

# Receptionist or Information Kiosk: How Do People Talk With a Robot?

Min Kyung Lee, Sara Kiesler, Jodi Forlizzi

Human-Computer Interaction Institute

Carnegie Mellon University

Pittsburgh, PA USA

{mklee, kiesler, forlizzi}@cs.cmu.edu

## ABSTRACT

The mental structures that people apply towards other people have been shown to influence the way people cooperate with others. These mental structures or schemas evoke behavioral scripts. In this paper, we explore two different scripts, receptionist and information kiosk, that we propose channeled visitors' interactions with an interactive robot. We analyzed visitors' typed verbal responses to a receptionist robot in a university building. Half of the visitors greeted the robot (e.g., "hello") prior to interacting with it. Greeting the robot significantly predicted a more social script: more relational conversational strategies such as sociable interaction and politeness, attention to the robot's narrated stories, self-disclosure, and less negative/rude behaviors. The findings suggest people's first words in interaction can predict their schematic orientation to an agent, making it possible to design agents that adapt to individuals during interaction. We propose designs for interactive computational agents that can elicit people's cooperation.

## Author Keywords

Agent, robot, schemas, scripts, dialogue, cooperation, human-robot interaction (HRI), conversational interface, speech interface, social robots, design

## ACM Classification Keywords

H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

## General Terms

Human Factors

## INTRODUCTION

The CSCW community has a longstanding interest in online agents, interactive devices, and robots used in collaborative interactions. For example, agents can assist collaborative learning and group coordination [11, 12]. Robots can

Permission to make digital or hard copies of all or part 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. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

*CSCW 2010*, February 6–10, 2010, Savannah, Georgia, USA.

Copyright 2010 ACM 978-1-60558-795-0/10/02...\$10.00.

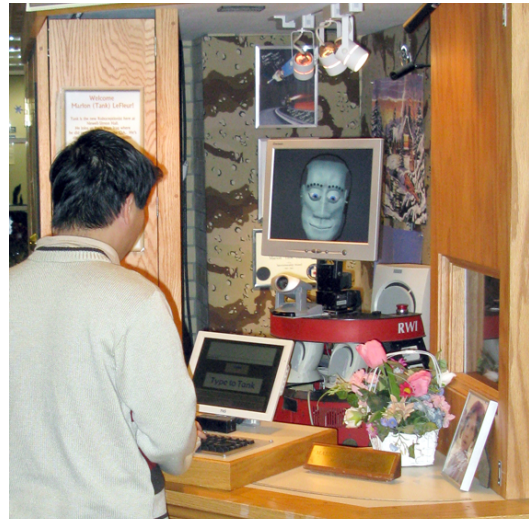


Figure 1. A photo of Roboceptionist, a receptionist robot located in a high-traffic entrance area in an academic building.

introduce and guide groups of visitors in a variety of settings such as museums, subways, airports, and other public places [7, 17, 28, 29, 42, 46].

Computer agents or robots that work in public settings raise some challenging design questions. To be successful in imparting guidance or answering questions, they must elicit cooperation from busy workers or visitors who are total strangers. Furthermore, these interactions are likely to occur in the presence of others. People care about their self-presentation to others in public [20]. If they feel nervous or embarrassed, those feelings may negatively impact their willingness to cooperate.

Researchers have suggested many directions for design to support interactions in public settings with agents or robots. For example, Bickmore et al. [5] sought to make interacting with agents in public comfortable by bringing the agents to human height and creating natural eye-gaze toward speakers. They involved bystanders by creating a back-screen that displayed the dialogue between the robot and a user interacting with the robot. To attract visitors to a museum guide robot, Thrun et al. designed it to express happiness through its facial expressions when more visitors approached the robot [45]. Shiomi et al. also found that a

robot can increase user engagement in a museum by referring to visitors by their names [43].

Some researchers have argued that creating user models, for example, by learning from peoples' repeated interactions over time, can support adaptivity in an agent or robot's interactions with people [27]. In line with the theory of regulatory fit [9], an agent that adapts to people's orientation to computational agents might elicit more cooperation than an agent that does not adapt to this orientation. For example, people may be oriented to treat an agent or robot as a humanlike being or, alternatively, as a computational tool. According to regulatory fit theory, the robot should act to support these different orientations. Nass et al. [38] showed that extroverts found an extroverted agent more attractive and credible than introverts did, whereas introverts found an introverted agent more attractive and credible. Goetz and Kiesler [19] showed that matching a robot's personality to users' serious or playful tasks elicited more cooperation from them.

One way to create adaptivity to people's orientation is to use learning algorithms or to detect people's demographic characteristics to build a model of people with different orientations. Doing so may be difficult in public settings, where many encounters will be new and where the population is diverse, mobile, and busy. In this paper, we argue that we can build reasonable adaptivity in an agent or robot if we can use people's initial verbal cues to estimate their schema for the agent or robot.

A schema, in our meaning, is a mental structure or representation of any object or phenomenon encountered in the world [1]. A schema can determine the way people approach a situation. Previous research suggests that the form factor of a computational agent influences their schema and willingness to cooperate with the agent. In one study, participants cooperated more with an agent that looked like a person than an agent that looked like a dog, and more with a realistic dog agent than with a cartoon dog agent [39]. In another study, participants took less responsibility for the successful completion of the task when working with a humanlike robot than a machine-like robot [24].

Even given the same external form factor, people have different orientations to agents and robots. Friedman et al. showed that some people think of AIBO, a dog-like robot, as a technological entity whereas others attribute more lifelike qualities to it [18]. In studies of hospital delivery robots, researchers observed that some employees anthropomorphized the robot whereas others regarded the same robot as a machine [36, 44].

Here, we explore two different schemas for an interactive robot. We used the archival data from the dialogue logs of the Roboreceptionist robot, a receptionist robot at Carnegie Mellon University that has been located in a high traffic area in the Newell Simon Hall building for about 5 years [21, 22] (Figure 1). The robot is partially humanlike in that

it can speak and has a screen "head" that turns to look at passers-by. It greets people, gives directions to rooms in the building, looks up weather forecasts, and tells visitors its personal stories. Visitors can type to the robot, and the logs of these interactions afford an opportunity for understanding people's natural, spontaneous interaction styles with a robot.

In our encounters with people, greetings indicate our willingness to engage socially with others [25]. For example, if you come into a store and silently put down your money, your behavior may indicate that you are not in the mood for social conversation. A greeting such as "Hi, nice day," signals a more social orientation. We argue that greetings may predict the schema people have for the receptionist robot. We categorized visitors who interacted with the robot in two groups, depending on whether they greeted the robot or not, and analyzed their subsequent verbal behavior. We show that whether people greeted the robot or not at the beginning of conversations predicted much of what followed.

### **From Schemas to Scripts**

We posit that when people encounter a computational agent or robot in a receptionist role, they will have one of two general orientations toward it, either a human social schema such as a service person or a computational tool schema such as an informational kiosk or display. We further posit that people's orientation will elicit different scripts for their subsequent behavior with the receptionist robot. Schemas activate specific behavior through scripts. A script is a "conceptual representation of stereotyped event sequences [1]." For example, when people enter a restaurant, they follow a standard sequence of events and typical activities such as making small talk with the serving person, placing an order, tipping, and collecting their belongings before leaving. Likewise, the script for interacting with a human receptionist is cordial whereas the script for interacting with an information kiosk is utilitarian.

Having scripts for daily activities reduces people's cognitive load and allows them to focus on more high-level activities. Scripts also guide people to appropriate behavior in different settings and cultures [23]. Thus scripts are very common in everyday behavior. We have scripts for eating in restaurants, for shopping in grocery stores, for visiting museums, for attending sports events, and for holiday dinners with family. People construct scripts for specific, particular contexts through direct and indirect means. Direct script acquisition involves learning through interaction experience with other people, events, or situations. Direct experience tends to initiate a script development process. Indirect script acquisition occurs by means of communication or media. Watching people interact with a robot in a movie or science fiction novels could give people a script for interacting with a robot.

People may consciously choose to perform scripts when facing new situations, although the scripts themselves might

be unconscious. Starting a script performance usually entails a commitment to finish it. For example, one does not readily leave a restaurant once seated or walk out of a dentist's office before the dentist is through.

### **Roboceptionist Scripts**

We argue that visitors' scripts for the Roboceptionist will have drawn on their prior interactions with other service personnel and human receptionists or with other computing machines in public settings, tilting them to have either a schematic orientation to a robot as human service person or as computational tool. In a relevant paper, Fischer reported that the orientations that people held toward mobile robots with varying anthropomorphic forms influenced their instructional strategies and the prosodic strategies that they used to give instruction to the robot [14]. Fischer proposed that users' choice of dialogue beginnings might have predicted speakers' concepts of the human-robot situation [15]. One of our goals in this paper is to follow up on this idea and to determine how people's initial dialogue predicts their orientation to a robot.

In our study, we focused on two alternative scripts that people might apply when interacting with the Roboceptionist robot, that is, the script for interacting with a receptionist or other service person or the script for interacting with an informational computational tool such as an information kiosk. We believe these scripts will arise from the schema that people have for the robot. According to social actor theory, people interact with machines as though they are other humans [41], but many studies show that this response depends on other factors, such as the form factor of the machine [39], whether people think their interaction is with a computer or person [36], the presumed gender of the agent [40], or even its nationality [32]. In the current study, we did not manipulate the form factor or any other attributes of the robot. Instead we assumed that people vary in their schematic orientation and aimed to predict this orientation.

What would be involved in a receptionist script? If this script is evoked when interacting with the Roboceptionist, we believe people would apply the sequence of activities common in everyday interaction with a receptionist or other service personnel. Typical sequences in these scripts might include casual greeting, small talk, instrumental questions, information exchange, and leave-taking [25]. The script also should follow general social norms for weak tie interactions, such as maintaining politeness, not insulting the other personally, and little personal disclosure. The script also should accommodate conversational grounding [10].

On the other hand, the same robot might invoke a more computational machine schema. People today have had experiences with interactive computational machines such as information kiosks or GPS car navigators. In many such systems, the agent or computer looks like a machine or exposes its mechanical parts such as a camera and a laser.

The machine's voice may have a mechanical tone, and people may have to type to the machine instead of speaking to it. These mechanical qualities of the computational machine could reinforce the feeling that this is a machine rather than a social actor.

The Roboceptionist, while somewhat humanlike because of its displayed face and conversational speech, also has many machinelike qualities. The face is a display on a computer screen, the voice has a mechanical quality to it, and users type to talk to it. Due to these mechanical qualities, people might draw analogies between the Roboceptionist and other computing machines. Such machines often act as computational tools that support people's utilitarian goals such as guidance in a museum or in a car. People interact with these devices by directly specifying their goals and instructions by using a graphical user interface, or typing or speaking keywords. In the case of a GPS car navigator, people specify their destination either by typing the destination on its screen or speaking a keyword. The GPS system provides direction in natural human language. When interacting with these types of devices, people typically use an instrumental script: instruct the machine, wait for its reply, and correct it if needed. The script is for communication of intent in a direct manner, and does not use relational conversational strategies.

### **Greeting as an Indicator of the Script**

From previous research and literature on scripts, we hypothesized that whether visitors greeted the Roboceptionist or not would predict which script they performed when they interacted with the Roboceptionist. One of the characteristics of a script is that, once people choose to enter a particular script, they are less likely to stop the script until its end, unless unexpected breakdowns happen [1]. As the greeting is the first interaction that happens in human social encounters, the greeting could predict whether or not people have followed a script for human social interaction with a receptionist or other service person or, instead, a script for interacting with a machine.

Fischer's study of people's instructions to a robot showed some evidence for this argument [14]. She reports that people who greeted the robot tended to instruct the robot in full sentences rather than phrases without verbs. Those who greeted the robot also tended to refer to the robot using personal pronouns, "he" or "she," rather than "it," and they used structuring words (e.g., "next," "then"). Encouraged by these findings, we developed hypotheses for people's conversation patterns depending on whether they initially greeted the robot or not.

### **Hypotheses**

From the above arguments, we predicted that people who greet a robot will follow social norms for human-human communication more than those who do not greet a robot. We developed the following specific hypotheses:

**H1.** People who greet the robot will exhibit more conversational grounding behaviors than people who do not greet the robot.

**H2.** People who greet the robot will use more relational conversation strategies than people who do not greet the robot.

**H3.** People who greet the robot will be less likely to use computer command input styles than people who do not greet the robot.

## **METHOD**

The method of this study entailed an analysis of utterances that people typed to a receptionist robot over a period of five and a half months. We grouped people into two groups, those who greeted the robot and those who did not, to show how using a greeting predicted subsequent conversation.

### **Roboceptionist**

As noted above, the Roboceptionist robot, named “Tank,” is situated in a booth in a lobby near the main entrance of the university building. The robot is built with a B21r mobile robot and a 15” flat-panel LCD screen mounted on a pan-tilt unit. It has a caricatured humanlike male face on the screen. It changes its facial expression and rotates its head to look at passers-by. The robot speech is generated from text using the Cepstral text-to-speech engine [8], and is automatically synched with its lip movements.

To interact with Roboceptionist, people type on a keyboard located in front of the robot. Upon a typed query, the robot gives directions to campus offices and buildings, looks up office numbers of employees, and reports on the weather. The robot also enacts its persona by describing some personal history and preferences if visitors ask. The examples of its personal story include its work experience at the CIA and in Afghanistan, and its family, girlfriend, and dog. The robot’s booth contains various props such as the robot’s photograph with soldiers in the desert to reinforce the robot’s persona.

The robot uses a rule-based, pattern-matching parser, modified from Aine [2] to generate responses to people’s input. During the study, the robot responded to every person’s initial input in the same manner, whether they gave a greeting or not. The robot is passive in that visitors always initiated a conversation, and the robot only responded to their utterances.

People who work in the building can swipe their ID cards or credit cards in a card reader so the robot can call them by name. However, our analysis showed that people rarely swiped their cards. For more details on the Roboceptionist, please refer Gockley et al. [21].

### **Data Collection and Coding**

#### *Dialogue log data*

We logged 1180 interactions over 5.5 weeks in March and April 2008. Each interaction was defined as a dialogue that occurred from the moment a person approached the robot

until he or she left, as detected by the laser. The unit of analysis is the interaction. When the same utterances were observed multiple times in one interaction, they were calculated as happening once, so that we do not over-count and can measure the percentage of persons who exhibited particular behaviors.

#### *Video data*

To protect people’s privacy, the dialogue log data did not contain any contextual information about persons who interacted with the robot. However, we obtained permission to record Roboceptionist-person interactions for one week in March and April 2009 using the security camera installed in the Roboceptionist booth. We coded persons’ gender, whether they were alone or with others, and guessed their ages. These codes were compared with the presence of greetings in dialogues with the Roboceptionist so we could evaluate whether gender, age, or being alone predicted greeting the robot.

#### **Measures**

We measured attributes of each interaction, and person utterances in each interaction. The unit of coding was an exchange between a person and the robot. A coding scheme for topic was based on coding 197 individual interactions collected over one week in March 2008 by Lee and Makatchev [31, 34]. A coding scheme for linguistic styles was drawn from the common ground and politeness literature [6, 10].

#### *Person Utterance Measures*

We coded whether people greeted the robot or not (such as “Hi,” or “What up”).

*Grounding behaviors* had four attributes: relevancy, acknowledgement, repair (rephrase), and misunderstanding. Relevancy was coded if a person built upon the robot’s previous utterance. Acknowledgement was coded if a person explicitly expressed his or her understanding of the robot’s utterance. Repair was coded if a person rephrased his or her previous utterances. No one misunderstood the robot’s utterance, so this factor was not considered in our results.

*Relational behaviors* were measured by people’s politeness, sociable behaviors, and negative behaviors. *Politeness* was counted when a person said farewell, thanked the robot, made an apology, or said phrases that express courtesy or etiquette (e.g., “please,” “Good evening Mr. Tank,” “would you mind telling me your name again”).

*Sociable behaviors* were measured by whether people made small talk, called the robot’s name during the interactions, made empathetic comments for the robot, introduced themselves or others to the robot, complimented the robot, or told a joke to the robot.

*Negative behaviors* were measured by whether people said nonsense or insulted the robot, or asked it intrusive questions (e.g., “What is your GPA?,” “Are you gay?”).

Topics were coded as instrumental, robot-related, and person-related, and others.

*Instrumental* topics were measured by whether people asked for information about the university where the robot was situated, locations of places (e.g., restaurant or bathroom), information about employees (e.g., office number, phone number, or email), travel information (e.g., how to get a taxi), information about Pittsburgh weather, or the current date and time.

*Robot-related* topics were measured by whether people asked about the robot’s stories and information about the robot (e.g., its name, age, preferences, family, friends, pets).

*Person-related* topics were measured by whether people talked about their feelings or events in their lives.

We used a code, “*other topics*”, for idiosyncratic comments and questions (e.g., “tell me how babies are born.”).

*Sentence structure* was a coding of sentences, whether they were imperative, interrogative, declarative, or contained no verb.

#### Interaction

We measured the total duration of each interaction and the total number of utterances a person said.

One coder performed all of the coding, and another coder coded ten percent of the data. They compared their results until they reached agreement.

## RESULTS

We conducted multi-level repeated measures analyses of variance (ANOVAs) to test each of the hypotheses, comparing people who greeted the robot with people who did not greet the robot as a between groups variable, code type as a within groups variable, and each person’s interaction as a random control. We report the percentage of the interactions that included behaviors relevant to our hypotheses.

On average, 43 interactions with the robot happened per day. The average interaction duration was 55 seconds and four interactive exchanges (turns) per person. Overall, half of the interactions included a greeting at the start and half did not (Table 1). Those who greeted the robot interacted with the robot longer (Greeting:  $\bar{M} = 78.4$  seconds, No Greeting:  $\bar{M} = 31.4$  seconds). This difference in the interaction duration happened because those who used a greeting typed more words (Greeting:  $\bar{M} = 12.6$  words, No Greeting:  $\bar{M} = 8.4$  words,  $p < .0001$ ), and took more turns (Greeting: Mean = 5.5 turns, No Greeting:  $\bar{M} = 3.2$  turns,  $p < .0001$ ). We ran additional regression analyses controlling for number of words or turns. These analyses show that the topics, tone (social, polite, and less negative behaviors), and sentence structures still differed across those who greeted versus those who did not.

Only 21.4% of those who greeted the robot left immediately afterwards. Some of these people just wanted to say hello

(P: “Good morning to you.” R: “Hello,” P: “Nothing, just wanted to say hi.”). In addition, 18.5 % of those who did not greet exhibited only abusive behaviors such as typing insults or nonsense. Those interactions, lacking conversation, had to be excluded for the subsequent analysis. Very few people swiped their cards, and the number did not differ across the two groups (G = 2%, NG = 3%).

Group	Percentage
<b>Greeting</b>	<b>49.5 % (N=585)</b>
Greet & leave	21.4 % (N=125)
Greet & converse	78.6% (N=460)
<b>No Greeting</b>	<b>50.5 % (N=595)</b>
Abusive behavior only	18.5 % (N=110)
No greeting & converse	81.5% (N=485)

**Table 1. Percentage of interactions that include greeting and those that do not include greeting at the beginning of their interactions with the Roboceptionist.**

#### Grounding Behavior

We predicted (H1) that those who greeted the robot would use more grounding strategies such as acknowledging, making relevant follow-up comments, and rephrasing. As predicted, people who greeted the robot exhibited grounding behaviors that they would do with other people more frequently than those who did not greet the robot. The ANOVA shows a significant greeting vs. no greeting main effect ( $F [1, 943] = 14.5, p < .001$ ), a grounding type (acknowledge, repair, relevancy) main effect ( $F [2, 1886] = 18.3, p < .001$ ) and a marginal interaction effect ( $F [2,1886] = 2.5, p = .08$ ).

Grounding Behavior	Greeting	No Greeting	<i>p</i>
Relevancy	0.35	0.23	<.001
Acknowledgement	0.23	0.15	<.01
Repair	0.27	0.23	n.s.

**Table 2. Comparison of percentage of grounding behavior between people who greeted the robot and those who did not. (The *p* value is calculated using ANOVA planned contrasts.)**

More people who greeted the robot than those who did not acknowledged that they heard and understood the robot’s utterances by typing “yeah,” “ok.” The planned contrasts show significant differences between those who greeted the robot and those who did not in acknowledging the robot’s response ( $F = 7, p < .01$ ).

People who greeted the robot tended to continue their dialogues, building on what the robot said to them as compared with those who did not greet the robot ( $F = 17, p < .001$ ). They also asked follow-up questions based on the stories that robot told them or made statements about the stories.

In the following dialogue, the robot did not understand most of what the person said, yet the person commented on the robot's response, rephrased his or her utterances, and continued the dialogue:

P: ok tank.  
 R: That's me!  
 P: how are you today?  
 R: Dr. Spitola suggested that a self defense seminar might help to reassemble my self esteem. He assures me that just because I'm armless doesn't mean I'm any less of a person. It doesn't mean I have to take this abuse. I don't have to settle for sexual harassment from that disgusting buffer. The good doctor has arranged for me to attend a self defense camp designed especially for abused machines. I'm looking forward to getting away for a while.  
 P: ouch. are you depressed?  
 R: I didn't understand that. It might help if you rephrase it.  
 P: are you depressed?

**Example 1. In this excerpt, the person acknowledged the robot's response and rephrased his utterance.**

**Relational Behavior**

We predicted (H2) that those who greeted the robot would use relational conversational strategies such as small talk, and empathetic and polite remarks more frequently than those who did not.

*Sociable Behavior*

The repeated measures ANOVA of social interaction used seven codes shown in Table 3.

Sociable Behavior	Greeting	No Greeting	p
Small talk	0.22	0.03	<.0001
Empathetic comment	0.02	0.00	n.s.
Joke	0.01	0.00	n.s.
Call robot's name	0.08	0.05	<.01
Introduce self	0.03	0.01	n.s.
Introduce others	0.004	0.00	n.s.
Compliment	0.03	0.01	n.s.

**Table 3. Comparison of percentage of sociable behavior between people who greeted the robot and those who did not. (The p value is calculated using ANOVA planned contrasts.)**

As predicted, the analysis shows a significant main effect of greeting ( $F [1, 943] = 85, p < .0001$ ), significant differences among the codes ( $F [6, 5658] = 57, p < .0001$ ) and a significant interaction ( $F [6, 5658] = 33.9, p < .0001$ ). The interaction reflects the fact that those who greeted the robot were much more likely to engage in small talk and also call the robot by name. People who greeted the robot initiated

small talk with the robot more frequently than those who did not greet the robot. They asked "How are you?" and "How's your day?" ( $F = 273, p < .0001$ ). Those who greeted the robot also called the robot by name during their conversation more frequently than those who did not ( $F = 9.3, p < .01$ ).

Although the difference was not statistically significant, those who greeted the robot also tended to make empathetic comments on what the robot said:

P: Do you have a girlfriend?  
 R: I had another date. Heinzilla, the score board at Heinz Field. She was a bit large. Okay, she's massive, but who am I to judge? I thought, if she can look past my lack of arms, I can look past her gargantuan proportions. She was a great conversationalist, and it was wonderful to talk with a woman who knew so much about sports. I went home and was very excited, but unfortunately I can never see her again.  
 P: Why not  
 R: I rented a video of great moments at Heinz Field. You know, just to see her in action, but there was a problem. She got way too much action. I couldn't believe how many people have scored with her. Two at the same time! No, Heinzilla is not the girl for me.  
 P: I'm sorry to hear that.

**Example 2. The excerpt shows a person's empathic comment to the robot.**

*Polite Behavior*

The repeated measures ANOVA of polite interaction used four codes shown in Table 4.

Polite Behavior	Greeting	No Greeting	p
Thank you	0.16	0.07	<.0001
Farewell	0.21	0.11	<.0001
Apology	0.02	0.01	n.s.
Courtesy	0.06	0.03	n.s.

**Table 4. Comparison of percentage of polite behavior between people who greeted the robot and those who did not. (The p value is calculated using ANOVA planned contrasts.)**

The analysis shows a significant main effect of greeting ( $F [1, 943] = 28, p < .0001$ ), significant differences among the codes ( $F [3, 2829] = 63.9, p < .0001$ ) and a significant interaction ( $F [3, 2829] = 7.4, p < .0001$ ).

The planned contrasts show those who greeted the robot, as compared with those who did not, were more likely to thank the robot ( $F = 24.3, p < .0001$ ), and to say farewell to the robot ( $F = 32.2, p < .0001$ ). When closing a conversation, those who greeted the robot tended to inform









