferred ways of satisfying the task.

6.2. *Co-adaptation and Trust*

In Section 4 we discussed the mechanisms that enable trusted co-adaptation processes and the related experimental studies on different co-adaptation mechanisms. However, there are still many areas for improvement. This is mainly related to three aspects. The first part is the definition of trust mechanisms. The trust we define at a generic level may not match well with the overall smooth communication process of human-robot interaction. It is likely that the feedback that satisfies the social needs of humans does not necessarily represent a trusting relationship between humans and robots at a higher level of meaning. In psychological terms, it is simply "basic trust". The study of higher-level trust relationships may require us to rethink and expand our definition of the human-robot interaction scenario.

The second aspect for improvement is the study of interaction patterns. A better interaction model should take into account all aspects of information in the interaction process. In previous studies, we have considered mostly a single model, but in future studies, multi-modality will be our main focus. This will include natural language, image, gaze, gestures, and other models related to the HRI process. Considering this aspect does not only mean that we will face more problems in data collection and deal with its complexity, but also our algorithms need to take in more inputs and sort out the relationship between them.

The third aspect is the study of the interaction algorithm itself. Most of the previous interaction algorithms for personalized adaptation used in HRI are based on general reinforcement learning algorithms, which are not optimized for the interaction modes and the co-adaptation required in the interaction modes. We will consider how to optimize the reinforcement learning algorithms to co-adapt faster during the HRI.

These three components form the focus of future work for co-adaptation and trust. They are also important problems to achieve better future human-robot interaction.

6.3. *Multi-Modal Interaction*

The solutions presented in Sections 5.1.1 and 5.1.2 address how feedback can be interpreted and mapped to an understanding of the user. What is not addressed is how the feedback that creates such an understanding is identified in the first place, and categorised as negative or positive feedback on various grounding levels. Buschmeier and Kopp have argued and created models for modelling user feedback over time as a Bayesian process where the user's feedback is combined with their previous state to create a new grounding state.[71] While an approach like this does not necessarily extend to our knowledge-aware scenario supported by a graph, since the grounding state would have to be quantified in terms of what parts of the knowledge graph a user is reacting to, the general sentiment that feedback is timing-dependent holds.

The modalities and types of feedback that are most important for a presentation agent that mostly holds the turn, taking responsibility for driving the interaction forward towards some goal, may be different from those that are important for a more open-ended conversational agent. Back-channel signals (short signals that do not aim to take the turn away, like "uh-huh", "mhm" or a short nod[72]) from the system towards the user are almost absent in the presentation scenario, but are crucial in more equal communication.

The knowledge graph representation presented in Section 5.1.1 has the advantage of generalising and offloading the generation of content to the knowledge on which the presentation is based. However, it fails to represent information that does not relate to grounding, but is also not quantifiable in terms of commonly accepted information. Attitudes towards facts or entities in the graph are not grounding information. A user must have accepted that the robot holds a piece of information to be true to have an opinion on it. Pitchl et al. address this issue by adding edges to their knowledge graph corresponding to information about the user or the system (which are added to the graph as nodes). However, they also represent what we would consider grounding information in this way, which is unlike our approach.[73]

In section 5.2, we presented a solution for learning and adaptation based on reinforcement learning for non-verbal human-robot collaboration settings. One limitation of the introduced method is that it requires manual design of the behavior tree. This limitation hinders the approach to adapt the action-selection policy to changes in the task without redesigning the behavior tree by an expert. An alternative way to this is to learn the behavior tree based on human user demonstrations. However, the main challenge is that data-collection processes which include humans are quite expensive. As a part of our future research, we will study the approaches that enable the robot to infer a task faster from human demonstrations.

Finally, in addition to future research directions in "Human-in-the-loop Planning and Control", "Co-adaptation and Trust", and "Multi-Modal Interaction", future work should also tackle the problem of combining each of these subsections together for an interconnected framework.

7. Conclusion

In this paper, we summarized a framework for co-adaptive human-robot co-existence. This body of work was the culmination of developments from the Swedish project *COIN: Co-adaptive human-robot interactive systems* by the Swedish Foundation for Strategic Research (SSF). We addressed three main concepts in the context of co-adaptation including safe planning/control of robot systems, trust, and multi-modal interaction. Regarding safe planning/control methods, we summarized a methodology that allows human input in both high and low-level stages of the plan-

ning/control setup and can adapt high-level tasks to human preference. We then discussed how to build multiple natural and trusted human-robot co-adaptation scenarios. By studying how to make the human-robot adaptation and interaction process smoother, we established different interactive algorithms that take human feedback into account in the robot's learning and decision-making process. Finally, verbal and non-verbal methods were addressed for co-adaptation in multi-modal interaction. For verbal-based approaches, we have shown how a robot can present information (e.g., related to a piece of art) to a human audience and adapt that presentation based on the verbal feedback it receives. We then addressed non-verbal interaction and proposed a reinforcement learning-based framework to anticipate and adapt to human/task objectives. The framework learns proactive behaviors by balancing between timely actions and the risk of making mistakes. The experiments show that this form of adaptation enables faster coordination between human and robot partners by eliminating unnecessary delays.

In addition to summarizing our methodologies, we have outlined avenues for future investigation regarding co-adaptation. With respect to safe planning/control, the next steps include extensions to heterogeneous systems, large-scale systems, addressing human-preferred methods of satisfying tasks, and translating human speech to temporal logic specifications. In the future implementation of large-scale robotic systems, effective computing is as important as physical human interaction and only if we are careful with the modeling process of human feedback, can we have the real human-in-the-loop thinking at the affective level. For multimodal presentation agents, the main question remains how to map feedback given by the audience to an internal representation of the user's level of understanding. Such a representation can in turn be used to adapt what the agent presents or how the agent presents it. A question we have not addressed here is what different types of feedback from the audience mean, i.e., if there are contexts where outwardly positive feedback has an internal negative meaning because of context, or vice versa, and if this changes depending on the presentation context (museum, classroom, etc.) or the number of people in the audience (one-on-one, a large classroom, a lecture hall, etc.). Future research will be focused on learning the behavior tree together with the Q-learning node and test the method on potentially more complex collaborative tasks.

Acknowledgments

This project was funded by the Swedish Foundation for Strategic Research (SSF).

References

[1] T. S. Tadele, T. de Vries and S. Stramigioli, The safety of domestic robotics: A survey of various safety-related publications, *IEEE Robotics Automation Magazine* **21**(3) (2014) 134–142.

[2] V. Villani, F. Pini, F. Leali and C. Secchi, Survey on human-robot collaboration in industrial settings: Safety, intuitive interfaces and applications, *Mechatronics* **55** (2018) 248–266.

[3] M. Hanafusa and J. Ishikawa, Mechanical impedance control of cooperative robot during object manipulation based on external force estimation using recurrent neural network, *Unmanned Systems* **08**(03) (2020) 239–251.

[4] M. Huo, H. Duan and X. Ding, Manned aircraft and unmanned aerial vehicle heterogeneous formation flight control via heterogeneous pigeon flock consistency, *Unmanned Systems* **09**(03) (2021) 227–236.

[5] J. Chen, Y. Ding, B. Xin, Q. Yang and H. Fang, A unifying framework for human-agent collaborative systems–part I: Element and relation analysis, *IEEE Transactions on Cybernetics* (2020) 1–14.

[6] C. E. Harriott, S. Garver and M. Cunha, A motivation for co-adaptive human-robot interaction, *Advances in Human Factors in Robots and Unmanned Systems*, ed. J. Chen *Advances in Intelligent Systems and Computing*, (Springer International Publishing, 2018), pp. 148–160.

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

[8] C. Baier and J. P. Katoen, *Principles of model checking* (MIT press, 2008).

[9] G. Jing, C. Finucane, V. Raman and H. Kress-Gazit, Correct high-level robot control from structured English, *IEEE International Conference on Robotics and Automation*, (2012), pp. 3543–3544.

[10] S. Andersson, A. Nikou and D. V. Dimarogonas, Control synthesis for multi-agent systems under metric interval temporal logic specifications, *IFAC-PapersOnLine* **50**(1) (2017) 2397–2402.

[11] S. G. Loizou and K. J. Kyriakopoulos, Automatic synthesis of multi-agent motion tasks based on ltl specifications, *2004 43rd IEEE Conference on Decision and Control (CDC)(IEEE Cat. No. 04CH37601)*, **1**, IEEE (2004), pp. 153–158.

[12] B. He, J. Lee, U. Topcu and L. Sentis, Bp-rrt: Barrier pair synthesis for temporal logic motion planning, *2020 59th IEEE Conference on Decision and Control (CDC)*, IEEE (2020), pp. 1404–1409.

[13] M. Srinivasan, S. Coogan and M. Egerstedt, Control of multi-agent systems with finite time control barrier certificates and temporal logic, *2018 IEEE Conference on Decision and Control (CDC)*, IEEE (2018), pp. 1991–1996.

[14] H. Kress-Gazit, G. E. Fainekos and G. J. Pappas, Temporal-logic-based reactive mission and motion planning, *IEEE Transactions on Robotics* **25**(6) (2009) 1370–1381.

[15] J. Tumova, A. Marzinotto, D. V. Dimarogonas and

D. Kragic, Maximally satisfying LTL action planning, *IEEE/RSJ International Conference on Intelligent Robots and Systems*, (2014), pp. 1503–1510.

[16] M. Guo, K. H. Johansson and D. V. Dimarogonas, Revising motion planning under linear temporal logic specifications in partially known workspaces, *IEEE International Conference on Robotics and Automation*, (2013), pp. 5025–5032.

[17] S. Sadraddini and C. Belta, Formal methods for adaptive control of dynamical systems, *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*, IEEE (2017), pp. 1782–1787.

[18] Y. Chen, J. Tůmová and C. Belta, LTL robot motion control based on automata learning of environmental dynamics, *IEEE International Conference on Robotics and Automation*, (2012), pp. 5177–5182.

[19] M. Hasanbeig, Y. Kantaros, A. Abate, D. Kroening, G. J. Pappas and I. Lee, Reinforcement learning for temporal logic control synthesis with probabilistic satisfaction guarantees, *2019 IEEE 58th Conference on Decision and Control (CDC)*, IEEE (2019), pp. 5338–5343.

[20] G. E. Fainekos, Revising temporal logic specifications for motion planning, *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, IEEE (2011), pp. 40–45.

[21] A. Girard and G. J. Pappas, Approximation metrics for discrete and continuous systems, *IEEE Transactions on Automatic Control* **52**(5) (2007) 782–798.

[22] J. A. Stork, C. H. Ek, Y. Bekiroglu and D. Kragic, Learning predictive state representation for in-hand manipulation, *IEEE International Conference on Robotics and Automation*, (2015), pp. 3207–3214.

[23] I. Leite, G. Castellano, A. Pereira, C. Martinho and A. Paiva, Empathic robots for long-term interaction, *International Journal of Social Robotics* **6**(3) (2014) 329–341.

[24] Y. Gao, W. Barendregt, M. Obaid and G. Castellano, When robot personalisation does not help: Insights from a robot-supported learning study, *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, IEEE (2018), pp. 705–712.

[25] R. Stower, N. Calvo-Barajas, G. Castellano and A. Kappas, A meta-analysis on children's trust in social robots, *International Journal of Social Robotics* (2021) 1–23.

[26] M. I. Ahmad and O. Mubin, Emotion and memory model to promote mathematics learning-an exploratory long-term study, *Proceedings of the 6th International Conference on Human-Agent Interaction, Xi'an, China*, (2018), pp. 214–221.

[27] G. Gordon, S. Spaulding, J. K. Westlund, J. J. Lee, L. Plummer, M. Martinez, M. Das and C. Breazeal, Affective personalization of a social robot tutor for children's second language skills, *Thirtieth AAAI Conference on Artificial Intelligence, Phoenix, Arizona USA, Vol. 30. No. 1*, (2016), pp. 65–75.

[28] A. Ramachandran and B. Scassellati, Fostering learning gains through personalized robot-child tutoring interactions, *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction Extended Abstracts*, ACM (2015), pp. 193–194.

[29] A. Ramachandran and B. Scassellati, Adapting difficulty levels in personalized robot-child tutoring interactions, *Proceedings of the Workshops at the 28th AAAI Conference on Artificial Intelligence, Québec, Canada*, (2014).

[30] A. Jones, S. Bull and G. Castellano, "i know that now, i'm going to learn this next" promoting self-regulated learning with a robotic tutor., *International Journal of Social Robotics* (2017) 10(4), 439–454.

[31] A. Jones and G. Castellano, Adaptive robotic tutors that support self-regulated learning: A longer-term investigation with primary school children, *International Journal of Social Robotics* (2018) 10(3), 357–370.

[32] F. Yang, Y. Gao, R. Ma, S. Zojaji, G. Castellano and C. Peters, A dataset of human and robot approach behaviors into small free-standing conversational groups, *PloS one* **16**(2) (2021) p. e0247364.

[33] F. Yang and C. Peters, Appgan: Generative adversarial networks for generating robot approach behaviors into small groups of people, *2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, IEEE (2019), pp. 1–8.

[34] Y. Gao, S. Wallkötter, M. Obaid and G. Castellano, Investigating deep learning approaches for human-robot proxemics, *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, IEEE (2018), pp. 1093–1098.

[35] S. Roy, E. Kieson, C. Abramson and C. Crick, Using human reinforcement learning models to improve robots as teachers, *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, ACM (2018), pp. 225–226.

[36] H. H. Clark and S. E. Brennan, Grounding in communication., *Perspectives on socially shared cognition*, (American Psychological Association, 1991), pp. 127–149.

[37] H. H. Clark and D. Wilkes-Gibbs, Referring as a collaborative process, *Cognition* **22**(1) (1986) 1–39.

[38] N. Lubold and H. Pon-Barry, Acoustic-prosodic entrainment and rapport in collaborative learning dialogues, *Proceedings of the 2014 ACM workshop on Multimodal Learning Analytics Workshop and Grand Challenge*, (2014), pp. 5–12.

[39] A. Ogan, S. Finkelstein, E. Walker, R. Carlson and J. Cassell, Rudeness and rapport: Insults and learning gains in peer tutoring, *International Conference on Intelligent Tutoring Systems*, Springer (2012), pp. 11–21.

[40] J. Lopes, M. Eskenazi and I. Trancoso, From rule-based to data-driven lexical entrainment models in spoken dialog systems, *Computer Speech & Language*

31(1) (2015) 87–112.

[41] G. Skantze, A. Hjalmarsson and C. Oertel, Turn-taking, feedback and joint attention in situated human–robot interaction, *Speech Communication* **65** (2014) 50–66.

[42] H. Clark, *Using language* (Cambridge University Press, Cambridge, UK, 1996).

[43] E. Sibirtseva, A. Ghadirzadeh, I. Leite, M. Björkman and D. Kragic, Exploring temporal dependencies in multimodal referring expressions with mixed reality, *International Conference on Human-Computer Interaction*, Springer (2019), pp. 108–123.

[44] J. Bütepage, A. Ghadirzadeh, Ö. Ö. Karadağ, M. Björkman and D. Kragic, Imitating by generating: Deep generative models for imitation of interactive tasks, *Frontiers in Robotics and AI* **7** (2020).

[45] J. Butepage, M. J. Black, D. Kragic and H. Kjellstrom, Deep representation learning for human motion prediction and classification, *Proceedings of the IEEE conference on computer vision and pattern recognition*, (2017), pp. 6158–6166.

[46] J. Bütepage, H. Kjellström and D. Kragic, Anticipating many futures: Online human motion prediction and generation for human-robot interaction, *2018 IEEE international conference on robotics and automation (ICRA)*, IEEE (2018), pp. 4563–4570.

[47] A. Ghadirzadeh, X. Chen, W. Yin, Z. Yi, M. Bjorkman and D. Kragic, Human-centered collaborative robots with deep reinforcement learning, *IEEE Robotics and Automation Letters* (2020).

[48] A. Ghadirzadeh, J. Bütepage, A. Maki, D. Kragic and M. Björkman, A sensorimotor reinforcement learning framework for physical human-robot interaction, *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE (2016), pp. 2682–2688.

[49] A. Czeszumski, A. L. Gert, A. Keshava, A. Ghadirzadeh, T. Kalthoff, B. V. Ehinger, M. Tiessen, M. Björkman, D. Kragic and P. König, Coordinating with a robot partner affects action monitoring related neural processing, *bioRxiv* (2021).

[50] H. Kress-Gazit, G. E. Fainekos and G. J. Pappas, From structured english to robot motion, *2007 IEEE/RSJ International Conference on Intelligent Robots and Systems*, (2007), pp. 2717–2722.

[51] H. Kress-Gazit and G. J. Pappas, Automatically synthesizing a planning and control subsystem for the darpa urban challenge, *2008 IEEE International Conference on Automation Science and Engineering*, (2008), pp. 766–771.

[52] M. Guo, S. Andersson and D. V. Dimarogonas, Human-in-the-loop mixed-initiative control under temporal tasks, *2018 IEEE International Conference on Robotics and Automation (ICRA)*, (2018), pp. 6395–6400.

[53] S. Ahlberg and D. V. Dimarogonas, Human-in-the-loop control synthesis for multi-agent systems under hard and soft metric interval temporal logic specifications, *2019 IEEE 15th International Conference on Automation Science and Engineering (CASE)*, IEEE (2019), pp. 788–793.

[54] S. Ahlberg and D. V. Dimarogonas, Mixed-initiative control synthesis: Estimating an unknown task based on human control input, *3rd IFAC Workshop on Cyber-Physical & Human Systems Beijing, December 3-5, 2020*, (2020).

[55] A. Y. Gao, W. Barendregt and G. Castellano, Personalised human-robot co-adaptation in instructional settings using reinforcement learning, *IVA Workshop on Persuasive Embodied Agents for Behavior Change: PEACH 2017, August 27, Stockholm, Sweden*, (2017).

[56] E. R. Hilgard, The trilogy of mind: Cognition, affection, and conation, *Journal of the History of the Behavioral Sciences* **16**(2) (1980) 107–117.

[57] P. A. Kragel and K. S. LaBar, Decoding the nature of emotion in the brain, *Trends in cognitive sciences* **20**(6) (2016) 444–455.

[58] C. Finn, P. Abbeel and S. Levine, Model-agnostic meta-learning for fast adaptation of deep networks, *International Conference on Machine Learning*, PMLR (2017), pp. 1126–1135.

[59] J. Schulman, S. Levine, P. Abbeel, M. Jordan and P. Moritz, Trust region policy optimization, *International conference on machine learning*, PMLR (2015), pp. 1889–1897.

[60] M. Salem, G. Lakatos, F. Amirabdollahian and K. Dautenhahn, Would you trust a (faulty) robot?: Effects of error, task type and personality on human-robot cooperation and trust, *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, ACM (2015), pp. 141–148.

[61] M. Colledanchise and P. Ögren, *Behavior trees in robotics and AI: An introduction* (CRC Press, 2018).

[62] N. Axelsson and G. Skantze, Using knowledge graphs and behaviour trees for feedback-aware presentation agents, *Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents (IVA '20)*, University of Glasgow, Scotland & Online (2020).

[63] D. Vrandečić and M. Krötzsch, Wikidata: a free collaborative knowledgebase, *Communications of the ACM* **57**(10) (2014) 78–85.

[64] H. Liu, T. Lin, H. Sun, W. Lin, C.-W. Chang, T. Zhong and A. Rudnicky, Rubystar: A non-task-oriented mixture model dialog system, *arXiv preprint arXiv:1711.02781* (2017).

[65] R. Jalota, P. Trivedi, G. Maheshwari, A.-C. N. Ngomo and R. Usbeck, An approach for ex-post-facto analysis of knowledge graph-driven chatbots–the dbpedia chatbot, *International Workshop on Chatbot Research and Design*, Springer (2019), pp. 19–33.

[66] Q. Chen, J. Lin, Y. Zhang, M. Ding, Y. Cen, H. Yang and J. Tang, Towards knowledge-based recommender dialog system, *arXiv preprint arXiv:1908.05391* (2019).

[67] W. J. Levelt, *Speaking: From intention to articulation* (MIT press, 1993).

[68] C. Bartneck, D. Kulić, E. Croft and S. Zoghbi, Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots, *International journal of social robotics* **1**(1) (2009) 71–81.

[69] F. Biocca and C. Harms, Networked minds social presence inventory (scales only version 1.2), *East Lansing: MIND Labs, Michigan State University. Retrieved from http://cogprints. org/6742* (2011).

[70] N. Axelsson and G. Skantze, Modelling adaptive presentations in human-robot interaction using behaviour trees, *Proceedings of the 20th Annual SIGdial Meeting on Discourse and Dialogue*, Stockholm, Sweden (2019), pp. 345–352.

[71] H. Buschmeier and S. Kopp, Towards conversational agents that attend to and adapt to communicative user feedback, *International Workshop on Intelligent Virtual Agents*, Springer (2011), pp. 169–182.

[72] V. H. Yngve, On getting a word in edgewise, *Papers from the sixth regional meeting of the Chicago Linguistics Society*, University of Chicago, Department of Linguistics, Chicago (April 1970), pp. 567–578.

[73] J. Pichl, P. Marek, J. Konrád, P. Lorenc, V. D. Ta and J. Šedivỳ, Alquist 3.0: Alexa prize bot using conversational knowledge graph, *arXiv preprint arXiv:2011.03261* (2020).