



Københavns Universitet



Robots Cannot Lie

Borggreen, Gunhild Ravn

Published in:
Sociable Robots and the Future of Social Relations

Publication date:
2014

Citation for published version (APA):
Borggreen, G. (2014). Robots Cannot Lie: Performative Parasites of Robot-Human Theatre. In J. Seibt, R. Hakli, & M. Nørskov (Eds.), *Sociable Robots and the Future of Social Relations : Proceedings from Robo-Philosophy 2014* (pp. 157-163). Amsterdam: IOS Press.

“Robots Cannot Lie”: Performative Parasites of Robot-Human Theatre

Gunhild BORGGREEN¹

Department of Arts and Cultural Studies, University of Copenhagen

Abstract. The presentation reflects on theories of performative speech acts from the British philosopher of language J.L. Austin’s *How To Do Things With Words* (based on a series of lectures in the 1950s). Using Austin’s notion of “parasitic” language of an actor on the stage as a starting point, the paper analyses the Japanese Robot-Human Theatre production *Hataraku Watashi* (I, Worker) by Hirata Oriza and Ishiguro Hiroshi from 2008 as a site for investigating the complex layers of intention, effect and cultural conventions in human-robot interaction. Linking the fiction of the stage production to laboratory testing that include robots in simulated real-life situations, the notion of how robots “lie” are discussed in terms of different types of participants: theatre audiences and test persons. The presentation will focus on how the robot acts “parasitic” upon normal circumstances and thus points out the troubled notions of “normal” or “natural” in robot-human interaction, as well as in human-human interaction.

Keywords. human-robot theatre, lie, Wizard of Oz, performativity, parasitic, speech act, art

Introduction

As the project manager of ROCA, robot culture and aesthetics, I represent a network of scholars who are interested in the cultural and artistic dimensions of robots. We are based at the Department of Arts and Cultural Studies at University of Copenhagen. ROCA is still at the stage of applying for funding and establishing a creative and dynamic research environment. Our aim is to be able to contribute to the field of human-robot interaction from the perspective of cultural and aesthetic theories. By applying a practice-based research dimension, we also hope to build bridges between artists, designers, and engineers in order to incorporate cultural and aesthetic dimensions into the technology and the interaction design from the very start. In this paper, I will analyse a robotic art product which is created as a collaboration between an artist and a robotic scientist as an example of the kind of inquiries cultural and aesthetic theories may pose to the field of robotics. I will point out how aesthetic elements especially related to theatre and performance in this art work may contribute to a broader understanding of human-robot interaction, and as well as human-human interaction.

The art work I will discuss is the Japanese robot-human theatre production entitled *Hataraku Watashi* (I, Worker) from 2008, written and directed by theatre director

¹ Department of Arts and Cultural Studies, University of Copenhagen, Karen Blixens Vej 1, DK-2300 Copenhagen S, Denmark; E-mail: gunhild@hum.ku.dk.

Hirata Oriza in collaboration with robot scientist Ishiguro Hiroshi. Professor Ishiguro, who is one of the distinguished key note speakers at this conference, is acclaimed for his android robots, and he has also collaborated with Hirata on robot-human theatre play that feature the android robot, but today I will discuss an art work including more conventional humanoid robots. I will first show a short extract of a video recording of the play, and then bring in screen shots for further elaboration.

In this short footage, a young man called Yuuji sits at a low table. In front of him, on the other side of the table, are two bright yellow robots. Yuuji is acting out an apology to one of the robots. He bows his head and says "I'm sorry about before. I thought it would cheer you up". The robot replies by dismissing the incident. "It is nothing. It is not your fault," the robots says. The other robot standing next to them adds, as a way of confirming the truth in this remark: "robots cannot lie"[1].

The play *I, Worker* includes verbal communication and interaction between robots and human beings, and the narrative displays an example of how relationship with humanoid robot may look like in the not to far future. In this paper I would like to explore the performative aesthetics of this kind of robot-human theatre by comparing the art work with another type of theatrical setting that includes robots. I am particularly intrigued by the line spoken by the robot in the play here, when it claims that robots cannot lie.

By default robots do not possess human properties such as intention or will, so it may seem obvious that robots cannot lie. On the other hand, in this play, the robots are programmed to appear as if they have human emotions. In other words, robots pretend to be humans, and thereby act parasitic upon the norms and appearances of human behaviour. By using the philosopher J.L. Austin's notion of infelicities in speech act theory as well as Jacques Derrida's comments on Austin, I will argue that *I, Worker* highlights the notions of citation or iteration through the aesthetics of theatre. I suggest that the aesthetics of theatre as well as other art forms can offer valuable insights for robotic science that are different from those obtained in the field of HRI (human-robot interaction), which is primarily based on sociology, psychology and cognitive science. It reveals the ambiguous and fragile foundation upon which humans perceive of themselves.

1. Robot-Human Theatre *I, Worker*

The play *I, Worker* focuses on dialogues and interactions between human beings and robots. The play features a young couple Yuuji and Ikue, and the narrative takes place in the living room of their home. Yuuji and Ikue have no children, but they have two humanoid robotic home companions to help out in the house: Takeo, a male-gendered robot, and Momoko, a female-gendered robot, wearing an apron during the play. The roles of Yuuji and Ikue are performed by two human actors, while the two robots are performed by two bright yellow Wakamaru robots from Mitsubishi Heavy Industry. Each robot is 1 m tall, they are mounted on wheels, and they can move their arms and rotate their heads and bodies. They speak Japanese with mediated synthetic-sounding

voices. Clearly, the two robots are programmed or remote controlled to move around the stage floor and to communicate with the two humans and with each other.

The gender roles and nuclear family structure represented in this play follow the conventional ideals in Japanese society, and hereby seems to confirm what anthropologist Jennifer Robertson has concluded in her study of the implementation of social robots in Japan, namely that traditional social values and structures are reinforced through the implementation of robotic technologies in care-taking industry and entertainment [2,3]. Parts of the conversation between the robots and the human beings in *I, Worker* reveal various aspects of identity crisis, both on parts of the humans as well as for the robots. The male robot Takeo, for example, does not want to leave the house to help Ikue go shopping. In the robot's self-understanding, this aversion to go outside in order to help the human being is a major crisis because it contradicts the very purpose of being a robot. It states: "a robot should be working, we are created to work." It turns out that Yuuji, the male human being, is out of job and also does not want to leave the house. This poses a problem for his wife Ikue in her relationship to her parents. In Ikue's conversation with the robot Takeo, she explains how her parents keep bothering Yuuji that he has no job: "My parents say that to Yuuji all the time: 'People are meant to work. You feed yourself by the sweat of your brow.'" [1] In this way, the dialogue on stage poses questions concerning issues such as gender roles, family values and social expectations, which reflects real-life issues that many Japanese face in the post-bobble era of economic recession.

The surprising effect of the play is when emotions related to existentialist concerns are expressed not by a human being, but by a machine. This *Verfremdung* effect created by the juxtaposition of human emotions expressed through machines is constantly and explicitly addressed in the play, and becomes the key to questioning the possible differences and similarities between humans and robots. In the scene mentioned at the beginning, Yuuji tries to cheer up the miserable robot Takeo by playing the theme song from RoboCop on his CD player, but then realizes that playing a heroic battle tune was inappropriate and he apologizes to the robot [4]. As audience, we identify with the situation and would have felt emotional too, and therefore do not really believe it when the robot replies "It is nothing". It is also difficult to believe when the other robots states "robots cannot lie."

2. Theatre-Based HRI

But of course robots lie. In fact, a lot of the testing of robots within science and technology are based on deception. The method known as the Wizard-of-Oz (WoZ) is often applied in the field of HRI, human-robot interaction. This method includes a person who operates the robot remotely, usually hidden from the test person who interacts with the robot. The Wizard method may be used in cases where the robots are not sufficiently advanced to interact autonomously with humans, or as a test of the robot at early stages of an interaction design process [5]. Furthermore, the notion of theatre is applied as a specific approach to such HRI research situations. A number of HRI research projects that apply theatre as a methodology present a rather conventional understanding of theatre and its aesthetics to argue for the relevance of theatrical scenarios for HRI. It is commonly thought that actors and robots have the goal to "get as close as possible to the unobtainable ideal", which is presenting the actual character to the audience. In this process, "the robot/actor must appear to be something it is not".

Hence, many HRI researchers think of actors and robots on equal terms and argue that the theatrical stage for the human actor is equivalent to the real-life scenarios for the social robot - both situations rely on the "as-if" effect[6].

One example that I will refer to here, is LIREC, an integrated project based at University of Hertfordshire in which researchers use a Theatre-based Human-Robot Interaction methodology (THRI). In the test situation, an actor plays the role of a robot owner in front of a group of volunteer test persons. The robot too is "acting" by remote control because it is made to simulate actions that no robot can actually perform yet, such as robust navigation in everyday environments, smooth natural language interaction, and "social intelligence" in interaction and dialogue with the human owner [7,8].

The scenarios with the actor and the robot includes extensive use of the Wizard of Oz method. In the left hand side of the room and off-stage, technicians operate the robots without the audiences' knowledge. According to the research report, the actor who pretends to be the robot owner will stay in character during the short scenarios and during the discussions afterwards in order to "make the audience believe that the experiment was a true representation of an owner and robot"[7]. Likewise, those technicians who control the robots are placed "out of the audience's direct line of sight to allow the audience to 'forget they are there'"[7]. This is an example of how scientific testing of human-robot relations apply performance as a means of producing knowledge about human response. However, as roboticist Laurel Riek notes, the robot in this kind of remote controlled puppeteering serves more as a proxy for humans and less as an independent entity. In fact, the situation may be described not as a human-robot interaction, but as a human-human interaction via a robot [5].

3. Performatives and Infelicities

In both of my examples robots are programmed or remote controlled to act as if they have a smooth and natural interaction with humans. There is a difference, however, between robots staged in a theatre production and the theatre-like scenarios for testing human-robot interaction. In order to analyse this difference I will turn to linguistic philosopher J.L. Austin's theory on performative speech acts. I should mention that I am not trained in philosophy, but have encountered Austin's speech act theory through the field of performance studies. In performance studies, Austin is credited for coining the term "performative", which denotes the kind of linguistic utterances that perform or "do" an action when they are spoken. One of Austin's examples that appear in his lecture notes *How To Do Things With Words* from the 1950s is the utterance "I do" in a wedding ceremony: by uttering the words under the proper circumstances a conjugal union is established[9]. A performative sentence is therefor not just simply to state something, but is to carry out an action through speech.

Austin's performatives are relevant for exploring social interaction because he is concerned about intentionality (of the speaker) and effect (on the listener) in everyday situations of linguistic communication. Hereby, Austin's linguistic theory is linked with anthropological and social performance through semiotics and notions of social and cultural conventions[10]. Discussing the properties of performatives, Austin is less concerned about whether an utterance is true or false; instead he focuses on the success or failure of an utterance, and creates a distinction between "felicitous" and "infelicitous" utterances. The utterance "I do" may thus only be successful or felicitous

if it fulfils certain conventions, for example if it is spoken by a person involved in a wedding ceremony, and by a person who is not already married. In his argument, Austin presents a number of infelicities, and one of them concerns situations where there is no true intention behind the utterance, for example when a performative is uttered on a theatre stage or in poetry.

Austin claims: "a performative utterance will, for example, be in a peculiar way hollow or void if said by an actor on the stage, or if introduced in a poem"[9]. Language in such circumstances, Austin continues, "is in special ways – intelligibly – used not seriously, but in ways parasitic upon its normal use." [9] In other words, Austin is only interested in performative utterances spoken in ordinary everyday circumstances, and he dismisses performative utterances that are spoken in what he calls "unserious" ways, that is aesthetic or fictional situations such as theatre and poetry.

4. Comparison: Two Types of Robot Theatre

If we look at the two types of theatrical settings I presented above, we can say that Austin's notion about the use "parasitic upon its normal use" takes different forms. In the case of the Hertfordshire test for robotic home companions, the test persons know of course that they are engaging in a scientific experiment, but they may not be aware of the theatrical framing of the entire event. The test persons may very well think that robots are able to actually behave and speak as they do in the scenarios.

Hirata Oriza's play *I, Worker* resembles the robot testing at Hertfordshire University: here too the robots are manipulated to act as if they can carry out robust navigation in everyday environments, engage in smooth natural language exchange, and display "social intelligence" in interaction and dialogue with the human owners, in this case Yuuji and Ikue. Because much of the acting on stage is based on linguistic dialogue, the smoothness of verbal exchange is convincing, exactly because the robots speak as most audiences would expect them to speak, namely in synthetic voices and with some slow response as if they need to perceive and process the utterances made by the humans.

There is, however, a significant difference in how these two types of staging are perceived. Because *I, Worker* takes place in a theatre, the audience engage knowingly and voluntarily in the "deception" of the theatrical framing, and they are not fooled in the same way as the test audience at the Hertfordshire laboratory.

Ishiguro Hiroshi, the robot scientist, explains the reasons for him to collaborate with Hirata Oriza in an essay on robot science. For Ishiguro, the naturalism in Hirata's play is the main reason for the collaboration because this is where the artistic and the scientific approach have central elements in common. One of Ishiguro's goals is to approach the notion of a human heart. "We robot scientists cannot programme a human heart into a robot", Ishiguro writes. "We can programme a function that makes it look as if the robot has a heart. But the problem is that we do not know what it looks like to have a heart." [11] Ishiguro notes that psychologists and cognitive scientists study such aspects of human nature, but their experiments are always carried out in the controlled environments of the laboratory, and therefore "cannot clearly explain how humans express their heart in the midst of multiple stimuli in everyday situations." [11] When scientific approaches such as robotics, psychology and cognitive science cannot give the answer, Ishiguro turns to the artistic creation of Hirata's theatre work.

5. Naturalism as a Set-up

In the play *I, Worker* the relationship between reality and fiction is challenged by the aesthetic format of the play. *I, Worker* resembles a naturalistic play because it contains what performance studies scholar Peter Eckersall describes as "visceral conditions of life through corporeality and sensation and without moral intervention"[12].

Hirata Oriza and his Seinendan Theatre Company is acclaimed for the use of contemporary colloquial language, and many critics praise the way in which Hirata dissects ordinary everyday situations in his plays. This is true for the play *I, Worker* too, except for the fact that the colloquial dialogues of everyday life here takes place not between human beings, but between humans and machines. The elements of improvisation and individual response usually associated with naturalistic theatre, however, is at stake in this play. The naturalistic mode is a set-up: the conversation is staged to appear natural. In the intense dialogue with machines, there is no room for the two human actors' emotional improvisation or personal response. Every single word, every slight movement, is conceived, staged, directed, and rehearsed. In Hirata's play, the actors do not need to have a heart, and this corresponds with Ishiguro's own views as an engineer when he creates robots. We as audience realize that the human actors are programmed and controlled in a way similar to the robots. Naturalism as an aesthetic framework becomes an experimental form: it is used as an epistemological tool for reflections that makes us aware of broader issues, such as questioning the fundamental issue of what is "natural" or "normal" in human beings.

6. Differentiation of Citationality

This points back to Austin's phrase about the "parasitic use" of language and the critique this particular part of his theory has generated. The philosopher Jacques Derrida offers a productive elaboration on Austin's remark. Rather than dismissing the infelicities uttered by an actor on stage or introduced in a poem such as Austin does, Derrida turns the argument around and instead suggests that there is a citational element, an iteration, in all performatives.

The words "I do" in the wedding ceremony is felicitous because this utterance is coded through repetition; the words have been uttered before, within an iterable set of circumstances and context. All performative speech is in some sense always a citation. Derrida does not dismiss citationality as a criteria for infelicities, instead he proposes a differentiation of citationality. He speaks of a "relative purity" of performatives that "does not emerge *in opposition* to citationality or iterability, but in opposition to other kinds of iteration within a general iterability"[13].

In similar veins, the robots in *I, Worker* appears at first sight to be an infelicitous and parasitic use of the norms of the human. However, using Derrida's argument, we may elaborate on this perspective and understand the robots as iterations or citations of human beings in movement, behaviour, and speech. The performative appearance of the robot is not in opposition to the "purity" of human intention and emotion, but in opposition to other kinds of iteration within a general iterability. If robots are citations of humans, humans too are citation of iterable norms and conventions. This aligns with Judith Butler's theories of performativity, in which the subject is constituted by performative repetitions of social codes[14]. A further investigation into the cross-sections between performativity and robotics may bring about surprising perspectives

on robotic subjectivity and identity, and hereby contribute to understanding the multifaceted aesthetic agencies and properties pertaining to robots in present and future everyday life.

So let me finish by returning to the ambiguous and paradoxical statement uttered by a robot on stage: "robots cannot lie." Of course robots lie. They do so every time aspects of human behaviour, social intelligence or emotional response are attributed to the machine, whether in a robot-human theatre production, or in simulated robotic home companion tests. The robot acts as a "parasite" on human interaction, because the robot is attributed intention by the human. Such performance of human-ness by something other than human is "in a peculiar way hollow or void", as Austin would say, but these ways "parasitic upon its normal use" is exactly what art contributes to the field of robotics: it opens for new insight of how humans project and respond to their own behaviour and emotions in their interaction with technology. We may realize that human beings, like robots, are constituted through a kind of social "programming" that makes us behave within certain norms and conventions. We may also learn from theories of performativity how such conventions may be negotiated and provide agency for robots and humans alike.

References

- [1] Hirata O., Hataraku watashi, *Robotto Engeki*, Osaka Daigaku Shuppankai, Osaka, 2010.
- [2] J. Robertson, Robo Sapiens Japonicus. Humanoid Robots and the Posthuman Family, *Critical Asian Studies* 39:3 (2007), 369-398.
- [3] J. Robertson, Gendering Humanoid Robots: Robo-Sexism in Japan, *Body & Society* 16:2 (2010), 1-36.
- [4] B. Poledouris, RoboCop themesong, 1987. <https://www.youtube.com/watch?v=D7rjLQuW2nQ>, accessed 14 August 2014.
- [5] L.D. Riek, Wizard of Oz Studies in HRI: A Systematic Review and New Reporting Guidelines, *Journal of Human-Robot Interaction* 1:1 (2012), 119.
- [6] D.V. Lu and W.D. Smart, Human-robot Interactions as Theatre, *2011 RO-MAN: 20th IEEE International Symposium on Robot and Human Interactive Communication*, July 31- August 2, 2011, Atlanta GA. <http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=6005241>, accessed 17 August 2014.
- [7] A.R. Chatley et al., Theatre as a Discussion Tool in Human-Robot Interaction Experiments. A Pilot Study, *2010 Third International Conference on Advances in Computer-Human Interactions*, <https://uhra.herts.ac.uk/dspace/bitstream/2299/4354/1/903863.pdf>, accessed 4 May 2012.
- [8] D.S. Syrdal et al., The Theatre methodology for facilitating discussion in human-robot interaction on information disclosure in a home environment, *2011 RO-MAN: 20th IEEE International Symposium on Robot and Human Interactive Communication*, July 31- August 2, 2011, Atlanta GA. http://www.researchgate.net/publication/224256263_The_Theatre_methodology_for_facilitating_discussion_in_human-robot_interaction_on_information_disclosure_in_a_home_environment, accessed 17 August 2014.
- [9] J.L. Austin, *How To Do Things With Words*, Harvard University Press, Cambridge, Mass., 1975.
- [10] Marvin Carlson, *Performance. A critical introduction*, Routledge, New York, 1996.
- [11] Ishiguro H., *Robotto kenkyū towa ningen no kokoro o shirukoto*, *Robotto Engeki*, Osaka Daigaku Shuppankai, Osaka, 2010.
- [12] P. Eckersall, Hirata Oriza's 'Tokyo Notes' in Melbourne: conflicting expectations for theatrical naturalism, *Outside Asia: Japanese and Australian identities and encounters in flux*, Japanese Studies Centre, Clayton, Vic., 2011.
- [13] J. Derrida, *Signature Event Context*, *Limited Inc.*, Northwestern University Press, Evanston, 1988.
- [14] J. Butler, *Bodies That Matter. On the Discursive Limits of 'Sex'*, Routledge, New York, 1993.