Robots as Confederates: How robots can and should support research in the humanities

Kerstin FISCHER[a,1]
[a] *University of Southern Denmark*

Abstract. This paper addresses the use of robots in experimental research for the study of human language, human interaction, and human nature. It is argued that robots make excellent confederates that can be completely controlled, yet which engage human participants in interactions that allow us to study numerous linguistic and psychological variables in isolation in an ecologically valid way. Robots thus combine the advantages of observational studies and of controlled experimentation.

Keywords. humanistic research, linguistics, psychology, confederates, matched guise technique, human-robot interaction

1. Introduction

When the object of study is human language, human interaction, or human nature, then the obvious thing to do is to study humans, humans speaking and humans interacting. In ethnomethodological Conversation Analysis (Sacks et al. 1974), for instance, this perspective is part of the core principles of the methodology, and thus 'ordinary conversation' is taken to be the most relevant, most natural starting point for the analysis of social interaction (Hutchby & Wooffitt 1998).

However, if we want to answer questions that concern possible causes and influencing factors, it may be necessary to go beyond the observation of behaviors in naturally occurring situations and to elicit data in controlled scenarios. For instance, in order to find out about how people take their communication partners into account, Schober & Brennan (2003) suggest that we have two options with advantages and disadvantages each: corpus studies and laboratory studies. By corpus studies they mean investigations of collections of spontaneous conversations occurring in real-life settings. The advantage of such corpus studies is their ecological validity, even though there may be some methodological problems, such as the question of sample size and the difficulty to infer the speakers' intentions from the transcript. The alternative, Schober & Brennan

[1] Paper submitted to the Robophilosophy 2016 Conference, Aarhus, Denmark.

suggest, are laboratory studies, which allow the researcher "to assess speakers' intentions and addressees' comprehension independently of the conversation" (Schober & Brennan 2003: 129). The disadvantages of such studies are their lack of ecological validity, such that the controlled psycholinguistic experiments in which many of the findings are obtained do not necessarily allow us to conclude what participants do 'in the wild', and their restrictedness to task-oriented situations.

Nevertheless, controlling certain aspects of the situation is what may be necessary to identify the role of possibly influential factors. For instance, lab studies can control the degree with which participants know each other, how much task-related knowledge they have, how much experience they have in the interaction with each other etc. (see, for instance, Clark & Wilkes-Gibbs 1986; Brown-Schmidt et al. 2008; Brown-Schmidt 2009). Thus, eliciting data in controlled lab settings may be the only way to tease out the effects of particular variables that influence people's behavior. In such studies, often confederates are used, research associates who behave in controlled ways in order to elicit a particular response (e.g. Brown & Dell 1987). Now, my point in this paper is that the role of the confederate is best taken over by a robot. I am going to illustrate this point on three case studies, concerning human language, human interaction and human nature.

2. Robots as Confederates in Studies of Human Language

In order to identify the functions of certain linguistic features, we can, for instance, analyze the contexts in which they occur and deduce from the way they are responded to what their effects may have been (i.e. conducting a corpus study). If the kinds of responses elicited are stable across contexts, we can assume that there is a causal relationship between the linguistic feature under consideration and a certain communicative function. However, once we enter the realm of interpersonal functions, i.e. functions that concern how much a person producing the feature is liked, accepted, respected etc., identifying the effects of linguistic features becomes increasingly difficult since very many different variables co-occur in a real life situation: the person, his or her appearance, his or her gender, choice of words, contents stated, tone of voice, pronunciation, non-verbal behavior and many more, which all come together and can potentially influence the way an utterance is perceived. Furthermore, the communicative effects of such signals cannot always be traced directly in the interaction; instead, they may influence interactions on a global level, if at all, and may thus be hard to trace. Linguistic research has thus developed techniques to isolate the effects of certain variables, notably the matched guise technique. In this method, a bilingual speaker produces the same or similar sentences in two ways, varying only one linguistic feature. For instance, one may want to find out who is rated as more intelligent, educated or friendly, someone with an RP accent or someone with a Standard American accent. So a

speaker who can speak in both a British RP accent and in a Standard American accent equally well will record the same sentence or two similar sentences in each accent. Then other speakers are recorded, saying the same sentence. The sentences are then played to participants who have to rate the respective speaker for those properties that are the suspected functions of the linguistic feature under consideration, here, for instance, social status, suspected intelligence, education, income, friendliness etc. The stimuli are presented as if they all belonged to different speakers, disguising the fact that two of the stimuli are produced by the same speaker. Participants then rate all 'speakers', and when they differ in the way they perceive the bilingual speaker who produced both stimuli, the differences can be inferred to be due to the different accents.

Now, the quality of such an investigation crucially depends on the availability of a bilingual speaker and his or her ability to produce authentic utterances that differ only in the respect under investigation. Furthermore, because the stimuli have to be presented audio only, the method is restricted to non-interactive, decontextualized, monological speech phenomena. Obviously, one can circumvent the first problem by manipulating his or her utterance on the computer to produce the contrast under consideration, yet the problems of the lack of interactivity and possible interrelationships with other features remain.

Here, robots are the much better confederates. We can presynthesize or record each utterance in advance, manipulate it the way we want and have participants interact with the robot, being able to not only ask in questionnaires *post hoc* how they perceive the robot, but can also observe behavioral consequences directly. For instance, in a recent study (Jensen et al. submitted), we had two identical Keepon robots greet each participant, where one used longer and the other shorter syllables. Previous work on free conversation (Pillet-Shore 2012) had hypothesized that speakers use longer greetings for people they know and shorter greetings for strangers. Our studies support this hypothesis: people rated the robot with the longer syllables as significantly more friendly, and they wanted to possess it more.

To sum up, robots make reliable, controllable, interactional confederates for the analysis of interpersonal effects of linguistic features.

3. Robots as Confederates in Studies of Human Interaction

In psychology, confederates are frequently used to study the effects of particular partner behaviors. For instance, in a study of partner effects in storytelling, Brown & Dell (1987) had participants tell a story to a confederate who either had a picture showing the relevant information or not. The authors find only a weak effect for taking the partner's needs into account. However, when Lockridge and Brennan (2002) repeated the experiment with a non-confederate, real participant, they found that speakers take

listeners with real information needs into the account all the time. Brennan, Galati & Kuhlen (2010) thus suggest that confederates cannot help using micro-cues that subtly inform speakers about their information states and thus influence the results of these investigations. So what confederates communicate implicitly is the lack of need of information. Here, robots are ideal because they will behave exactly in the way programmed and identically for each participant, irrespective of how often they have heard a story being told before.

Furthermore, there are limits to how much can be controlled in the first place; most linguistic behaviors are routinized and produced subconsciously, which makes it difficult to control them voluntarily. In comparison, using robots we can control for subtle interactional features, such as the timing of a listener's behaviors. Such manipulations are far beyond anything a human confederate could control. For robots, in contrast, such manipulations can be programmed such that they will be used consistently across participants. For instance, several studies have investigated the effects of contingent response, i.e. the temporal connectedness of the robot's and the human tutor's behavior (Lohan et al. 2011; Fischer et al. 2013, 2014; Fischer in press). In these experiments, the robot's gaze and pointing behavior was either coordinated with the human tutor's behavior or played at random, creating the impression of a lively robot. The experiments reveal that participants interacting with the contingently responding robot attribute more understanding to the robot by reducing the complexity of their utterances more, by involving the robot more and by resting their explanations more on situationally available information (Fischer 2016).

For investigations of human interactions, robot confederates can thus contribute to our understanding of complex interactional processes and psycholinguistic processing. Robots consequently make excellent confederates also in experiments on processes that constitute interaction.

4. Robots as Confederates for the Study of Human Nature

Much previous research has shown that people respond to robots as if they were social beings. Most influentially, Nass and colleagues have shown in various publications that people respond to computers and robots as if they were people. For instance, they may respond to flattery from a computer in the same way as to flattery from another person (Reeves & Nass 1996), they may respond to computer voices in similar ways as to human voices (Nass & Brave 2005), and they may produce similar politeness behaviors for humans and computers (Nass 2004). Nass and Moon (2000) argue that the reason is mindless transfer, i.e. an automatic, fast misjudgment of the human-robot interaction as a social situation. Nass (2004) suggests that this error rests on our evolution in social environments. The phenomenon extends to the interaction with robots (Groom et al. 2009). At the same time, people do not simply exhibit the same behavior when

interacting with robots as when interacting with other people (e.g. Fischer et al. 2011). Thus, if people behave as if robots were other people, they only do so sometimes (Fischer 2011), or to different degrees (Shechtman & Horowitz 2003), or under special circumstances. However, irrespective of where exactly this behavior is coming from, robots help us identify the human tendency to attribute sociality into all kinds of beings, revealing the considerable human bias for social relationship. Human-robot interaction experiments like those cited above shed light on human nature and those processes that define human relationships.

Similarly, implementing robots to copy human behavior entails that we get to increasingly more detailed understandings of human behavior and its underlying causes in the first place. The large amount of recent work on implementing human-like capabilities into robots (e.g. on gaze aversion (for example Andrist et al. 2014) or on trust (Grigore et al. 2011) has in fact enriched our understanding of these human behaviors considerably.

5. Possible Problems with Robots as Confederates

There are generally three objections against robots as confederates; a) that the interactions investigated are reduced to a degree that makes them unnatural and consequently that the processes observable in human-technology interaction are very different from processes in natural conversation and thus irrelevant; b) that people will attribute intentions and properties to robots that they do not possess; and c) that understanding is interactionally achieved and that therefore interactions are not comparable.

One part of the answer to the first objection, that the approach is reductionistic, is that yes, robots are still very much restricted in the amount of social signals they process and produce, which is however not necessarily a disadvantage; if interactions with robots were as seamless and complex as interactions between people, they would not serve their purpose, to allow the study of the effects of independent variables that constitute interactions. Thus especially if interactions do not go smoothly, they are particularly revealing since they constitute so-called 'deviant cases' (e.g. Hutchby & Wooffitt 1998: 98), i.e. situations in which it becomes apparent what participants would have expected and how they make sense of what they are confronted with (see Fischer forthcoming).

Another part of the answer is that because conversation is such a sense-making activity, which has, for instance, been described by means of the cooperative principle (Grice 1975), people will make sense out of the robot's behavior, irrespective of how minimal it may be, as numerous studies of HRI have documented (e.g. Youssef et al. 2015; Salem et al. 2015). And finally, in some circumstances, the reduction of social signals produced by a robot in comparison with a person is wanted, for instance, in the

interaction with autistic children (Scasselati et al. 2012), or in interactions in which (implicit) evaluation is to be prevented, for example, in conflict mediation (Jung et al. 2015). Here robots have been shown to be possibly even superior to humans.

The second objection is that people will take all kinds of preconceptions into the interactions with robots to fill in the gaps and thus that they attribute properties to robots that they do not have. This is certainly the case, but equally so in interactions with other, human, communication partners, such as infants (see Kaye 1980), where parents have been found to make up for the lack of interactional competence of their young communication partners. Similarly, Edwards (2001) argues that we always attribute properties and intentions to other beings, including other people. In fact there is no independent evidence that people have intentions that motivate their behaviors; instead, it is as likely that people ascribe intentions and desires to others based on their behaviors, and even to themselves (see Cialdini's 2010 discussion of commitment). In this way, there is no principal difference between the interactions with other people, dogs, computers and robots (Edwards 1994). Moreover, in human-robot interaction, other than in conversation between humans, we can investigate people's sense making and can thus take it into account, whereas in human conversation we tend to treat it as a given. For instance, in Fischer (2011) I have shown that people indeed differ in their ways of understanding human-robot interaction; while some treat it as if it was an interaction with another person, along the lines suggested by Nass and colleagues, some treat the robot as a mechanical tool and withhold all social signals. Similarly, while conversation analytical studies of human conversation, for instance, do not normally quantify their findings, if they do, they also report instances in which a person's personal agenda influences social practices (e.g. Schegloff 1968). Furthermore, considerable interpersonal differences have been found in the interaction with a foreigner (Smith et al. 1991). So while the objection is correct that people attribute intentions and properties to robots that they do not have, people are likely to do that in interactions with other people, too, just that it is less obvious.

The third objection, that meanings are interactively achieved, is accurate, too, and especially scripted human-robot interactions are vulnerable against this criticism because participants take the robot's behavior as occurring in response to their particular utterance. So a robot utterance like 'I did not understand' may lead to very different conclusions if it occurs after an utterance like 'I would like to you to go to the second object in the middle of the four red ones' than when it occurs after an instruction like 'go straight'. In such cases, different people may experience the robot differently overall, dependent on what they started out with; on the other hand, since the robot's utterances are taken to occur in response to speakers' prior utterances, at least there is a strong connection between what speakers start out with and how they interpret the robot's responses. My findings, that speakers' behavior at the beginning of interactions predicts their behavior much later in the interaction to a high degree (Fischer 2011), suggests that the randomness of the robot's responses do not influence the interaction as a whole

very much. Nevertheless, such instances limit the comparability of the interactions with the same robot with scripted behavior – however, in the same way as interactions with human confederates do (see Smith et al. 1991); if confederates respond to participants' utterances, they also involuntarily contribute to the interactional negotiation of meaning in one way or other – which limits the comparability of the interactions with different participants. The solution here can only be to refrain from using confederates completely, human and robot, which then limits the range of possible methodologies.

6. Conclusion

To sum up, robots as confederates are extremely helpful tools for research in the humanities on various different levels. Being completely controllable in ways humans are not makes them perfect confederates for experimental research about human language, interaction and nature. At the same time, robots allow us to carry out humanistic research in a political climate that attaches little value to the study of humanity and a high value to the study of technology.

References

[1] Andrist, S., Tan, X.Z., Gleicher, M. & Mutlu, B. (2014). Conversational Gaze Aversion for Humanlike Robots. *Proceedings of HRI'14,* Bielefeld, Germany.

[2] Brennan, S. E., A. Galati, and A. K. Kuhlen (2010). Two minds, one dialog: Coordinating speaking and understanding. *Psychology of Learning and Motivation* 53, 301–344.

[3] Brown-Schmidt, S. (2009). Partner-specific interpretation of maintained referential precedents during interactive dialog. *Journal of Memory and Language* 61, 171-190.

[4] Brown-Schmidt, S., C. Gunlogson, and M. K. Tanenhaus (2008). Addressees distinguish shared from private information when interpreting questions during interactive conversation. *Cognition* 107, 1122–1134.

[5] Brown, P. and G. Dell (1987). Adapting production to comprehension: The explicit mention of instruments. *Cognitive Psychology* 19, 441–472.

[6] Cialdini, R. (2010). *Influence*. Collins.

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

[8] Edwards, D. (1994). Imitation an artifice in apes, humans, and machines. *American Behavioral Scientist* 37 (6), 754–771.

[9] Edwards, D. (2001). *Discourse and Cognition*. London, Thousand Oaks, New Delhi: Sage.

[10] Fischer, K. (2011). Interpersonal variation in understanding robots as social

actors. In *Proceedings of HRI'11*, March 6-9th, 2011. Lausanne, Switzerland, pp. 53–60.

[11] Fischer, K. (in press). The Situatedness of Pragmatic Acts: Explaining a Lamp to a Robot. Allan, K. /Capone, A. /Kecskes, I. (Eds): *Pragmemes and Theories of Language Use*. Berlin, Springer.

[12] Fischer, K. (2016): *Designing Speech for a Recipient: The Roles of Partner Modeling, Alignment and Feedback in So-Called 'Simplified Registers'*. Amsterdam: John Benjamins.

[13] Fischer, K., K. Foth, K. Rohlfing, and B. Wrede (2011). Mindful tutors – linguistic choice and action demonstration in speech to infants and to a simulated robot. *Interaction Studies* 12 (1), 134–161.

[14] Fischer, K., K. S. Lohan, J. Saunders, C. Nehaniv, B. Wrede, and K. Rohlfing (2013): The impact of the contingency of robot feedback on HRI. *Cooperative Technological Systems,* San Diego.

[15] Fischer, K., Lohan, K.S., Rohlfing, K. & Foth, K. (2014) Partner Orientation in Asymmetric Communication: Evidence from contingent robot response. HRI '14 *Workshop on Humans and Robots in Asymmetric Interactions*, March 3rd, 2014, Bielefeld, Germany.

[16] Grice, P. (1975). Logic and Conversation. In Cole, P. and Morgan, J. (eds.): *Syntax and Semantics: 3. Speech Acts*. New York: Academic Press, pp. 41-58.

[17] Grigore, E.C., Eder, K., Lenz, A., Sergey Skachek, S, Pipe, A. and Melhuish, C. (2011). Towards Safe Human-Robot Interaction. Towards Autonomous Robotic Systems - 12th Annual Conference, TAROS 2011, Sheffield, UK, August 31 - September 2, 2011.

[18] Groom, V., L. Takayama, P. Ochi, and C. Nass (2009). I am my robot: The impact of robot-building and robot form on operators. In *Proceedings of the Human-Robot Interaction Conference: HRI 2009*, San Diego, CA. , pp. 31–36.

[19] Hutchby, I. and R. Wooffitt (1998). *Conversation Analysis*. Cambridge: Polity.

[20] Jensen, L.C., Fischer, K. and Niebuhr, O. (submitted). Keepon Greeting: Phonetic Variance in First Encounter Meetings.

[21] Jung, M.F., Martelaro, N. & Hinds, P.J. (2015). Using robots to moderate team conflict: the case of repairing violations. *Proceedings of the Tenth Annual ACM/IEEE Conference on Human-Robot Interaction*, Portland, Oregon.

[22] Kaye, K. (1980). Why we don't talk baby talk to babies. *Journal of Child Language* 7, 489–507.

[23] Lockridge, C. B. and S. E. Brennan (2002). Addressees' needs influence speakers' early syntactic choices. *Psychonomic Bulletin & Review* 9 (3), 550–557.

[24] Lohan, K.S., K.J. Rohlfing, K. Pitsch, J. Saunders, H. Lehmann, C.L. Nehaniv, K. Fischer and B. Wrede (2012): Tutor spotter: Proposing a feature set and evaluating it in a robotic system. *International Journal of Social Robotics* 4(2):131-146.

[25] Nass, C. and S. Brave (2005). Wired for Speech. How Voice Activates and Advances the Human-Computer Relationship . Cambridge, MA., London: MIT Press.
[26] Nass, C. and Y. Moon (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues* 56 (1), 81–103.
[27] Nass, C. I. (2004). Etiquette equality: Exhibitions and expectations of computer politeness. *Communications of the ACM* 47 (4), 35–37.
[28] Pillet-Shore, D. (2012). Greeting: Displaying stance through prosodic recipient design. Research on Language & Social Interaction, 45, 4, 375–398.
[29] Reeves, B. and C. Nass (1996). *The Media Equation*. Stanford: CSLI and Cambridge: Cambridge University Press.
[30] Sacks, H., E. A. Schegloff, and G. Jefferson (1974). A simplest systematics for the organization of turn-taking for conversation. *Language* 50 (4), 696–735.
[31] Salem, M., Lakatos, G., Amirabdollahian, F. and Dautenhahn, K. (2015). Would You Trust a (Faulty) Robot? Effects of Error, Task Type and Personality on Human-Robot Cooperation and Trust. In: *Proceedings of the 10th ACM/IEEE International Conference on Human-Robot Interaction (HRI 2015),* Portland, Oregon, USA.
[32] Schober, M. F. and S. E. Brennan (2003). Processes of interactive spoken discourse: The role of the partner. In A. C. Graesser, M. A. Gernsbacher, and S. R. Goldman (Eds.), *Handbook of Discourse Processes*, pp. 123–164. Hillsdale: Lawrence Erlbaum.
[33] Scasselati, B., Admoni, H. and Mataric, M. (2012). Robots for use in autism research. *Annual Review of Biomedical Engineering* 14: 275-294.
[34] Schegloff, E.A. (1968). Sequencing in Conversational Openings. *American Anthropologist*, 70, 6:1075-1095.
[35] Shechtman, N. and L.M.Horowitz (2003). Media inequality in conversation: How people behave differentlywhen interacting with computers and people. In *CHI 2003*, April 5-10, 2003, Ft. Lauderdale, Florida, Volume 5, pp. 281–288.
[36] Smith, S. W., N. Scholnick, A. Crutcher, and M. Simeone (1991). Foreigner talk revisited: Limits on accomodation to nonfluent speakers. In J. Blommaert and J. Verschueren (Eds.), *The Pragmatics of International and Intercultural Communication*. Amsterdam: Benjamins.
[37] Youssef, K., PR De Silva, M Okada (2015). Exploring the Four Social Bonds Evolvement for an Accompanying Minimally Designed Robot. *Social Robotics*, 2015