Partner Orientation in Asymmetric Communication: Evidence from contingent robot response

Kerstin Fischer  
Institute for Design and Communication  
University of Southern Denmark  
6400 Sonderborg, Denmark  
kerstin@sdu.dk

Katrin Lohan  
Katharina Rohlfing  
Bielefeld University  
Technical Faculty  
K.Lohan@hw.ac.uk  
kjr@uni-bielefeld.de

Kilian Foth  
University of Hamburg  
Department of Informatics  
22527 Hamburg, Germany  
foth@informatik.uni-hamburg.de

ABSTRACT
In this paper, we present evidence from empirical studies of human-robot interaction with different kinds of robot feedback, which suggests that speakers take their communication partners into account on the basis of partner models, which are constantly and concomitantly updated on the basis of the partner’s feedback. Thus, the partner’s contributions are not reacted to directly, but instead feed into a partner model, on the basis of which utterances are produced.

Categories and Subject Descriptors
H.1.2 User/Machine Systems

General Terms
Human Factors

Keywords
Asymmetric communication, human-robot interaction, partner orientation, partner model, alignment, feedback

1. INTRODUCTION
The problem addressed in this study is the possible mechanism by means of which speakers adjust to their communication partners. Currently, there exist many explicit and implicit proposals on how humans take their partners into account. Some hold that there are conventional ways of interacting with our partners; for instance, we know how to talk to children because we have encountered such speech in early childhood ourselves (e.g. Ferguson 1982). A recent version of this concept of (conventional) association is Pickering and Garrod’s (2013) suggestion that we predict our partners’ behaviors, especially if they are dissimilar to us in some way or other, by association with certain behaviors. That is, if previous experience tells us that this is how you talk to children, then we use these ways to talk to children. Another proposal is that we adjust to our partners by interactive alignment (Pickering & Garrod 2004), i.e. by automatic priming. That is, the partner uses particular behaviors in certain ways, so we do so, too. A proposal to account for partner orientation assumes cognitive modeling of the partner and strategic choice of behavior on the basis of the partner model (e.g. Clark 1996). Proponents of this view assume that we build up rich models of our partners and the context and take these factors into consideration during speech planning (see also Brown-Schmidt (2009) and Brennan et al. (2010), for example).

Again others assume that such modeling occurs only later in processing and only if there is time, suggesting two-stage models in which speakers initially do not include knowledge about their partners (Shintel & Keysar 2009). And finally, some authors suggest that people choose their behaviors based on their psychological needs, for instance, the need to distance oneself from, or to express solidarity with, the other person (e.g. Giles et al. 2006).

To sum up, numerous approaches to account for how people adjust to their communication partners exist, all suggesting different mechanisms. While these different proposals have been developed on human-human interaction, asymmetric communication in general and human-robot interaction in particular provide us with useful tools to investigate how people take their partners into account since a) robots are very different from humans and thus the adjustments made for artificial communication partners, such as robots, will be different than those for other human adults (cf. Metzing and Brennan 2003), which renders the processes and mechanisms involved identifiable; and b) robots can be controlled in ways human communication partners cannot, which allows the experimentation with properties and behaviors of communication partners in controlled ways, which would be impossible to this extent in interactions between humans. Human-robot interaction is thus a useful methodological resource for the investigation of human communication processes (Fischer 2010).

2. EMPIRICAL STUDY
In order to determine the model of partner orientation that accounts for asymmetric communication best, we carry out an empirical study that allows us to distinguish between the predictions of the different models proposed. In particular, we set up HRI experiments in which the robot responds in controllable ways to a defined set of the tutor’s behaviors and vary these responses, which results in two conditions that differ in the robot’s behavior. Using this design, we can identify the adjustments people make when taking their artificial communication partner into account and thus determine the mechanisms involved in partner orientation. We reason that if automatic processes are responsible for the adjustment to the
partner, the adjustments in users’ behavior should be local, straight-forward stimulus-response patterns and not configurations of adjustments. Furthermore, the timing of the robot’s behaviors in the two conditions should not make a difference. In contrast, if a partner model mediates people’s choices, the adjustments should be global and concern configurations of features. Furthermore, the timing of the robot’s behaviors should have (global) effects, which would then lead to different behaviors in the two conditions.

2.1 Participants
Participants were 38 native speakers of German; Condition 1, in which the robot responds contingently to the tutors’ instructions (see below), consists of 19 tutors, 19-68 years old (median 24), seven male and twelve female. Condition 2, in which the robot exhibits the same behaviors as in Condition 1, yet employed randomly, consists of 19 participants, 20-55 years old (median 25.5), nine of whom are male and ten female.

2.2 Tasks
Participants’ tasks were to explain certain objects to the robot, in particular, how to switch on a lamp, how to stack cups, how to arrange blocks to form a toy house, how to ring a bell, how to put rings into a box and how to dispense salt.

2.3 Robot
The robot used is the iCub (see Figure 1). The robot was equipped with the tutor spotter (Lohan 2011), a system that monitors the tutor and allows it to respond contingently to the tutor’s eye gaze and demonstrating actions. The contingency detection is calculated based on temporal co-occurrence of visually detected ostensive signals of human and robot behavior. The classification of the eye gaze was obtained by geometrical calculations, resulting from locating the intersection point between gazing orientation and the object plane or face plane of the robot. In other words, the eye gaze module detects whether the tutor is looking towards the object (Fig 1b), to the face of the learner (Fig 1a) or somewhere else (Fig 1c).

The robot then responds to the tutor’s gazing behavior by taking its own current state and the tutor’s gaze direction into account (Lohan 2011). Furthermore, if the tutor moves objects closer to the robot in order to demonstrate the objects to the robot, the robot will respond by pointing at the objects.

2.4 Data Encoding
The data were analyzed for properties describing the kinds of instructions tutors produce for their artificial communication partner. In particular, we looked for the following indicators of linguistic complexity in the instructions (see Fischer et al. 2013ab): Speakers’ mean length of utterance has been found to be adjusted reliably in tutoring in child-directed speech (Roy et al. 2009). Another useful indicator of complexity is embedding, i.e. the number of structures that are embedded into others, such as appositions, subclauses, relative clauses and the like.

Furthermore, since in child-directed speech references to the past are rare (Snow 1972), we have also analyzed whether there are differences in the amounts of references to the past.

Finally, we encoded whether and how tutors involve the robot by means of tag questions, such as ne?, okay? or verstanden? (understood?).

2.5 Results
The ANOVA results for all features investigated shows significant differences between the two conditions (F(1,120) = 1.9664, p<.05). Post hoc analyses show that these differences are due to differences in mean length of utterance, embedding, relative clauses and the number of tag questions produced, i.e. the amount of understanding checks with which the robot is addressed.

Thus, while there is higher complexity of utterances in Condition 2 (see Figure 3), there are more tag questions in Condition 1, indicating that tutors check the robot’s understanding more often when the robot responds contingently to their gaze and demonstrating behavior (see Figure 4). However, regarding references to the past, no differences could be found since talking about past events was rare in both conditions.

![Figure 1: Robot Response Patterns: When the tutor is gazing towards iCub, iCub gazes back (1a); when the tutor is gazing at the object, iCub looks at the object (1b); when the tutor looks elsewhere, iCub follows the eye gaze (1c).](image)

The robot’s behavior is identical in both conditions, yet in Condition 1, the behaviors are synchronized with the tutor’s actions such that they occur within the usual human response rate of 3 seconds, whereas in Condition 2, the behaviors are coordinated with the object. That is, if there is an object, the robot looks at the object, if there is none, the robot will look at the tutor’s face. The pointing gestures were produced randomly.

![Figure 2: Example session: The participant explains iCub how to dispense salt](image)
3. DISCUSSION

The results show that the contingent robot response has an impact on tutors’ linguistic behavior, especially concerning the complexity of their utterances. That is, while the two conditions differ in the timing of the robot’s gaze and pointing behaviors, it is not the tutor’s gaze and demonstrating or pointing behavior that changes, but higher level interpretations of these behaviors. In particular, not those behaviors are adjusted that correspond to the robot’s behavior, but that instead people’s overall behavior towards the robot changes. The timing of the partner’s responses is thus taken as indicator of the partner’s understanding of the current state of talk (Fischer et al. 2013ab). People’s responses to this behavior are consequently not immediate, direct and automatic reactions (by alignment or association) but instead mediated by a partner model. Therefore, the studies of different types of robot feedback show that correctly timed non-verbal robot behavior contributes to a certain mental model of the robot as an interaction partner, which then licenses certain behaviors.

4. CONCLUSION

To sum up, our findings on tutors’ adjustments to the socially contingent robot show that action is chosen on the basis of a partner model, which in turn is sensitive to the partner’s behavior, which is taken into account concomitantly and subtly. Importantly, the effects from the partner’s feedback concern not isolated features of the tutors’ behavior but instead configurations of such features, such that socially contingent robot response has an impact on the tutors’ partner models with respect to interactivity, complexity and capability of the robot in general. The model emerging is thus one in which linguistic choices are made on the basis of a partner model, yet which is updated online during the interaction.

ACKNOWLEDGEMENTS

This research was partially funded by the EU project ITALK.

5. REFERENCES


