Modeling Interaction for Understanding in HRI

Thomas Hellström
Umeå University
Sweden
thomas.hellstrom@umu.se

Suna Bensch
Umeå University
Sweden
suna.bensch@umu.se

ABSTRACT

As robots become more and more capable and autonomous, there is an increased need for humans to understand what the robots do and think. In this paper we investigate what such understanding means and includes, and how robots are and can be designed to support understanding. We present a model of interaction for understanding. The aim is to provide a uniform formal understanding of the large body of existing work, and also to support continued work in the area.

ACM Reference format:
Thomas Hellström and Suna Bensch. 2018. Modeling Interaction for Understanding in HRI. In *Proceedings of Explainable Robotic Systems Workshop at HRI 2018, Chicago, USA, March 2018,* 2 pages.
https://doi.org/10.475/123_4

1 INTRODUCTION

The importance of understandable robots has been acknowledged by the HRI community for a long time, with terms such as *readability*, *anticipation*, *intelligibility*, *intent communication*, *legibility*, *transparency*, or *predictability*. While understandability as such is the goal of a lot of HRI research, an analysis of what the concept really means and how it can be formalized is to the authors' knowledge largely missing. This paper aims at filling this gap, thereby providing a tool for continued research.

In our work we use to the terms "understandability" and "understanding", with the latter defined as "... a psychological process related to an abstract or physical object, such as a person, situation, or message whereby one is able to think about it and use concepts to deal adequately with that object" [3]. More specifically, we focus on what enables humans to successfully interact with robots.

One important aspect of understanding concerns goal-directed actions and intentions of a robot [15]. However, understanding of a robot also includes entities such as desires, knowledge and beliefs, emotions, perceptions, capabilities and limitations of the robot [14], and also task uncertainty [7], and task progress [2, 4]. We refer, somewhat loosely, to all such entities collectively as the *state-of-mind* (SoM) of the robot. We introduce the following definition:

Definition 1 An agent's understanding of another agent is the extent to which the first agent has knowledge about the other agent's SoM in order to successfully interact with it.

Explainable Robotic Systems Workshop at HRI 2018, March 2018, Chicago, USA

ACM ISBN 123-4567-24-567/08/06…$15.00
https://doi.org/10.475/123_4

An agent can support another agent's understanding by performing *communicative actions* [9], that we define as:

Definition 2 A communicative action is an action performed by an agent, with the intention of increasing another agent's knowledge of the first agent's SoM.

It is sometimes sufficient for a robot to generate static communicative actions, such as reported in [1, 2, 5, 10–12]. However, sometimes communicative actions have to be designed to fit the current perspective and needs of the human. For this, the robot benefits from inferring a model of the human's mind by utilizing a first-order *theory of mind* (ToM) [13]. For example, a robot should normally not inform an interacting human about the same thing more than once. To manage this, the robot needs to estimate the human's current knowledge, i.e. it needs a ToM of the human. Other cases require the robot to be equipped with a second-order ToM, such that the robot assumes not only that the human has a mind, but also that she has a ToM of the robot. One (rare) example of how this has been used in earlier research is [8], in which the authors describe how a robot models how a human infers the robot's objectives from observed behavior, and then chooses the most informative behavior to communicate its objectives to the human.

2 INTERACTION FOR UNDERSTANDING

To formally describe existing work in the area, and to provide a tool for research, we propose a model of how robots and humans generate, interpret, and exchange communicative actions aiming at supporting understanding. The model is illustrated in Fig. 1. The robot's SoM M_R contains a model m_H of the human's SoM M_H. In a symmetric fashion, M_H contains a model m_R of M_R. By Definition 1, human understanding of the robot relates to the mismatch between m_R and M_R. We denote this mismatch $|m_R - M_R|$. Communicative actions are generated with the goal of reducing $|m_R - M_R|$. m_R and M_R do not necessarily have to be identical, but the important parts (application dependent) should match.

Human understanding of the robot is established and supported by sequential execution of the three modules I_R, N_R, and G_R:

- I_R The robot infers m_H by using M_R, communicative actions A_H generated by the human, and general interaction Ix between human and robot.
- N_R The robot compares its mind M_R with its estimation of m_R (this estimation is part of m_H). If $|m_R - M_R|$ is too large, the robot identifies which information the human needs in order to reduce $|m_R - M_R|$.
- G_R The robot generates and executes communicative actions A_R aiming at communicating the needed information.

The interacting human's cognitive process is modeled symmetrically in the three modules I_H, N_H, and G_H.

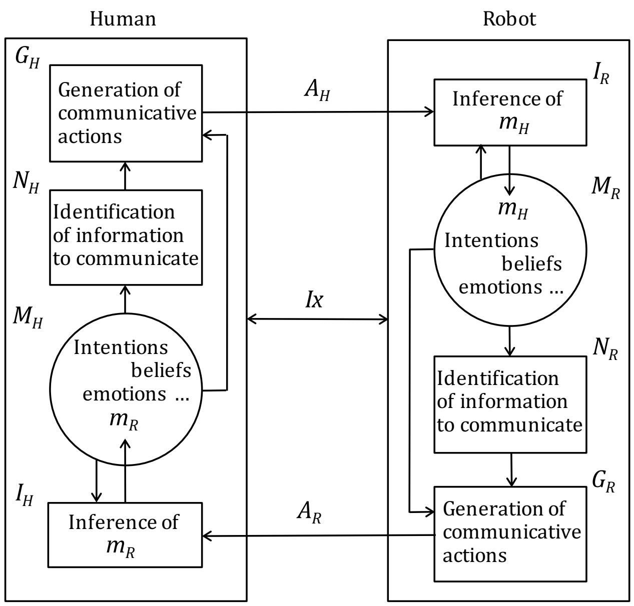

Figure 1: Model of interaction for understanding

Example

Consider an autonomous car driving on the highway. The car's perception system detects a pedestrian approaching the roadway in front of the car. Based on the traffic situation the system decides not to slow down, and performs the following cognitive operations:

- I_R infers that the pedestrian believes that the car intends to slow down, since the pedestrian is entering the road.
- N_R concludes that there is a serious mismatch between the car's intention and the inferred belief of the pedestrian. Communicating the car's intention to the pedestrian is chosen as a means to reduce the mismatch.
- G_R honks and flashes the headlights as means to communicate the intention.

The pedestrian performs the following cognitive operations:

- I_H interprets the honking and flashing headlights as signals indicating that the car does not intend to slow down, but rather expects the pedestrian not to proceed crossing the road. This is also the new decision made (outside of the model) by the pedestrian who consequently stops.
- N_H estimates that there is no serious mismatch between m_H (the car's belief that the pedestrian will not cross) and M_H (the fact that the pedestrian does not intend to cross). Hence, there is no need to communicate any information to the car.
- G_H performs no communicative actions.

3 USING THE MODEL

Development of understandable robots may benefit from thinking in terms of the concepts and modules suggested by the presented model. In particular, development can be guided by answering the following questions related to the modules I_R, N_R, and G_R:

- Q1 (I_R) **How should the robot represent and infer the human's mind?**
- Q2 (N_R) **What information (if any) should be communicated to the human?**
- Q3 (G_R) **How should communicative actions be generated to communicate the required information?**

These are all non-trivial questions. One important sub question of Q2 is how to compute the mismatch $|m_R - M_R|$, and how to determine if it is large enough to generate communicative actions.

The proposed model applies to cases in which both human and robot utilize a first or higher-order ToM to understand each other, and also to simpler cases in which the robot performs static communicative actions in order to support the human's understanding of the robot. It is our hope that the model will serve as inspiration for continued research, in particular by the identification of general concepts and principles for understandable robots. A specific insight provided by the model is the conceptual separation of information to be communicated, from the means to communicate, i.e. the communicative actions. More information on the material presented in this paper can be found in [6] (submitted).

REFERENCES

[1] Svante Augustsson, Jonas Olsson, Linn Gustavsson Christiernin, and Gunnar Bolmsjö. 2014. How to Transfer Information Between Collaborating Human Operators and Industrial Robots in an Assembly. In *Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational (NordiCHI '14)*. ACM, New York, NY, USA, 286–294. https://doi.org/10.1145/2639189.2639243

[2] Kim Baraka and Manuela M. Veloso. 2017. Mobile Service Robot State Revealing Through Expressive Lights: Formalism, Design, and Evaluation. *International Journal of Social Robotics* (16 Oct 2017). https://doi.org/10.1007/s12369-017-0431-x

[3] Carl Bereiter. 2002. *Education and mind in the Knowledge Age*. L. Erlbaum Associates.

[4] F. Broz, A. Di Nuovo, T. Belpaeme, and A. Cangelosi. 2015. Talking about task progress: Towards integrating task planning and dialog for assistive robotic services. *Paladyn - Journal of Behavioral Robotics* 6, 1 (2015), 111–118.

[5] Ravi Chadalavada, Henrik Andreasson, Robert Krug, and Achim J. Lilienthal. 2015. That's on my mind! Robot to Human Intention Communication through on-board Projection on Shared Floor Space. In *Proceedings of European Conference on Mobile Robots*.

[6] Thomas Hellström and Suna Bensch. 2018. Understandable Robots - What, Why, and How. *Submitted to Paladyn - Journal of Behavioral Robotics* (2018).

[7] Julian Hough and David Schlangen. 2017. It's Not What You Do, It's How You Do It: Grounding Uncertainty for a Simple Robot.. In *In Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction (HRI '17)*.

[8] S. H. Huang, D. Held, P. Abbeel, and A. D. Dragan. 2017. Enabling Robots to Communicate their Objectives. *ArXiv e-prints* (2017). arXiv:1702.03465

[9] Ross A. Knepper. 2016. On the Communicative Aspect of Human-Robot Joint Action. In *IEEE International Symposium on Robot and Human Interactive Communication Workshop: Toward a Framework for Joint Action, What about Common Ground?* New York, NY, USA.

[10] Kazuki Kobayashi and Seiji Yamada. 2009. Making a Mobile Robot to Express its Mind by Motion Overlap. In *Advances in Human-Robot Interaction,*, Vladimir A. Kulyukin (Ed.). InTech. https://doi.org/10.5772/6829

[11] J B Lyons. 2013. Being transparent about transparency: A model for human-robot interaction. In *Proceedings of AAAI Spring Symposium on Trust in Autonomous Systems*. 48–53.

[12] Milecia Matthews and Girish V. Chowdhary. 2015. Intent Communication between Autonomous Vehicles and Pedestrians. In *Robotics: Science and Systems*.

[13] M. Michlmayr. 2002. *Simulation Theory versus Theory Theory: Theories concerning the Ability to Read Minds*. Master's thesis. University of Innsbruck.

[14] Hugo Romat, Mary-Anne Williams, Xun Wang, Benjamin Johnston, and Henry Bard. 2016. Natural Human-Robot Interaction Using Social Cues. In *The Eleventh ACM/IEEE International Conference on Human Robot Interaction (HRI '16)*. IEEE Press, Piscataway, NJ, USA, 503–504.

[15] D. Vernon, S. Thill, and T. Ziemke. 2016. The Role of Intention in Cognitive Robotics. In *Toward Robotic Socially Believable Behaving Systems - Volume I*, Anna Esposito and Lakhmi C. Jain (Eds.). Springer International Publishing, 15–27.