Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 192 (2021) 3487–3496

Procedia Computer Science

www.elsevier.com/locate/procedia

25th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems

Explainability in Human-Robot Teaming

Maryam Banitalebi Dehkordi[a,*], Reda Mansy[a], Abolfazl Zaraki[a], Arpit Singh[a], Rossitza Setchi[a]

[a]*Research Centre in AI, Robotics and Human-Machine Systems (IROHMS), Cardiff University, Cardiff CF24 3AA, UK*

Abstract

In human-robot teaming, one of the crucial keys for the team's success is that the robot and human teammates can collaborate accordingly in a coordinated manner. Each teammate should be aware of what the other teammate is going to perform and likely to need. In this context, a robot is expected to understand human teammate intention and performance and explain its actions and decisions and its rationale to its teammate. In addition, the capability to model the expectation of a human teammate empowers the robot to collaborate with human understandably and expectedly, leading to effective teaming. Through forming mental modelling, the robot can understand the impact of its own behaviour on the mental model of the human. In addition, the desirable traits in human-robot teaming, including fluent behaviour, adaptability, trust-building, effective communication, and explainability, can be achieved through mental modelling. In this work, we introduce a scenario for human-robot teaming considering all the five desirable traits in teaming with the main focus on explainability and effective communication. Using a general model reconciliation, the expectation of the human teammate of the robot can be modelled, and the explanation can be generated. In a considered scenario including Care-O-bot 4 service robot and a human teammate, we assume that the robot detects the human's current task (analysing his body gesture) and predicts his following action and his expectation from the robot. In a reciprocal interdependence task, the robot coordinates his behaviour and acts accordingly by picking up the relevant tool. Through explanation and communication robot further offers the outcome of his decision to the human teammate and adapts its action by handing the tool to the human upon his desire.

Keywords: Human-robot teaming; mental model; explanation; expectation; trust.

1. Introduction

Human-robot teaming is the collaboration between humans and robotic systems whereby they work together to perform a joint activity [18]. Humans and robots traditionally worked separately in manufacturing and related fields, even when there was potential for collaborative work. This has limited the potential of what the robots can do as they were used only for repetitive tasks. Due to the rise of autonomous systems, the collaboration between robots and

* banitalebidehkordim@cardiff.ac.uk

10.1016/j.procs.2021.09.122

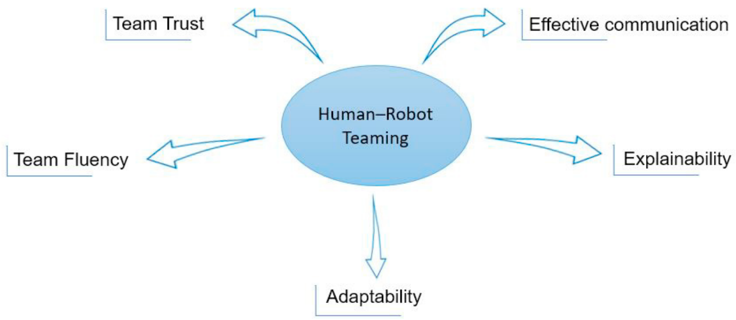

Fig. 1. Desirable traits that can be achieved through mental modeling in human-robot teaming [27].

humans becomes more accessible and possible, as the robots' ability to interact with humans is improved [27]. This brings into question the robot's awareness of the human while carrying out collaborative tasks [32].

For this to be done, it is important to consider mental modelling methods in order to obtain effective teaming between a robot and human teammates. Through mental modelling, a robot can behave as a trustworthy teammate in collaboration with human. Mental models are mental representations that humans use in order to interact with and predict events in their environment and can be extended to robots for effective teaming [27]. Furthermore, it is vital that human-robot interactions occur in an explainable and logical manner as these characteristics are shown to be successful in human-human teaming [5]. Similarly, robotic teammates need to not only understand their human partners but also provide explanations for their own decisions and behaviours when necessary [31]. In a team, explanations help to build shared situational awareness, thus strengthening the trust between teammates [31]. Explainability in human-robot teaming is especially important as robots are geared towards solving problems that require high computational abilities that humans do not possess. And in order to achieve effective teaming and to build trust, explainability is necessary for the human to understand their robotic teammate [31].

How tasks are shared between robots and humans is also necessary to be considered for effective teaming. Zhao et al. [37] explored the effects of different levels of task interdependence with reference to human-robot teaming. Task interdependence can be defined as the effect one's behaviour can have on the performance of others in their team and the extent to which team members interactions affect their tasks. The levels outlined are pooled, sequential, and reciprocal, with findings suggesting that the reciprocal level is ideal for human-robot collaboration as measured by the participants' decreased stress levels, more efficient output, and their view of the robot as a collaborator rather than a tool to complete tasks. Although the outcome of this work is beneficial for human-robot teaming, they did not consider the theory of mind and mental modelling techniques which are crucial in effective team working. Tabrez et al. [27] posit that mental models can be used to create an effective team working, as members can predict what their teammates need, which further facilitates the coordination of their tasks. This is the basis of the Shared Mental Model (SMM), which helps describe, explain, and predict a team's behaviour. It further outlines desirable traits in team working that can be achieved through mental modelling in human-robot teaming as fluent behaviour, adaptability, trust-building, effective communication, and explainability (see Figure 1).

In this work, we first present crucial aspects in human-robot teaming, including *mental models*, *plan explanation* and *task interdependence*, and then taking into account these concepts, we present our customised conceptual model for human-robot teaming. We then introduce our designed scenario, with the main focus on an explanation between the robot and its human teammate. We report the separate semi-autonomous modules of our framework and the current progress on track to the final system integration and autonomous human-robot teaming. We considered using a Vicon motion tracking system for human activity recognition and the Care-O-bot 4 service robot as a human teammate in the teaming scenario. The idea is that the robot should be able to detect the human teammate's current task and predict the following action and coordinate its behaviour accordingly. The robot needs to explain and describe its decisions and actions to the human teammate to generate a shared awareness by creating mental modelling.

2. Human-Robot Teaming

This section describes three important aspects in human-robot teaming, including mental models, plan explanation and task interdependence, following which in section 2.4, we present our conceptual model for human-robot teaming in this work.

2.1. Mental models

To enable a fluent and effective human-human teamwork, an important aspect is to coordinate actions among teammates. For that, each teammate should be aware of what others are going to do and likely to need. Through communication, humans can effortlessly understand other teammates plans and desires [16]. Contrary to humans, due to the lack of intuition in robots, they require mathematical models and algorithms to obtain an approximate mental model of their human teammates and plan and behave accordingly.

Mental models, or as called in psychology *mental representations*, are categorised knowledge structures that enable humans to interact with their surrounding real-world environment [30, 27]. Mental models play an important role in assisting humans in relating, explaining, and also predicting real-world environmental events [20]. They are powerful and effective in supporting teamwork fluency for the execution of complex tasks in different applications [19].

Mental models usually appear reciprocally with Theory of Mind (ToM) in developmental psychology [27]. ToM is defined as the capability of attributing opinions, intentions, targets and preferences to others [23]. ToM is a fundamental concept of human-human social interactions, following which in human-robot interaction and teaming, the aspects of ToM are integrated into various architectures [17, 22].

In human-robot teaming, mental models and ToM are correlated as the robots making a model of other teammates means it is able of forming a ToM. On the other hand, the human teammate directs a ToM to the robot, which in turn leads to the robot creating a mental model of the human's mental model of the robot [27]. In contrast to first-order (standard) mental modelling, this is known as second-order mental modelling and outlines the impact of robot's behaviour on the mental model of human [4]. Literature shows mental models and team performance are mutually affected each other, where the similarity among the teammates' mental models can be resulted in improving the team performance [3, 20]. In other words, when team members can detect other members intentions and expectations (through predicting other members actions and requirements), they can boost the concurrence and compatibility of their actions. Therefore, team members should have a shared understanding, and accordingly, a shared mental model [27]. The shared mental model (SMM) posits that in order for a team to be successful, team members need to have comparable mental models, which results in similar expectations for performing shared tasks [7, 13]. In summary, as stated in [27] *"if a mental model helps in describing, explaining, and predicting the behaviour of a system, a shared mental model serves the purpose of describing, explaining, and predicting the behaviour of a team"*.

Various studies in human-robot collaboration have used mental models to extract the positive qualities of the team working (i.e. decision making in crucial situations). A survey by Tabrez et al. [27] described five categories of preferable attributes that can be attained through mental modelling in human-robot teaming:

Team Fluency: Hoffman [10] described Fluency as a "*coordinated meshing of joint activities between members of a well-synchronized team*" [27]. To measure fluent behaviour among team members, several subjective metric scales and objective measures can be employed (i.e. the idle time of human and robot teammates, the time lost between one team member performing a task and another member continuing that action). As demonstrated by [11] when the robot predicts the action of other teammates, no improvement in the efficiency of the task can be detected; however, the sense of fluency in human users was raised.

Adaptability: In a real-world human-robot teaming, all teammates are required to be able to adapt their plans and actions accordingly depends on the team's current status and plan. For this reason, in order to enable a prompt adaptation among team members, shared mental models are leveraged to alter the requirements of the task [7, 21].

Team Trust: Trust is one of the critical factors that should exist among teammates. Trust can be measured in different ways, for example, in health care scenarios, it can be measured through the willingness to cooperate, physical/visual contact, and emotional feedback [24]. Literature indicates through a shared mental model, a human can trust a robot teammate: when they have no concern about the robot's capabilities and have a clear view of the robot's decision making methods [1, 26].

Effective communication: To enable a human-robot teaming, all members should be able to communicate either verbally or non verbally. Klingspor et al. [14] stated that human-robot communication should be enhanced in order to allow humans to take advantage of robots. Through communication, teammates can inform each other whether they need assistance to accomplish a task [25, 28].

Explainability: As mentioned in our previous work [24] an explanation is a set of statements containing beliefs, facts, context, clarifications of causes and possible consequences. A rapidly emerging research area of explainable AI has shown the importance of converting a black box into a glass box and explaining how AI models work and how their decisions are made, which can increase the trust, transparency, security, and confidence among teammates, which in turn can improve the team performance [12].

2.2. Plan explanation generation

In human-robot teaming, the goal is to create the interaction in the same way as in human-human teaming in a coherent manner [5, 8]. A robot teammate is expected to understand other humans teammates [34, 35] in addition to having the capability of explaining its actions or the decisions it made. In a team working, explanation results in creating a shared situation awareness and boosting trust among team members [9, 31].

Considering both models of *explainee* and *explainer*, Chakraborti et al. [6] introduced a general model reconciliation in order to generate explanation. They demonstrated that the explanation in human and AI interaction is best studied considering the differences between their models. Assume M^R is the model of the robot, based on which the robot behaves and generates an optimal plan π_M^R (see Figure 2). Human teammates, in turn, interpret this model based on their model M^H, and the explanation will be required if π_M^R is different from what human expected. Following that, Zakershahrak et al. [31] stated that using M^H human teammates develop their expectations of robot's actions, which is in fact generated based on M^R and therefore it is different from M^H. It means that π_R^M which is generated with respect to M^R might be different from π_H^M which is generated with respect to M^H. Again anytime these two plans are not similar, the robot needs to explain its plan to the human teammate.

The capability of modelling the expectation of a human teammate empowers the robot to collaborate with other members in an understandable and expected manner [15], which leads to an effective teaming [8]. To enable that, according to [32] one of the main challenges is that the robot needs to learn the preconceptions of human teammates about the robot's model (see Figure 2). Zhang et al. [36] assumed that human teammates are able to understand the behaviour of other teammates through relating abstract tasks with teammate's actions [32]. If a human teammate's expectation differs from the actions of the robot, the human member can not relate some of its behaviours with the labels of the tasks. In order to learn the labelling process, the Conditional Random Fields (CRFs) can be employed [32]. Afterwards, utilising the learned model, a new plan of the robot can be labelled to calculate its explicability score [36].

2.3. Task interdependence

In human-robot teaming, another essential aspect is task interdependence between a robot and a human teammate. Zhao et al. [37] outlines three levels of task interdependence and explores their effectiveness in human-robot teaming. In pooled interdependence, teammates complete tasks separately and with minimal interaction with one another. The actions of a team member do not directly affect others in their team; as such, this is considered to be of low task interdependence. In the sequential level, teammates carry out tasks together, however, in a specific order or sequentially. This means that teammates rely on each other heavily. This is because the order in which they do tasks directly affects their teammates, but there is no focus on coordination between them. Therefore, this is considered to be a moderate level of task interdependence. At the reciprocal level, each teammate has a speciality in a specific skill and can carry out certain aspects of the task. This means portions of tasks can be done in no specific order, and teammates are required to coordinate together to achieve their goals. This means that the reciprocal level is considered to be of high task interdependence. Teaming at a high interdependence level has shown to be successful between human teammates in terms of goal commitment, and therefore performance [2]. This can be extended to human-robot teammates in terms of increased performance as well as the human's view of the robot as a teammate and collaborator. Through an experimental study, Zhao et al. [37] compared these three levels of interdependence in human-robot teaming and proved that higher interdependence groups performed better as expected and felt less

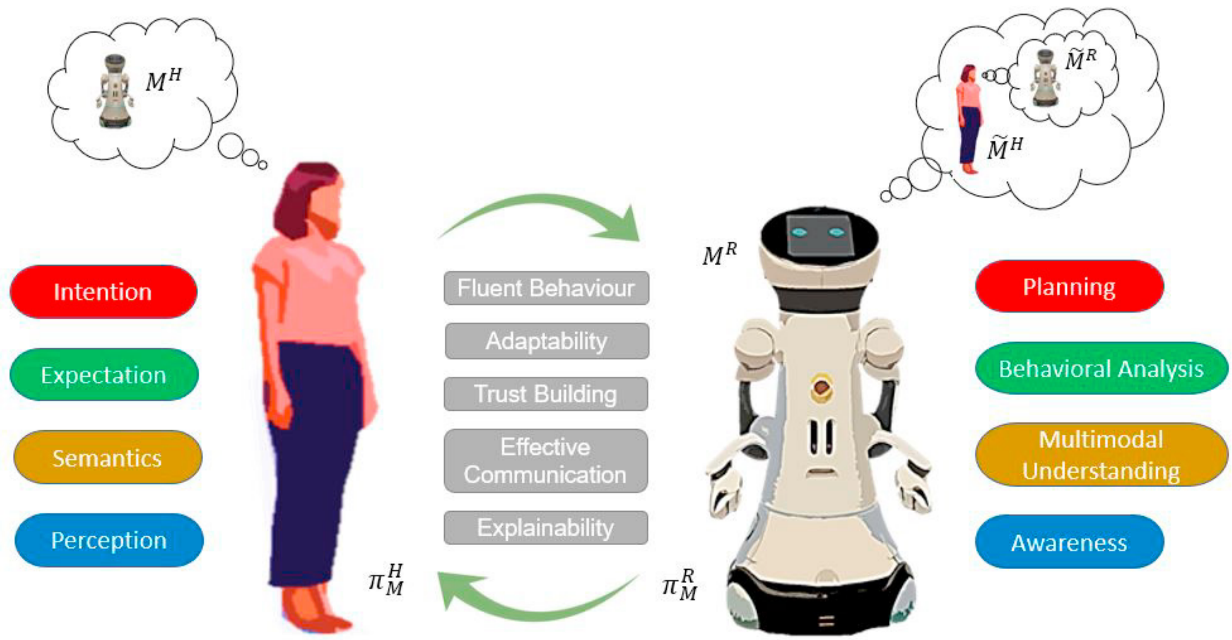

Fig. 2. A human-robot teaming model considering five desirable traits in human-robot teaming, which can be achieved through mental modelling [27]. Robot behaves and generates a plan π_M^R based on its model M^R and an approximate human model $\tilde{M}^H$. Human teammate interprets this plan model using their model M^H and an approximate robot model $\tilde{M}^R$ [32]. An explanation is required whenever π_M^R is different from what human expected [6].

stressed than their counterparts. Although the difference in stress levels was not statistically significant, qualitatively, participants consistently viewed the robot as merely a tool in the lower interdependence levels and as a collaborator and teammate in the higher interdependence level. They felt as though they were actively working with the robot towards a shared goal, with the robot contributing to the success of the task. It can be seen from this study that higher interdependence can significantly improve human-robot teaming and enables humans to view the robot as part of their team, which is crucial for effective collaboration and successful teaming.

2.4. Conceptual model

Figure 2 presents our conceptual model for human-robot teaming. Following the concepts of mental modelling and desirable traits in teaming which are outlined in [27] and section 2.1, we consider enabling fluent behaviour, adaptability, trust-building, effective communication, and explainability in our human-robot teaming framework. Following our previous work on explainable robotics, [24] our main focus will be on explainability and effective communication, which in turn can increase trust among teammates.

To achieve the team's targets, the robot needs to be proactive in collaborating with human teammates. Therefore, employing the algorithms of plan recognition through observations, the robot should be aware of the human teammate's intention, expectations and takes actions accordingly

Finally, following [37] as the task interdependence conditions, we consider the reciprocal interdependence case, in which the robot and its human teammates are in charge of accomplishing certain aspects of the task accordingly. We will consider the goal scenario in which the robot needs to autonomously detect the human current task and intention. It then coordinates his behaviour and performs a piece of the task to accomplish the team target. The robot needs to predict human intention, following actions and expectations from the robot in order to adapt its decision and actions. Through explanation and communication,the robot needs to continually exchange its plans and decisions with human teammates, which will increase the team fluency and trust.

Fig. 3. Vicon motion tracking system mounted all around the laboratory.

3. Proof of Concept

Our final scenario will be the case that the robot proactively collaborates with human teammates at a reciprocal interdependence level. Through observation, the robot needs to detect and predict human teammate intention and their expectation from the robot, and then make a decision and act accordingly. It then needs to explain its plans and decisions to the human teammates.

This section presents the infrastructure of this scenario and the current progress on developing the semi-autonomous modules in line with the final goal of autonomous human-robot teaming. We used a Vicon motion tracking system for human gestures and activity monitoring and Care-O-bot 4 robot as the human's teammate in a teaming scenario.

3.1. Vicon motion tracking system for human activity recognition

In our scenario, we consider using an optical motion tracking system instead of robot's built-in cameras or external RGB-D sensor (e.g. [33]). The main advantage of such a system is that the human teammate should not be necessarily in the field of view of the robot, which enables a reciprocal interdependence teaming. In fact, in this way, a human can even perform the task in a separate room while the robot detects its actions and coordinates its behaviour to accomplish the complementary task to achieve the team target. This teaming has several potentials in healthcare and manufacturing (i.e. when the robot works in a contaminated area and humans collaborate from a safe environment).

In the human-robot interaction laboratory of IROHMS, the Vicon motion tracking system is installed on top of the walls all around the room, as shown in Figure 3. The system includes 10 Vero cameras (2.2 MPX @ 330 FPS) and 1 Vue camera (60 Hz and 1080p resolution for video capturing), using which the labelled objects can be precisely detected in the entire area of the room. The recorded data can be further analysed using Vicon Nexus software. Figure 5 shows the view from the Vue camera, in which a human is waving in front of the robot.

Figure 5.a shows how markers can be placed on the human user's body for full body skeleton detection. The detected skeleton of the human user is shown in Figure 5.b. In this work, we focus on the detection of hands movements, in which we assume that the robot aims to detect one of these defined three tasks: the first task is mainly performed using the right hand, the second task using the left hand, and the third task using both hands. Figure 5.c shows the detected hands of the user (right hand in the green and left hand in red). As shown, the user is waving his right hand, which includes significant motion by the right hand and slight motion by the left hand. The system will receive the centroid and orientation of both right and left hands, evaluating which the robot can detect the first action is performing by the user, and it will respond accordingly.

3.2. Human-robot teaming test

To illustrate human-robot teaming, we considered a scenario that involved Care-O-bot 4 and a human teammate. Figure 6 shows the scenario in which a human (Reda) is repairing a hardware piece in the existence of his robot teammate. We assume that the robot detects the human task and his current action: *applying thermal paste*. The robot then predicts the human succeeding action/intention as: *screwing*. Afterwards, it estimates Reda's expectation from the robot for that action: *handing in the screwdriver*. Therefore, among the available tools on the worktop, a robot picks up a screwdriver. The following action for the robot is to explain his plan and decision to Reda. Through communication with human, a robot can adapt his plan and behaves accordingly to the approximate mental model of a human teammate.

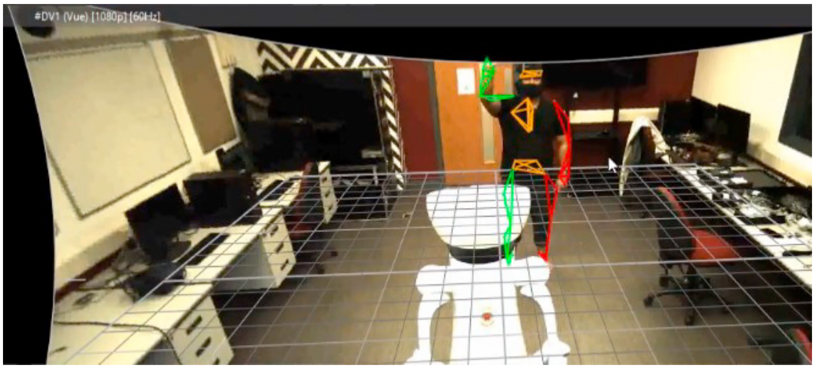

Fig. 4. The view from Vicon VUE camera, in a human-robot teaming scenario.

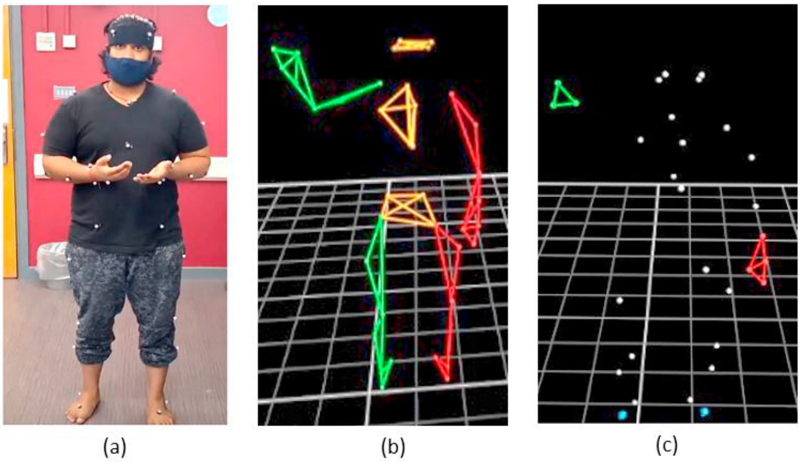

Fig. 5. Vicon motion tracking system. (a) human users wears Vicon markers over his body. (b) full skeleton tracking of the user. (c) tracking of the users hands.

Figure 7 illustrates the communication between the robot and human teammates. As shown robot explains its plan about handing in the screwdriver to the human. It then coordinates its action base on the human desire on whether or when to take the tool from the robot. The robot is capable of real-time speech recognition [29]. Specifically, it can detect “No”, following which the robot detects it is not a proper action/time to pass a screwdriver to the human. The second crucial word is “Need”, following which the robot offers the outcome of its decision making (passing screwdriver) to the human user. Once the robot detects “yes” it proceeds to hand the tool to the human. Detecting the “Thank you” robot returns to its original position, ready to assist in the next task.

In the current implementation of the scenario, there were issues with the accuracy of word detection, as closely related words such as “now” and “no” were often confused by the software, therefore triggering parts of the code that were unintended. This may have been due to the quality of the built-in microphone, as testing on an external microphone on a PC yielded more accurate output (it will be further investigated in the future development).

4. Conclusion

In human-robot teaming, each teammate should be able to understand other teammate actions and expectations. For this reason, a robot needs to detect and predict humans’ action, intention, and expectations from the robot. Using this capability, the robot can coordinate its actions with the human teammate, which results in effective teaming. The preferable attributes in human-robot teaming (fluent behaviour, adaptability, trust-building, effective communication,

Fig. 6. In a human-robot teaming scenario, Care-O-bot 4 offers using of screwdriver to the human teammate.

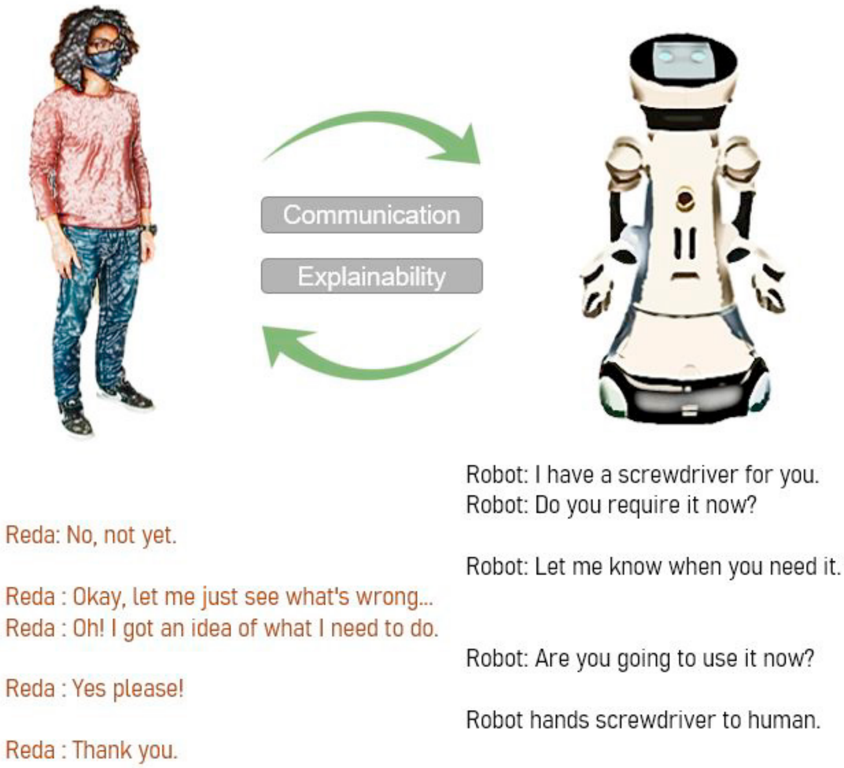

Fig. 7. The communication between human and robot, in which Care-O-bot 4 explains its plan to Reda and coordinates its actions accordingly.

and explainability) can be achieved through mental modelling. In this work, we initiated a human-robot teaming scenario considering desirable traits in teaming, focusing on explainability and effective communication. For that, employing a general model reconciliation, the robot can obtain an approximate model of the human expectation of the robot and generate an explanation to behave accordingly. The preliminary test reported a teaming between Care-O-bot 4 and a human teammate, in which we assumed the robot detected the human teammate is repairing a hardware piece using thermal paste. The robot then predicts the human's following action (using a screwdriver) and his expectation from the robot (passing the screwdriver to the human). The robot explained its decision and planned to the user and adapted its behaviour through communication with humans. In the future development, we aim to increase the level of robot's autonomy in which the robot will be able to identify different objects, detect more human actions, and behave accordingly. In addition, we are going to use the Xsens 3D motion tracking system, which allows human-robot teaming in the real-world setup (e.g. manufacturing environment), where we have no control over the environmental factors.

Acknowledgements

The authors would like to acknowledge the support of the Centre for Artificial Intelligence, Robotics and Human-Machine Systems (IROHMS), operation C82092, part-funded by the European Regional Development Fund (ERDF) through the Welsh Government.

References

[1] Arnold, M., Bellamy, R.K., Hind, M., Houde, S., Mehta, S., Mojsilović, A., Nair, R., Ramamurthy, K.N., Olteanu, A., Piorkowski, D., et al., 2019. Factsheets: Increasing trust in ai services through supplier's declarations of conformity. IBM Journal of Research and Development 63, 6–1.

[2] Aube, C., Rousseau, V., 2005. Team goal commitment and team effectiveness: the role of task interdependence and supportive behaviors. Group Dynamics: Theory, Research, and Practice 9, 189.

[3] Bolstad, C.A., Endsley, M.R., 1999. Shared mental models and shared displays: An empirical evaluation of team performance, in: proceedings of the human factors and ergonomics society annual meeting, SAGE Publications Sage CA: Los Angeles, CA. pp. 213–217.

[4] Brooks, C., Szafir, D., 2019. Building second-order mental models for human-robot interaction. arXiv preprint arXiv:1909.06508 .

[5] Chakraborti, T., Kambhampati, S., Scheutz, M., Zhang, Y., 2017a. Ai challenges in human-robot cognitive teaming. arXiv preprint arXiv:1707.04775 .

[6] Chakraborti, T., Sreedharan, S., Zhang, Y., Kambhampati, S., 2017b. Plan explanations as model reconciliation: Moving beyond explanation as soliloquy. arXiv preprint arXiv:1701.08317 .

[7] Converse, S., Cannon-Bowers, J., Salas, E., 1993. Shared mental models in expert team decision making. Individual and group decision making: Current issues 221, 221–46.

[8] Cooke, N.J., 2015. Team cognition as interaction. Current directions in psychological science 24, 415–419.

[9] Endsley, M.R., 1988. Design and evaluation for situation awareness enhancement, in: Proceedings of the Human Factors Society annual meeting, Sage Publications Sage CA: Los Angeles, CA. pp. 97–101.

[10] Hoffman, G., 2019. Evaluating fluency in human–robot collaboration. IEEE Transactions on Human-Machine Systems 49, 209–218.

[11] Hoffman, G., Breazeal, C., 2007. Cost-based anticipatory action selection for human–robot fluency. IEEE transactions on robotics 23, 952–961.

[12] Holzinger, A., Plass, M., Holzinger, K., Crisan, G.C., Pintea, C.M., Palade, V., 2017. A glass-box interactive machine learning approach for solving np-hard problems with the human-in-the-loop. arXiv preprint arXiv:1708.01104 .

[13] Jonker, C.M., Van Riemsdijk, M.B., Vermeulen, B., 2010. Shared mental models, in: International Workshop on Coordination, Organizations, Institutions, and Norms in Agent Systems, Springer. pp. 132–151.

[14] Klingspor, V., Demiris, J., Kaiser, M., 1997. Human-robot communication and machine learning. Applied Artificial Intelligence 11, 719–746.

[15] Knepper, R.A., Mavrogiannis, C.I., Proft, J., Liang, C., 2017. Implicit communication in a joint action, in: Proceedings of the 2017 acm/ieee international conference on human-robot interaction, pp. 283–292.

[16] Levesque, H.J., . Task-oriented dialogue as a consequence of joint activity .

[17] Leyzberg, D., Spaulding, S., Scassellati, B., 2014. Personalizing robot tutors to individuals' learning differences, in: 2014 9th ACM/IEEE International Conference on Human-Robot Interaction (HRI), IEEE. pp. 423–430.

[18] Ma, L.M., Fong, T., Micire, M.J., Kim, Y.K., Feigh, K., 2018. Human-robot teaming: Concepts and components for design, in: Field and service robotics, Springer. pp. 649–663.

[19] Marks, M.A., Zaccaro, S.J., Mathieu, J.E., 2000. Performance implications of leader briefings and team-interaction training for team adaptation to novel environments. Journal of applied psychology 85, 971.

[20] Mathieu, J.E., Heffner, T.S., Goodwin, G.F., Salas, E., Cannon-Bowers, J.A., 2000. The influence of shared mental models on team process and performance. Journal of applied psychology 85, 273.

[21] Nikolaidis, S., Nath, S., Procaccia, A.D., Srinivasa, S., 2017a. Game-theoretic modeling of human adaptation in human-robot collaboration, in: Proceedings of the 2017 ACM/IEEE international conference on human-robot interaction, pp. 323–331.

[22] Nikolaidis, S., Zhu, Y.X., Hsu, D., Srinivasa, S., 2017b. Human-robot mutual adaptation in shared autonomy, in: 2017 12th ACM/IEEE International Conference on Human-Robot Interaction (HRI, IEEE. pp. 294–302.

[23] Premack, D., Woodruff, G., 1978. Does the chimpanzee have a theory of mind? Behavioral and brain sciences 1, 515–526.

[24] Setchi, R., Dehkordi, M.B., Khan, J.S., 2020. Explainable robotics in human-robot interactions. Procedia Computer Science 176, 3057–3066.

[25] Tabrez, A., Agrawal, S., Hayes, B., 2019. Explanation-based reward coaching to improve human performance via reinforcement learning, in: 2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI), IEEE. pp. 249–257.

[26] Tabrez, A., Hayes, B., 2019. Improving human-robot interaction through explainable reinforcement learning, in: 2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI), IEEE. pp. 751–753.

[27] Tabrez, A., Luebbers, M.B., Hayes, B., 2020. A survey of mental modeling techniques in human–robot teaming. Current Robotics Reports , 1–9.

[28] Tellex, S., Knepper, R., Li, A., Rus, D., Roy, N., 2014. Asking for help using inverse semantics .

[29] University, C.M., . Pocketsphinx - an open-source continuous speech recognition engine. URL: https://github.com/cmusphinx/pocketsphinx.

[30] Wilson, J.R., Rutherford, A., 1989. Mental models: Theory and application in human factors. Human Factors 31, 617–634.

[31] Zakershahrak, M., Marpally, S.R., Sharma, A., Gong, Z., Zhang, Y., 2020. Order matters: Generating progressive explanations for planning tasks in human-robot teaming. arXiv preprint arXiv:2004.07822 .

[32] Zakershahrak, M., Sonawane, A., Gong, Z., Zhang, Y., 2018. Interactive plan explicability in human-robot teaming, in: 2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), IEEE. pp. 1012–1017.

[33] Zaraki, A., Giuliani, M., Dehkordi, M.B., Mazzei, D., D'ursi, A., De Rossi, D., 2014a. An rgb-d based social behavior interpretation system for a humanoid social robot, in: 2014 Second RSI/ISM International Conference on Robotics and Mechatronics (ICRoM), IEEE. pp. 185–190.

[34] Zaraki, A., Mazzei, D., Giuliani, M., De Rossi, D., 2014b. Designing and evaluating a social gaze-control system for a humanoid robot. IEEE Transactions on Human-Machine Systems 44, 157–168.

[35] Zaraki, A., Pieroni, M., De Rossi, D., Mazzei, D., Garofalo, R., Cominelli, L., Dehkordi, M.B., 2016. Design and evaluation of a unique social perception system for human–robot interaction. IEEE Transactions on Cognitive and Developmental Systems 9, 341–355.

[36] Zhang, Y., Sreedharan, S., Kulkarni, A., Chakraborti, T., Zhuo, H.H., Kambhampati, S., 2017. Plan explicability and predictability for robot task planning, in: 2017 IEEE international conference on robotics and automation (ICRA), IEEE. pp. 1313–1320.

[37] Zhao, F., Henrichs, C., Mutlu, B., 2020. Task interdependence in human-robot teaming, in: 2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN), IEEE. pp. 1143–1149.