This is the published version of a paper presented at HRI 2018, Chicago, USA, March 2018.

Citation for the original published paper:


N.B. When citing this work, cite the original published paper.

Permanent link to this version:
http://urn.kb.se/resolve?urn=urn:nbn:se:umu:diva-149102
Modeling Interaction for Understanding in HRI

Thomas Hellström
Umeå University
thomas.hellstrom@umu.se

Suna Bensch
Umeå University
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:
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 understandability, 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.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

© 2018 Copyright held by the owner/author(s).
ACM ISBN 123-4567-24-567/08/06... $15.00
https://doi.org/10.475/123_4

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_H and M_R. We denote this mismatch |M_R − M_H|. Communicative actions are generated with the goal of reducing |M_R − M_H|. However, |M_R − M_H| 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 I_x between human and robot.

N_R The robot compares its mind M_R with its estimation of M_H (this estimation is part of M_H). If |M_R − M_H| is too large, the robot identifies which information the human needs in order to reduce |M_R − M_H|.

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.
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_H|, 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


