Tracking Anthropomorphizing Behavior in Human-Robot Interaction

KERSTIN FISCHER, University of Southern Denmark, Sønderborg, Denmark

Existing methodologies to describe anthropomorphism in human-robot interaction often rely either on specific one-time responses to robot behavior, such as keeping the robot's secret, or on post hoc measures, such as questionnaires. Currently, there is no method to describe the dynamics of people's behavior over the course of an interaction and in response to robot behavior. In this paper, I propose a method that allows the researcher to trace anthropomorphizing and non-anthropomorphizing responses to robots dynamically moment-by-moment over the course of human-robot interactions. I illustrate this methodology in a case study and find considerable variation between participants, but also considerable intrapersonal variation in the ways the robot is anthropomorphized. That is, people may respond to the robot as if it was another human in one moment and to its machine-like properties in the next. These findings may influence explanatory models of anthropomorphism.

CCS Concepts: • **Human-centered computing** → **HCI theory, concepts and models**; **Empirical studies in HCI**;

Additional Key Words and Phrases: Anthropomorphism, interaction analysis, methodology, mindless transfer, computers-are-social-actors

ACM Reference format:

Kerstin Fischer. 2021. Tracking Anthropomorphizing Behavior in Human-Robot Interaction. *ACM Trans. Hum.-Robot Interact.* 11, 1, Article 4 (October 2021), 28 pages.
<https://doi.org/10.1145/3442677>

1 INTRODUCTION

Over the past decades, it has been widely documented that people may respond to robots like they respond to other humans. This phenomenon has been discussed under the headings anthropomorphism, mindless transfer or media equation, or as the 'computers are social actors paradigm.' The phenomenon can be understood from at least three different perspectives, and thus an initial terminological clarification is in order:

The research presented here rests to a large extent on a fruitful collaboration with Herbert H. Clark, supported through a research stay at Stanford University funded by the Carlsberg Foundation. I am also much indebted to Jaap Ham and Andreas Matthias for many interesting questions and comments on a previous version of this paper, and to the anonymous reviewers of this paper, who made many very useful suggestions. Remaining errors and inconsistencies are self-evidently my own.

Authors' addresses: K. Fischer, Department of Design and Communication, University of Southern Denmark, Alsion 2, DK-6400 Sønderborg, Denmark; email: kerstin@sdu.dk.

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. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2021 Copyright held by the owner/author(s). Publication rights licensed to ACM.

2573-9522/2021/10-ART4 \$15.00

<https://doi.org/10.1145/3442677>

- (a) First, anthropomorphism is discussed as a particular psychological phenomenon, in which people attribute human-like traits to robots [e.g. Nass & Moon 2000]. For instance, Ruijten [2015: 13] defines anthropomorphism as “an automatic cognitive process of attributing human nature or human uniqueness characteristics to nonhuman technological, spiritual or natural objects and entities, causing a mixture of affective, behavioral and cognitive responses”. Nass and Moon [2000] distinguish between three types of effects: 1) over-use of human social categories, like, for instance, attributing female characteristics to a robot with a female voice; 2) over-learned social behaviors, for instance, applying politeness and reciprocity maxims to interactions with robots, and 3) premature cognitive commitments, for instance, when robots are trusted blindly if labelled ‘specialists.’
- (b) Second, anthropomorphism can be understood as a property of robots, such that they may carry human-like design features [e.g. Rosenthal-von der Pütten 2014]. These properties of robots may be understood as certain ‘triggers’, such as voice, face, emotion, interactivity, engagement with and attention to the user, unpredictability and the filling of traditional roles [see Nass 2004].
- (c) Third, anthropomorphism can refer to the description of people’s behaviors towards robots as similar to responses to human actors.

From a terminological perspective, I therefore distinguish between a) *anthropomorphism*, the socio-cognitive phenomenon to be explained, i.e. the psychological mechanisms that determine people’s behavior towards robots; b) *anthropomorphic design*, which I use here to describe the social cues and ‘triggers’ built in by robot designers [cf. Fink 2012]; and c) *anthropomorphizing behavior*, i.e. the observable behavior by means of which people respond to robots in ways similar to ways in which they respond to humans.

Different methods have been used to shed light on different aspects of anthropomorphism, such as on the relationship between anthropomorphic design and trait attribution; for instance, brain imaging techniques show activations in similar brain regions for human and robotic partners depending on the degree of anthropomorphic design [e.g. Krach et al. 2008]; reaction time studies show that anthropomorphic robots may give rise to similar inhibitions as other humans [e.g. Stenzel et al. 2012; Klapper et al. 2014]; questionnaire studies show that human-like characteristics are attributed to robots [Bartneck et al. 2009] and that robots are subjected to human stereotypes [e.g. Nass & Moon 2000] and to in-group versus out-group biases, for example [e.g. Eyssel & Kuchenbrandt 2012]; and analyses of people’s linguistic behavior show that people describe robots using the same linguistic expressions and intentionalist language as when they describe other people [Fussell et al. 2008] (more on the different approaches below).

At the same time, however, people do not interact with robots exactly like with other people. For instance, Kory Westlund et al. [2016], who tested the effects of framing a robot as machine versus as social actor on child-robot interactions, found that in comparison with interactions with the experimenter, children asked fewer questions and made fewer comments and laughed less in both robot conditions. There is also brain imaging research that reports different responses to human and robotic actors [Chaminade et al. 2010; Kuz et al. 2015]. In addition, not all people anthropomorphize robots to the same extent [Fischer 2011]; for instance, lonely people have been found to anthropomorphize robots more [Lee et al. 2006; Epley et al. 2007]. Ruijten et al. [2019] have also developed a questionnaire that allows them to identify different degrees or types of trait attributions; in particular, they find that traits that are unique to humans are more difficult to attribute than traits that also other beings may have. It is also possible that anthropomorphism changes over time; Lemaignan et al. [2014], for instance, suggested that people may respond with

anthropomorphism to robots in first encounters but come to treat them as more machine-like later [cf. however Turkle et al. 2006; Paetzel et al. 2020 for suggestions of different time courses]. Furthermore, besides anthropomorphism, behaviors not typical of human face-to-face interactions, such as flaming, challenging or swearing, can be observed in human-robot interactions [e.g. Kopp 2006; Bršćić et al. 2015]. Moreover, the amount of anthropomorphism observable has been found to depend to some degree on the method of analysis and on the time participants have available to reflect on their behavior [Fussell et al. 2008]. For instance, the amount of cognitive load a user is experiencing has been proposed to influence anthropomorphism [e.g. Lee et al. 2010; Ham et al. 2012].

While there is robust evidence for certain psychological responses to anthropomorphic design, and while anthropomorphizing behavior is observable across domains and using a plethora of different methods, it is still not clear how the different findings can be brought together. One reason could be that still little is known about what happens dynamically over the course of an interaction (and beyond). So far, anthropomorphism has been studied mostly either post hoc or in highly controlled studies that focus on one isolated behavioral measure; very few studies have looked at the progression of the expression of anthropomorphism over time in interaction. However, if we could describe and track anthropomorphizing behavior in human-robot interactions directly, we may be able to shed light on how the anthropomorphizing behavior observed and the findings on interpersonal variation, anti-social behaviors and effects of time courses come together. Such a method would constitute a useful addition to the methods to study anthropomorphism that are already available. In the remainder of the paper, I therefore present a method for the description of the dynamics of anthropomorphizing behavior in interaction, and illustrate its use in a case study. In the last section, I discuss how analyses of anthropomorphizing responses to robots in interaction may affect explanatory models of anthropomorphism.

2 PREVIOUS WORK: METHODS FOR DESCRIBING ANTHROPOMORPHISM

Anthropomorphizing responses to robots have been demonstrated using a broad range of different methods to analyze the effects of robot appearance, robot behaviors and the framing of robots. The methodological spectrum comprises, most importantly, questionnaires, experimental elicitation of behavioral responses, brain imaging techniques and reaction time studies.

2.1 Questionnaires

Questionnaires have been used ubiquitously in the study of anthropomorphism, for instance, to investigate the effects of robots' appearance and design [e.g. Hegel 2012; Philipps et al. 2017; Reeves & Hancock 2020]. Questionnaires allow the researcher to ask explicitly or implicitly about the effects of certain design decisions regarding robots' appearance or the implementation of social behaviors, such as gaze following, presence or absence of gestures, body postures and the like [e.g. Ham et al. 2012]. The questions asked may concern anthropomorphism directly by addressing topics like the attribution of mind, intention and social characteristics; for instance, Sims et al. [2005] had participants rate images that varied according to robot characteristics (movement method, body shape, presence of limbs, etc.) in a questionnaire addressing the degree of perceived aggression, intelligence, and animacy. Another study using questionnaire data is Rosenthal-von der Pütten & Kramer [2014], in which questionnaire responses to 40 robot pictures were elicited and clustered. The authors found six clusters emerged, depending on how threatening, submissive, likable and unfamiliar the robots were perceived. More recently, Mieczkowski et al. [2019] asked people to rate 342 images of robots concerning warmth and competence, and they elicited behavioral tendencies, such as participants' willingness to help or harm the robot. Several standardized

questionnaires have been developed to investigate anthropomorphism; very common is the God-speed questionnaire developed by Bartneck et al. [2009]. More recent proposals are the Robotic Social Attributes Scale [Carpinella et al. 2017], Ruijten et al.'s [2019] survey that distinguishes between different types of attributions, and Damgaard et al.'s [2020] questionnaire, which distinguishes between anthropomorphing and sociomorphing. Thus, the first use of questionnaires is to understand what features people attribute to robots (and to what extent), based on design properties of the robot or based on the ways in which it is introduced or presented.

A related use of questionnaires is to test for social categorization. For example, Eyssel & Kuchenbrandt [2012] presented a robot to participants as either an in-group or an out-group member and asked them to rate the robot regarding anthropomorphism (mind attribution, warmth), psychological closeness, contact intentions, and design. Participants rated the in-group robot more favorably and anthropomorphized it more strongly.

Questionnaires are also used to measure the effects of certain robot behaviors or properties of behaviors. For instance, Nass and Brave [2005] investigated whether people respond to characteristics of robot voices in similar ways as to human voices.

Another use of questionnaires is to study effects of robot framing on the extent to which people anthropomorphize robots; Paepke & Takayama [2009], for instance, presented participants with a Pleo robot, introducing it either as highly sophisticated or as rather basic. They found that people's experience of the interaction as measured by a post-experimental questionnaire was influenced by the way the robot was framed. A slightly different use of questionnaires is not to ask people for evaluations of the robot, but instead to have them assess the robot's suspected knowledge. Lee et al. [2005], for instance, presented participants with a robot that was introduced to them as either built in China or in the US. Participants furthermore witnessed a brief interaction between the experimenter and the robot in either Mandarin or English, before they had to rate the likelihood that the robot knew landmarks from China, the US and other regions of the world in a questionnaire. The results showed that participants assumed the robot to know more landmarks of the country in which it was built, and thus that participants overused social categories, like the assumption of common ground [Clark 1996].

Besides asking about trait attributions, attributions of group membership and cultural knowledge to robots based on their design, description or behavior, questionnaires may also elicit other kinds of attributions, such as moral capability. For instance, they may ask about the extent to which participants hold robots morally accountable for their actions and the consequences of these actions [e.g. Arnold & Scheutz 2017], which then provides indirect evidence for anthropomorphism.

If questionnaires are used to study anthropomorphic design, they often make use of images or of videos of robots [e.g. Philipps et al. 2017; Rosenthal-von der Pütten & Kramer 2014; Hegel 2012]. If questionnaires are used in human-robot interactions, they are usually administered post hoc, due to the fact that filling them out constitutes an activity by itself. Sometimes questionnaires are administered between different phases of an experiment [e.g. Ham et al. 2012], yet they do not capture people's responses in the here and now of the interaction.

2.2 Behavioral Measures

Another methodology for studying anthropomorphism is to elicit behavioral effects of robot design, robot behavior or robot framing. In comparison with questionnaires, studies targeting behavioral effects have the advantage of studying anthropomorphism in the moment, i.e. during the interactions in which they arise. A considerable body of work rests on previous findings on the effects of social cues in human interaction, which can be replicated in human-robot interaction such that when produced by a robot, social cues are similarly responded to as when such cues are produced by humans. Nass [2010, page 12] describes the method he and his collaborators have

been using [e.g. Reeves & Nass 1996; Nass & Moon 2000; Nass & Brave 2005; among many others] as follows:

“I would find any conclusion by a social science researcher and change the sentence ‘People will do X when interacting with other people’ to ‘People will do X when interacting with a computer.’ I constantly challenged myself to uncover ever more unlikely social rules that applied to technology in defiance of all common sense.”

Examples of observable human behaviors that may serve as indicators for the respective degree of anthropomorphism include their preferences for spatial distance towards a robot. For instance, Ham et al. [2012] measured how closely participants approached the robot to read numbers on a screen mounted on the robot depending on the ostensible approachability of the robot’s body posture and depending on their cognitive load. Other studies investigate the effects of anthropomorphic design or behavior on people’s willingness to lie for the robot [Kahn et al. 2015], on the persuasiveness of a robot [Andrist et al. 2013; 2015], or on compliance with robot requests, such as not to turn it off when the robot is afraid of losing its memory [Seo et al. 2015].

Behavioral studies are generally designed to elicit one measurable response, such as the distance to the robot chosen, secrets kept or revealed, or whether the robot is turned off or not. This single behavior is then taken as an indicator of, for instance, mindless transfer [Nass & Moon 2000]. Sometimes, not only the single behavior itself but also factors that possibly influence the respective response behavior are analyzed in order to identify the conditions under which anthropomorphism occurs, for example, by varying participants’ cognitive load [Ham et al. 2012] or by analyzing user characteristics [e.g. Lee 2010].

An exception from the more narrow focus on one type of response are studies that analyze participants’ verbal behavior in response to social clues from the robot. One variant of this method is to let participants witness human-robot interactions and have them summarize what they have observed in their own words. An example of this method is the study by Fussell et al. [2008, page 145] who gave participants scripts from healthcare interviews that were allegedly carried out either by human or by robot interviewers; the manipulation was supported by presenting participants with images of either a person or a robot. Results show that participants use similar vocabulary to describe the robotic and the human interviewers’ behavior, which in turn is taken as an indicator for anthropomorphism.

Another variant of this method is to have people interact with a robot via speech, and to analyze the degree to which their speech is anthropomorphizing [cf. Fischer et al. 2012]. For instance, Fischer [2006; 2011] finds people to differ in their initial responses to the robot’s greeting, such that their behavior shows more or less anthropomorphism, which correlates with their behavior later in the interactions. Using this approach, Lee et al. [2010] analyzed people’s verbal behavior for indicators of anthropomorphism depending on their initial greeting, and they observed specific behavioral patterns, for instance, concerning politeness, grounding, attention to relevancy and negative behaviors. This method relied on the fact that we always design our utterances in such a way that our intended recipients can understand it; therefore, utterances show traces of who they have been intended for, i.e. they allow ‘reverse recipient design’ [Fischer et al. 2012; Fischer 2016]. Thus, the method also provides indicators for how people understand the robot and the extent to which they anthropomorphize it.

Both types of studies of verbal behavior focus on participants’ overall use of anthropomorphizing language; that is, they analyze the quantitative distribution of features with respect to the total number of words spoken. Consequently, while they rely on the investigation of a broad range of linguistic behaviors, their focus is not on the dynamics of anthropomorphism over the course of

the interaction, nor on its development over time, but on the pervasiveness of anthropomorphism in participants' verbal behavior.

Participants' (mostly) verbal behaviors also form the basis of ethnographic analyses of anthropomorphism [cf. Chun & Knight 2020]. Specifically, participants' verbal behavior observed and documented in field notes or elicited in interviews can be analyzed for traces of anthropomorphism; for instance, Forlizzi [2007] reports instances of social attributions to a Roomba robot. Chun & Knight [2020] additionally asked the participants in their study, employees at a robot manufacturing company, directly for reports or stories about "interesting interactions" with the robots, as well as about the robots' gender and personality, in addition to a survey in which social attributions were elicited in open and Likert-scale questions. These analyses target stakeholders' "holistic understandings of integrated technologies" [Chun & Knight 2020, p. 16:5], i.e. how particular social groups of people, such as families or the different teams in the manufacturing company, make sense of robots depending on their cultural situatedness, and thus aim at understanding the general concepts and practices of groups, rather than at the dynamic unfolding of anthropomorphizing behavior in interaction.

Another type of behavioral study measures how quickly people answer questions about social attributions to robots in comparison with attributions to people [e.g. Fussell et al. 2008] or how quickly they respond when primed with images of either humans or robots (Złotowski et al. [2018]). A variant of this type of investigations is to study the conditions that slow reaction times down, in particular, interference effects, such as the 'Social Simon Effect.' This phenomenon describes the motor interference from observing another person carrying out a slightly different behavior. That is, if a person tries to draw a straight line or hit a button as quickly as possible, she may be more inaccurate or slower when watching another person carrying out a slightly different task [cf. Förster et al. 2020]. Stenzel et al. [2012] investigated to what extent robots are perceived as similar enough to cause such an interference. In these experiments, participants were seated next to a quite mechanically looking, humanoid robot, where they had to hit the left key when a square appeared on the screen, whereas the robot hit the right key when a diamond appeared. Participants were told either that the robot was active and intelligent and responding autonomously, or that it was completely controlled by the computer, and thus mechanically executing motor commands. The results showed that only when participants believed that the robot was active and intentional, their own behavior was influenced by the robot's behavior. In contrast, if they believed the robot to be an unintelligent machine, no interference was observed.

Similarly, Klapper et al. [2014] investigated the degree to which anthropomorphic design and beliefs about the robot interferes with participants' own finger movements. They found that human form and the belief to be observing a human hand together created inhibition effects, but not if only one of them is present. The two reaction time studies thus suggested that while automatic responses to anthropomorphic design can be observed, these responses are mediated by participants' mental models.

To sum up, behavioral measures provide indirect evidence for the extent to which people anthropomorphize a robotic interaction partner. Usually the baseline is the human communication partner; for instance, if people's reaction times, degrees of interference, ways of speaking or responses to a particular cue (like the request to lie for the robot) are similar to those observable in interactions with humans, then these behaviors are taken as evidence for anthropomorphism. This approach produces information about anthropomorphism in the moment, but studies usually measure only a single behavior and only once. This approach thus cannot inform us about the time course of anthropomorphism in interaction dynamically because not enough data points are elicited.

2.3 Brain Imaging

Brain imaging techniques have been used to investigate the extent to which people respond to robots like to other people. For instance, Krach et al. [2008] had people play a computer game and made them believe that they were playing against a human partner, a humanoid robot, a functional robot and a laptop computer. In fact, participants interacted with the same computer program in all four cases. However, the participants not only rated their interaction partners as increasingly more fun, intelligent, competitive and sympathetic with increasing degrees of human-likeness, but their brain activations in interactions with the artificial interaction partners were also increasingly similar to those in interactions with the human partner the more human-like the communication partner was expected to be. Kuz et al. [2015] compared people's recognition of movements by two robotic agents, one mechanical, the other anthropomorphic, and found that while the actions by both agents were generally recognized, in contrast to Krach et al.'s [2008] results, they did give rise to different brain patterns. In a review paper on people's responses to embodied artificial agents in general, Wykowska et al. [2016] concluded that responses to such agents are so similar to responses to humans that they can even be used as testbeds to study human social cognition.

To sum up, while previous work has demonstrated anthropomorphism in a large variety of situations and has shed much light on the factors conditioning anthropomorphizing behavior and the attribution of human-like properties into robots, the methods used in previous work are not designed to trace anthropomorphism over the course of an interaction. To have such a method to complement the set of methods already available would, however, be very useful to address inter- and intrapersonal variation in response to robots designed as social actors, to reconcile findings about anthropomorphism with findings on anti-social behaviors towards robots, and to understand the relationship between the automatic, mindless characteristics of anthropomorphism on the one hand and its cognitive mediatedness on the other. In the following, I thus suggest a method to describe the dynamics of anthropomorphism in interaction.

3 A METHOD FOR DESCRIBING ANTHROPOMORPHIZING BEHAVIOR IN INTERACTION

Anthropomorphism concerns the fact that an artefact is treated as if it had human-like characteristics even though it does not possess these properties. In this section, I suggest a method to track anthropomorphizing behavior in interaction over time.

3.1 Tracking Anthropomorphizing Behavior in Interaction

In order to describe how people anthropomorphize robots over the course of an interaction, we need a model that can track whether and, if so, in which ways or to what extent, people respond to a robot in a human-like way. Such a procedure would have to rely on people's *behavior*, since their beliefs and attitudes are not directly available, and stopping them to fill out a questionnaire would disrupt the very interaction the analysis is setting out to explain. Similarly, we cannot rely on robots' anthropomorphic design or behavior since it is the responses to these designs and behaviors that are at issue. Thus, any method that aims to investigate anthropomorphism in interaction has to rely on the observation of people's behaviors, and not just on single behaviors, such as whether or not a secret is kept [Kahn et al. 2015] or whether or not people agree to the robot's memory being erased [Seo et al. 2015], but potentially on all behaviors observable in interaction. That is, since anthropomorphism may express itself in mimics, gesture, polite verbal behavior and in many other ways, we would want to account for anthropomorphic responses with respect to all of these modalities, not just single actions. A method to track anthropomorphizing behavior in interaction over time thus has to rely on observing and classifying a broad range of people's behaviors towards robots in interaction.

Now, a minimalistic classification would mean to classify people's behavior as either anthropomorphizing or not, whereas a maximalist version of such a classification might classify the type or extent of anthropomorphization. One might, for instance, identify, based on each of a person's observable behaviors, what kind of trait is attributed, how uniquely human it is and correspondingly, how easy or difficult it is to attribute this trait to the artifact in question [cf. Ruijten et al. 2019]. This would however mean that specific attributions can be identified and intersubjectively classified, given that all we have is how people behave. The question is whether such a classification is feasible and how the relevant judgements can be made based on what is observable; it seems therefore safer to (a) keep the number of distinctions small, and (b) to allow an easy identification of the level oriented to.

The proposal made here is to use a model of depiction [Clark 2016] because it allows us to describe people's responses to robots without presupposing that we know what this behavior is caused by. For instance, Nass & Moon [2000] suggest that people respond to technological artifacts like robots like to other humans out of mindlessness; Nass [2004] argues that anthropomorphizing behavior is basically a mistake, based on our human evolution in which we were only faced with human social actors. In contrast, Ruijten et al. [2019], for instance, regard anthropomorphism to be a predisposition instead, whereas social-constructivist approaches may take it as an instance of social practice, so that anthropomorphizing behavior would just be 'how things get done' [cf. Edwards 1994; Hutchby 2001]. In order to avoid circularity, a model that describes observable behaviors should not presuppose the one or other explanation and should thus be largely theory neutral. Therefore, I analyze the robot's behavior based on Clark's three levels of depiction [Clark 2016], and the person's behavior in terms of orientation to these levels. From this perspective (Clark & Fischer under revision), the robot's design is viewed as an act of staging, in which the robot is depicted as a social actor.

We understand by depiction a very common strategy of human interaction in which one person stages a behavior for another person to evoke a temporally and spatially removed situation. This strategy is most obvious in plays and other kinds of fictional staging (in movies, video games, pretend play etc.), but also occurs frequently in conversation in general when people report on events involving other people, for instance, a person telling her friend about her tyrannical boss. Clark's model involves (a) a base scene, which consists of the concrete means by which the depiction is carried out, e.g. gesture, mimics, loud voice etc., i.e. the respective 'trigger', (b) the proximal scene, which comprises the scene depicted, e.g. an angry person with a loud voice and exaggerated gestures, and (c) the distal scene, for instance, the speaker's boss getting outraged about a mistake someone made. Robots are special kinds of depictions since the scene evoked is not necessarily spatially or temporally removed, but rather concerns a fictitious being (cf. Clark & Fischer under revision for more details on the full model).

Concerning human-robot interaction, the base scene is relatively straight-forward to define: it concerns the mechanical properties of a robot as a machine. Less straight-forward is the definition of the proximal scene, which can be understood as the encoded meaning of a robot's behavior. That is, the relationship between base and proximal scene can be understood as one of signaling [Peirce 1998], such that the base scene (e.g. a robot moving its actuators) may evoke a particular meaning (e.g. the robot waving) by means of indexical, iconic or symbolic relations between the movement and the interpretation. An index is a sign in which the form consists of a pointer to a particular referent; an example is the English word "I," which indexes the speaker, but it does not 'mean' John, Mary, Herb or Kerstin itself. For instance, directing the robot head towards a participant's position *indicates* attention towards him or her. An icon is a sign where the form is connected to its meaning by means of similarity; for instance, the robot moving its gripper from left to right and back is similar to a human waving gesture, i.e. it can be taken to *iconically* represent waving. Finally, a

symbol is a sign with an arbitrary, conventional relationship between form and interpretation, the prime example being language, where the sequence of the letters c-a-t are conventionally taken to mean ‘cat’. When a robot says “hello,” it conventionally *symbolizes* a greeting.

In contrast, the distal scene goes beyond the encoded meaning and concerns the pragmatic interpretation of an action. The definition of the distal scene is thus also relatively unproblematic since it concerns the completely fictional, live-like character evoked by the robot, for instance, as a companion, friend or collaborator. For example, the robot moving its hand encodes a waving gesture, but the interpretation that it sees me, recognizes me, wants to attract my attention and initiate an interaction by waving at me is an interpretation that goes beyond what is encoded. Both the proximal and the distal scene thus involve anthropomorphizing behavior; however, while the distal scene evokes the robot as a character, i.e. as a social actor, the proximal scene only assumes that the robot is engaged in meaningful action, i.e. that its behavior *means something*.

The two levels of depictions assumed thus differ in the underlying processes: whereas the proximal scene relies on semiotic processes, i.e. on the interpretation of signs, the distal scene is essentially fictional and relies on imagination [e.g. Walton 1993] and thus involves a certain degree of pretending on the side of the recipient [Clark 1999]. The distinction is useful for our analysis because it helps distinguish between features that are encoded in the (behavior) design of the robot and features that go beyond this; for instance, using a voice with features characteristic of a human female, for example, indexes a female speaker; that is, there is an objective, indexical relationship between certain voice characteristics and human gender. In contrast, if people associate gender stereotypes to a (human or robotic) speaker with these voice characteristics, then these stereotypes are not encoded in, but associated with the voice [e.g. Tay et al. 2014; Bryant et al. 2020; Law et al. 2020]. That is, by means of its speech characteristics, the robot is depicted as female, but an expected higher competence in a certain occupation is not part of the robot’s (behavior) design, but rather “in the eye of the beholder”, i.e. in the distal scene evoked.

In our model (Clark & Fischer under revision), a robot is thus similar to Mickey Mouse in Disney Land: The man in the costume constitutes the basis for the depiction; the mouse-shaped costume evokes the depiction proper, a large mouse with eyes, ears and a mouth. The distal scene is then that Mickey Mouse, a character one is familiar with from numerous comics, is present in the here and now. Regarding Mickey Mouse, kids can orient to the man in the costume (e.g. “is it hot in there?”), to the depiction of the mouse (e.g. “you have a large head!”) or to the character (e.g. “Mickey, where is Minnie?”). To which of the three scenes they orient can thus be identified from their behavior.

Similarly, the robot has a mechanical base that enables the robot’s functionality; the robot’s design, functionality and presentation in context suggest that it can be interacted with by talking to it. And finally, the staging of the robot evokes a distal scene in which the robot can be understood as a human-like character. Like kids’ interactions with Mickey Mouse, people’s behavior in response to a robot’s actions is revealing regarding to which of the three scenes they orient.

Table 1 illustrates how the depiction model applies to various different kinds of depictions [see also Clark 2016].

Thus, the depiction model assumes that depictions allow attention to three different perspectives or scenes, which exist in parallel, and all three can be in focus at different points in time. For instance, concerning the conversational depiction of the angry boss, the listener can focus on the base scene by saying ‘not so loud’, on the proximal scene by saying ‘you sound more cynical than angry’, or on the distal scene by saying ‘your boss is really not a nice person’.

Thus, what we are proposing is the anatomy of a ‘social cue’, where on the side of the robot, we understand a social cue as the staging of a design choice or behavior that is intended to signal a certain human-like behavior, which is meant to evoke a particular character or fictional being (for the explanatory power of this model, see Clark & Fischer under revision). In a second step, I

Table 1. Examples of Depictions

	base	proximal	distal
conversational depiction	speaker's voice and body	someone with loud, agitated voice	the speaker's tyrannical boss
puppet show	wooden puppet, speaker's voice and movement	puppet moving about and talking	character
Hamlet (on stage)	Kenneth Branagh	young man in conflict	prince in 16 th century Denmark
Mickey Mouse	man in costume	human-size mouse	Mickey Mouse character
human-robot interaction	robot moves gripper left and right	robot waves its hand	robot greets participant

Table 2. Examples of Human-Robot Interactions as Depictions

	base	proximal	distal
Nakata et al. [1998]	the robot lifts its 'head' when sensing human touch	repelling tactile reaction	rebellious
Tielmann et al. [2014]	robot speaks with loud, high pitched voice with higher speech rate	arousal	the robot is as excited about something as the child
Huang & Mutlu [2013]	robot produces recorded gesture	robot makes an iconic hand gesture	robot is competent

employ qualitative methods to analyze to which of these three levels people's behavior is oriented (see Section 3.3 below).

3.2 Describing Robot Behavior

For the description of robots' behavior, we can use the depiction model to account for the discrepancy between their mechanical nature as artefacts, the behaviors designed and the intended effects. In particular, the base scene corresponds to the fact that robots are artefacts that operate mechanically or are preprogrammed or remotely controlled and describes the robot's actions on the mechanical level, such as moving its joints. Responses to the base scene are thus clearly non-anthropomorphic. The proximal scene corresponds to the recognition of this scene as a purposeful action by the robot, such as lifting its arm. Responses to this scene are anthropomorphic since they rely on the attribution of certain properties that the robot does not have, but which are signaled by the design or behavior of the robot, such as having eyes or looking at something. The distal scene finally accounts for the possible function of the respective behavior in context, which may in this case be that the robot is trying to reach for something. It concerns a fictional scene in which the robot is a character, for instance, a human-like being.

Let us consider some examples from human-robot interaction (see Table 2); for instance, Nakata et al. [1998] introduced subjects to a fish-shaped experimental robot which is sensitive to touch. In one of the three behaviors implemented, the robot lifts its 'head' when sensing human touch. The authors define this behavior as "a repelling tactile reaction" and label it "rebellious." Using the depiction model, we can distinguish between the base scene, in which the robot's motors push up the upper part of the robot when the input to the touch sensor exceeds a threshold; the proximal scene, in which the robot 'pushes back;' and the distal scene, in which the robot's behavior is understood as expressing a rebellious intention.

As another example consider Tielmann et al.'s [2014] model of robot emotional expression in child-robot interaction, in which the robot's behavior differs concerning features like speech volume, pitch height, speech rate, head pose, eye color, arm and trunk pose. The robot is designed to match the child's measured emotional state, which is analyzed in terms of arousal, valence and

level of extroversion. The authors describe the relationship as follows: “The voice of the robot will be influenced by its arousal. The higher arousal, the louder the robot will speak, the higher pitched its voice will be and the higher the speech rate. Speech volume is also influenced by extroversion, the higher the extroversion, the louder the voice.” Thus, according to our model, the base scene is constituted by a machine producing speech in a loud and high-pitched voice, the proximal scene is to signal arousal, and the distal scene is to indicate to the child that the robot is as excited about something as the child is.

As a last example, consider the implementation of different types of gestures in a narrating robot [Huang & Mutlu 2013]. The robot’s gestures were carefully modeled after human gestures previously elicited in human narrations and coordinated with particular lexical material in the robot’s speech stream. The authors find, for instance, that the “robot’s use of iconic gestures significantly predicted males’ perceptions of the robot’s competence”. Using our model, we can distinguish between the robot producing a previously demonstrated and recorded posture of its ‘hand’ at the same time as a particular word is used, which constitutes the base scene; the interpretation of the hand posture as iconic gesture, which constitutes the proximal scene, and the impression of a competent robot that can accompany its words with appropriate gestures that indicate that it understands what it is saying, which constitutes the distal scene.

By conceptualizing the robot’s behavior as an act of depiction, we can analyze in the next step what level of the robot’s behavior users rely on while interacting with the robot, moment by moment. We thus apply the depiction model for describing the robot’s behavior, and in a second step use it to describe the ways in which people respond to the robot’s actions.

3.3 Describing Human Behavior in Interaction with Robots

For the analysis of participants’ responses to the robot in interaction, we take the micro-analytic perspective on participants’ actions developed in the framework of ethnomethodological conversation analysis [cf. Garfinkel 1972; Sacks 1992]. In this methodological framework [cf. Sacks et al. 1974], like in other collaborative models of communication [such as Clark 1996; Bangerter & Mayor 2013], it is assumed that in order for interaction to succeed, participants need to display to each other continuously how they understand each others’ contributions and what they consider the relevant context of the current action to be [cf. Clark & Schaefer 1989]. For instance, if speaker A says “sit down”, this utterance *per se* is highly ambiguous; it could be a suggestion, an invitation, or a command, among many other things. However, once speaker B has responded to it, for instance by saying “thank you”, the interpretation of A’s utterance is specified as an invitation. In the next turn, speaker A may reject this interpretation, for instance, by saying “this was an order!”; however, if she does not challenge speaker B’s interpretation, ‘invitation’ is the intersubjectively established interpretation of A’s utterance [cf. Sacks et al. 1974]. Because speakers in interaction need to display to each other the interpretations of each others’ utterances in order to arrive at intersubjectively available common ground [cf. Clark & Schaefer 1989], their interpretations of each others’ behaviors are also available to the observer [see Sacks 1984].

Correspondingly, in the method proposed, we look at the ways in which the participants display how they make sense of the robot’s behavior.¹ Our micro-analysis of the users’ responses to the robot’s depiction of human social behavior can indicate to which scene the user attends in the sense

¹Obviously, in most current human-robot interactions, robots are not designed to display their interpretation of a user’s utterance, especially not if the dialog is scripted or even deterministic. Nevertheless, even the non-verbal compliance with the request to move forward constitutes such a display of the interpretation of the user’s behavior – to the user. That is, if the robot complies, the user will take her instruction to be understood and processed successfully, and she will move on to the next instruction, which she will design based on what she assumes has previously been successful, i.e. what she takes to be common ground (for details and further discussion, please see Fischer [2016]).

that the human is displaying an orientation to a certain aspect of the situation; for instance, in the example above, by saying “thank you,” speaker B displays her orientation to speaker A’s utterance as an invitation. While in this example, the display concerns the interpretation of the function of the utterance, people also design their behaviors specifically for their current interaction partners. That is, each utterance is so designed that the particular recipient can understand it [Levinson 2006]; who the speaker takes the recipient of the utterance to be, for example, whether he or she assumes to be talking to a machine or to a live-like character, can thus also be identified from the way the utterance is designed [i.e. by means of ‘reverse recipient design’, cf. Fischer et al. 2012]. Using this approach, we can analyze participants’ behaviors towards robots by classifying them according to (a) whether they display an orientation to the base scene by treating the robot as a mechanical tool; (b) whether they display an orientation to the proximal scene by displaying an understanding of the robot’s behavior as being intentional and meaningful; or (c) whether they attend to the distal scene by displaying an understanding of the robot’s behavior as human-like.

In particular, for each response to a robot’s actions, we identify to which of the three scenes proposed it is a response to. For instance, a robot’s greeting “Hello, how are you?”, can be responded to in several different ways [see Fischer 2011]: One possibility is to reciprocate the greeting by replying “I’m fine, thanks, how are you?”. This is a response at the distal level, where two people are having a polite conversation. A second possibility is to respond to the robot’s utterance as the production of canned speech by a machine in order to indicate the beginning of the interaction. In this case, people may say “drive forward”. This utterance signals an orientation towards the base scene. Furthermore, people may acknowledge that the robot was asking a question and producing a greeting by responding “fine” or “hello”. In this case, people orient towards the proximal scene, in which the robot is taken to produce some type of social action. However, the human interlocutor does not enter into the joint pretense [cf. Clark 1999] of having a polite conversation, but acknowledges that the robot’s behavior may have a social function, such as opening the interaction (‘hello’) and asking a question (‘fine’). These responses would therefore be located at the proximal level.

We can thus classify people’s behavior as being oriented to the base, proximal or distal scene, where responses to the proximal and distal scene are anthropomorphizing since they involve sense-making by drawing on the human domain. While the distal scene clearly evokes a fictional, human-like being, the proximal scene also involves human-like categories, such as the attribution of human functions (e.g. the robot “sees” something), and can be (and has been) considered as anthropomorphism [e.g. Fussell et al. 2008], yet the mechanisms underlying these two scenes are most likely quite different (that is, semiotic versus inferential processes, [cf. Fischer 2016]), and thus we rather keep them distinct in our classification.

Given the multimodal nature of interaction, users’ displays of how they understand the partner’s actions may potentially comprise a very broad range of signals. For instance, on the one hand, we can look at what people say, on the other, also the timely response to the partner’s behavior, i.e. attending to the response time typical of human interaction, about 300msecs, [see Jefferson 2004], can indicate to the analyst that the person is orienting to the robot in similar ways as to a human communication partner [cf. Hutchby 2001]. Similarly, people may or may not gaze towards robots at times when they would also gaze towards human interaction partners [e.g. Andrist et al. 2014]. Anthropomorphizing responses may consequently become evident in all kinds of communicative behaviors.

The method proposed thus targets anthropomorphism in terms of anthropomorphizing behavior, i.e. behavior that attributes human-like qualities to a robot. It is consequently in line with work investigating behavioral responses, such as keeping a robot’s secret [e.g. Kahn et al. 2015], saving a computer’s face [Nass 2004] or describing a robot in intentionalistic ways [Fussell et al. 2008]. However, the method proposed does not make claims regarding people’s subjective

understandings of robots, i.e. the psychological basis for the observable behavior. For instance, the method proposed does not make assumptions about psychological mechanisms such as mindless transfer [Nass & Moon 2000], automatic trait attribution [Roebroek 2014], or the real belief that the robot under consideration has a set of human-like characteristics. That is, our study is neutral with respect to possible explanations for the anthropomorphizing behavior observable. Nevertheless, as we shall see (Section 6), the results of tracking people's anthropomorphizing behaviors in interaction over time will have considerable consequences for explanatory models of anthropomorphism in human-robot interaction. Bridging the gap between people's cognitive processes and their observable behavior goes unfortunately far beyond the limits of this paper (but see Clark & Fischer under revision). What the method proposed here however does allow us to do is to describe people's responses to robotic action and track anthropomorphizing behavior dynamically over the course of an interaction.

4 CASE STUDY

In the following, I illustrate the method proposed in an example analysis of human-robot interactions based on Clark's [2016] model of depiction; that is, I illustrate the application of the methodology on a case.

4.1 Data

The data re-used here are spontaneous interactions gathered in the lobby of Willow Garage for another study [Fischer et al. 2014]. In these interactions, participants were left alone in the lobby to fill out an initial questionnaire. While participants were busy filling out the form, a PR2 robot, placed directly opposite the participants across the counter on which they were writing, addressed them either with speech or with a beep, and performed a series of gestures intended to be understood as gesturing for help.

The data set comprises 19 interactions (nine men, nine women, one unreported gender) of participants between 19 and 63 years old and from various occupations and educational backgrounds. They had been recruited to participate in human-robot interaction experiments but did not know at the time of the interaction that it was part of the experiment (but were debriefed later). They were on average not very acquainted with robots (3.4 on a 7-item scale), and they scored $M = 82$ out of 110 on other-oriented empathy ($sd = 7.3$) and $M = 25.5$ out of 40 ($sd = 4.3$) for helpfulness [cf. Fischer et al. 2014] in a standardized questionnaire on prosocial personality [Penner 2002].

The robot's behaviors were all prerecorded and were initiated by the experimenter remotely from another room according to a fixed script. Robot behavior was not fully automatized in order to ensure participants' safety. The experimenter came back into the lobby when the participants had stopped interacting with the robot for at least 30 seconds.

4.2 Analysis of the Robot's Behavior

In Table 3, we analyze the robot's behavior in the speech condition on the basis of our depiction model; the robot behaviors were implemented by the engineer using the so-called 'puppet-interface' [cf. Pantofaru et al. 2013]. This interface allows the researcher to program different robot movements and to store them in sequence. These actions constitute the base scene, which comprises the concrete means by which the depiction is carried out, such as gesture, head orientation or presynthesized, canned speech. This behavior is meant to evoke a proximal scene, such as the robot reaching for something or turning towards the participant. The third level is the distal scene evoked, such as the robot intending to collect a plastic cup from the shelf or turning to the participant for help. The distal scene evoked thus corresponds to the robot behavior designer's intentions (see Table 3). That is, from a social interaction perspective (i.e. from the perspective of

Table 3. Analysis of the Robot's Behavior in Terms of Depiction

base	proximal	distal
turns upper part 60 degrees	turns head towards participant	attends to participant
plays canned speech	initiates contact, produces human speech	greet participant, opens conversation
plays canned speech	polite utterance	polite welcome
keeps upper part turned	extends gaze	continued attention
plays beep series	typical robot behavior	aims to get participant's attention
turns upper part	moves head towards shelf	attends ostensively to shelf
lifts and lowers arm	reaches for object	ostensively attempts to reach for object on shelf
lowers upper part	looks down	indicates sadness
moves upper part to previous position	looks towards participant	asks participant for help

the distal scene evoked), the robot first attempts to attract the participant's attention, then gestures towards an object on the shelf that is out of its reach, and finally looks back and forth between the object and the participant in order to get help.

Eight participants saw the robot in the speech condition, while eleven saw it in the beep condition, in which the robot's greeting was replaced by a beep series; thus, in the second condition, the robot leaves the polite welcome message away and instead simply uses a series of beeps to attract the participant's attention; in addition, it looks back and forth between the desired object and the participant one more time than in the speech condition.

4.3 Analysis of the Interactions

The interactions were analyzed concerning the scene attended-to in response to each of the robot's behaviors. That is, we investigated whether each participant displays an orientation to the base scene, the proximal scene or the distal scene and encode each response as one of the three. In particular, we code a user's behavior as attention to the base scene (see Table 3) if there is no reaction to the robot's behavior at all (because in interactions between humans, since interactants need to signal to each other how they understand each others' behaviors, lack of response to a partner's action is rare),² or when the participants draw attention to the material properties of the robot, for example, by pointing out that it cannot do anything, or by walking around it, reading posters behind it, etc.

In contrast, we code as attention to the proximal scene when people attend to the robot (e.g. by watching it) and thus display an understanding of the robot being in the course of carrying out an extended activity, i.e. when they observe it for longer periods of time. Another example of attention to the proximal scene is when participants assume that the robot is doing something meaningful, but do not attribute emotions or its own agenda into it, for instance, when the robot is turning to the shelf to get help while participants wave at it to make it perceive them and thus display their interpretation of the robot's movement as an attempt to scan the room.

²As one anonymous reviewer pointed out, there may be intercultural differences concerning what behaviors are conventionally expected and also how they are interpreted across cultures, and there may be cultures in which it is acceptable or even normatively expected not to respond to another person's verbal statements, possibly also depending on social roles or social status. This is of course possible and would have to be taken into account in the analysis. That is, when analyzing what understandings people display of the robot's behavior, we have to base our analysis on our knowledge of the social meanings of the cultural community in which the interaction occurs. The method proposed relies on people's own sense-making processes, which are always shaped by socio-cultural norms and expectations. The current analysis relies on the conventional practices uncovered in the conversation analytic literature [e.g. Sacks et al. 1974] for the same cultural context in which the data were elicited, namely the U.S. [cf. Fischer et al. 2014].

Finally, we code as attention to the distal scene when participants indicate that they expect the robot to have its own agenda (for instance, by asking what it wants), when they suspect human wishes, such as hoping to get a glass of water, or when they respond politely and reciprocate social behavior [Nass & Moon 2000].

Below is an informal rendering of an example interaction; because of spatial constraints only one analysis is presented in detail, and to make it as random as possible, the analysis presented concerns the first interaction elicited (participant H101)³ (see Table 4). The participant's behavior can now be analyzed according to which scene it is oriented to.

The robot's first action is to move its upper part, which was intended to indicate that the robot sees the participant and thus turns its head towards him or her. The participant does not notice this action, however. While we cannot be certain that participants would have noticed such a move from a human in the same position, the robot is still treated like any other piece of furniture in the room by not being attended to (i.e. at base level).

The robot's next action is to play canned speech, namely a polite greeting. During the greeting, the participant briefly looks up and down again. After a pause of about two seconds, the participant responds to the robot's greeting by reciprocating it, thus attending to the distal scene in which two equals participate in a polite exchange. At the same time, he does not look at the robot while it is speaking (which is unlike communication among humans, since listeners usually look at speakers while they are speaking [cf., for instance, Andrist et al. 2014]), and also the long pause before the greeting would indicate trouble in human interaction [e.g. Levinson 1983] – the usual response time in conversation is about 300 to 500 milliseconds [Jefferson 2004], with some slight intercultural variation [cf. Stivers et al. 2009]. Thus, while the participant's greeting itself is oriented towards the distal scene, other interactional behaviors are rather oriented at a communication partner that does not require attention to the social norms of human conversation (and thus relate more to the base scene). To sum up, the participant's looking at the robot during the greeting is encoded as attention to the proximal scene, the long pause is understood as oriented to the base scene, while the reciprocal greeting is coded as orientation to the distal scene.

In response to the robot's next utterance, the participant does not react at all, thus withholding feedback necessary for the joint management of interaction [cf. Sacks et al. 1974] by not providing evidence of his understanding of the partner's utterance. This behavior is consequently coded as attention to the base scene.

The robot produces a series of beeps then, which the participant repeats after he finishes his questionnaire; he thus displays that he recognizes it as attention-getting, but treats it as non-social (compare this to the ways R2D2's beeps are responded to in Star Wars). We therefore suggest this behavior to respond to the proximal scene: a robot signaling to get attention.

The robot's turning to the shelf is then interpreted as 'looking,' as can be seen from the waving gesture and the question "do you see me?", i.e. it is treated as a very basic, but directed behavior; thus, the participant responds to the proximal scene. Similarly, he does not understand the robot's gesture towards the shelf as based on its own agenda; instead, he provides the robot with a reason to move the arm towards the shelf by giving an instruction ("pick a book"). This means that he does not attribute intention or volition to the robot, which corresponds to attention to the proximal scene.

However, he then offers the robot, not just the cup, but "a glass of water," i.e. addressing a human need, hence responding to the distal scene. In the next step, the participant exaggerates the action, squeezing the cup into the robot's gripper and indicating the futility of the operation. Thus, he responds first to the proximal and then to the base scene. Finally, he walks away, indicating no

³Transcription conventions: "word" = verbal utterance; <par> word </par> = parallel speech; <P> = pause.

Table 4. Interaction between Participant H101 and the PR2

<p>H101: is writing.</p> <p>Pr2: turns head towards participant.</p> <p>Pr2: "Hello, my name <par> is Pr2. </par> How are you?"</p> <p>H101: <par> looks up – looks down </par></p> <p>H101: <P> "I'm okay. How are you?"</p>	<p>(base)</p> <p>proximal</p> <p>base</p> <p>distal</p>
<p>Pr2: "It's great to meet you. Enjoy your visit at Willow Garage."</p> <p>H101: keeps filling out form.</p> <p>Pr2: "beep beep – beep beep"</p> <p>H101: drops pen, looks at robot. "beep beep – beep beep."</p>	<p>base</p> <p>proximal</p>
<p>Pr2: turns head towards shelf.</p> <p>H101: "Do you see me?" Waves at robot. "Do you answer my question?"</p>	<p>proximal</p>
<p>Pr2: lifts <par> and lowers arm, lifts and lowers and lifts arm </par></p> <p>H101: <par> "pick a book. pick a book. pick a book." </par>.</p> <p>"Do you want a glass of water?" Walks over to robot.</p>	<p>proximal</p> <p>distal</p>
<p>Pr2: looks down</p> <p>H101: "can you pick this up? Here." Squeezes cup into the robot's hand.</p>	<p>proximal</p>
<p>Pr2: looks at participant's original position.</p> <p>H101: "what are you gonna do with that?" Turns away. "Do something." Walks away without another look.</p>	<p>proximal</p> <p>base</p>

Fig. 1. Progression of H101's responses to the robot's actions (y-axis) in terms of orientation to the base, proximal and distal level.

social obligation to negotiate the end of the interaction [cf. Schegloff & Sacks 1973 in contrast], thus orienting towards the base scene again.

The participant's behavior is visualized in Figure 1, which illustrates how his behavior changes between the base, proximal and distal scenes (y-axis) in response to the different robot behaviors (x-axis). A line diagram was chosen to highlight the progression of anthropomorphizing behavior over time (even though there are no transitions between the three levels as the connections might suggest), because the novelty of the method proposed here is that it can document the dynamic unfolding of anthropomorphizing behavior over the course of the interaction.

As the diagram summarizes, the participant switches back and forth between attention to the distal, proximal and base scene within this one minute of interaction (59 seconds).

4.4 Analysis of the whole dataset

All 19 interactions were analyzed based on the depiction model in the same way as described above. In order to measure the interrater reliability of the coding scheme presented, four interactions were analyzed by an independent coder, a master student in an IT program on web communication design, on the basis of the text in section 3 and the example analyses in section 4, which served as his only instruction. The comparison of the analyses by the first author and the independent coder reveals 93.74% agreement and a Cohen's kappa of 0.874.

Data analysis shows that every single interaction is unique in terms of attending to the robot, following the robot's attention to the shelf, moving towards the robot, and helping.

For example, participant H111 behaved very differently from participant H101 by first ignoring the robot (base scene – the robot is a machine that needs no attention) and then by trying to make sense of the robot's actions (proximal scene – the robot's actions indicate something) (see Table 5).

Another example showing another type of response is produced by H112, who responds to the robot in social ways initially, but then largely ignores it (see Table 6):

Table 5. Participant H111

Interaction	Code	Explanation
H111: filling out form Pr2: turns head towards participant. Pr2: ‘‘Hello, my name is Pr2. How are you?’’ (4secs) Pr2: ‘‘It’s great to meet you.’’	(base) base proximal	no response no response understands robot to be in the process of doing something
H111: looks at robot Pr2: Enjoy your visit at Willow Garage. Pr2: ‘‘beep beep – beep beep’’ H111: looks down at form Pr2: turns head towards shelf. 111: looks at robot. Pr2: lifts and lowers arm, lifts and lowers arm	 base proximal	 no response understands robot to be in the process of doing something
H111: looks down at form Pr2: lifts arm Pr2: looks down Pr2: looks at participant. (8secs) H111: pushes away form, looks at robot and moves towards shelf Pr2: turns head towards shelf. H111: looks at robot, looks at shelf, looks at robot. looks at shelf, steps towards shelf, looks at robot, looks at cup/shelf, inspects robot Pr2: turns head towards participant’s original position H111: steps back a little, looking at robot	 proximal proximal proximal	 no response / prioritizes own work understands robot to attend to the shelf understands robot to be in the process of doing something
Pr2: turns head towards shelf. H111: keeps looking at robot, then moves around robot and inspects back of robot Pr2: turns head towards participant’s original position Pr2: turns head towards shelf. H111: keeps looking at robot, then follows robot’s extended arm, looking at shelf, then looks down looks at shelf, then steps back, looking at lowest shelf. bends down towards lowest shelf, then keeps inspecting shelf. looks at robot, then down at floor, back at robot, looks at watch, back to robot, towards shelf, then around, then walks around and away from the robot	 proximal base proximal base	 treats robot as artifact understands the robot to index something on the shelf treats robot as one artifact out of others (like the shelf)

Table 6. Participant H112

H112: filling out form Pr2: turns head towards participant. Pr2: ‘‘Hello, my name is Pr2. How are you?’’ (6secs)	(base)	no response
H112: looks at robot, smiles, looks down at form Pr2: ‘‘It’s great to meet you. Enjoy your visit at Willow Garage.’’	base distal	delayed response understands robot as producing a friendly greeting
H112: nods, looks up at robot, waves at robot Pr2: ‘‘beep beep – beep beep’’ Pr2: turns head towards shelf.	distal base proximal	understands robot to be in the process of doing something directed at the shelf
H112: smiles, looks at shelf Pr2: lifts and lowers arm, lifts and lowers and lifts arm Pr2: looks down Pr2: looks at participant.	proximal proximal	interprets robot’s head movement as directed towards her, but without intention
H112: smiles, waves at robot Pr2: turns head towards shelf. Pr2: looks at participant. Pr2: turns head towards shelf. H112: looks towards shelf, then back to robot Pr2: looks at participant. Pr2: turns head towards shelf. (31secs)	proximal	understands that the robot’s behavior is directed
H112: looks away, looks around, glances briefly at robot twice, then looks down at form.	base	treats robot as one aspect of context

The variability in participants’ behaviors towards the robot is reflected in Figure 2, which represents the eight participants’ responses to the robot’s nine different behaviors (on the x-axis) in the speech condition. As Figure 2 illustrates, the eight participants in the speech condition employ a wide range of behaviors in unique combinations in interaction with the robot. That is, no robot behavior is responded to in the same way by all participants. It shows that most, but not all, participants respond to the distal scene evoked when the robot produces a polite greeting, and again many but not all participants attend to the distal scene after the robot’s gesturing for help. However, as Figure 2 additionally illustrates, all participants also produce behaviors in between that indicate their interpretation of the robot as machine-like, and in general the time courses of their responses to the base, proximal and distal scene vary considerably. The analysis shows that no direct correspondence between robot behavior and participant responses is emerging, so that participants’ actions are not ‘triggered’ by the robot’s actions in similar ways.

Comparing the behavior of the participants in the speech condition now with the participants in the ‘beep’-condition, in which the robot did not greet the participants but attracted their attention by means of a series of beeps, shows that even though much intrapersonal variation can be observed here, too, and no single robot behavior is responded to by the participants in a way that would appropriately be described as ‘triggered’, the overall distribution indicates that the initial use of speech in the speech condition has an impact on people’s overall behavior. Thus, while the robot’s behaviors do not trigger any specific responses, the anthropomorphic design of its behavior in Condition 1, i.e. the choice of human language, has an effect on how the robot is responded to in general.

Fig. 2. Attention to base, proximal and distal scene by the 8 participants in the speech condition (represented by a line each) over the course of the interaction (the nine actions of the robot).

Figure 3 shows that when the verbal greeting by the robot is replaced by a beep series, responses at the distal level occur later in the interaction, if at all. While some participants do not respond to the robot's depictions at the distal level at all, five participants do, but at different moments in the interaction. In fact, like the analysis of participants' behavior in the speech condition, the analysis of the beep condition indicates unique patterns and frequent switches between levels by the participants.

The different distributions of responses at the base, proximal and distal level in the two conditions can now be quantitatively compared. An independent sample Chi-Square test was used to compare participants' behaviors in the speech versus the beep condition; the results ($X^2 = 12.94$, $df = 2$, $p = .00155$) show a significantly different distribution in the extent to which people treat the robot as a machine compared to the extent to which they show anthropomorphizing behavior, based on the presence or absence of the verbal framing utterance at the beginning of the interaction.

5 DISCUSSION

In this paper, I have proposed a qualitative methodology that addresses people's observable behaviors in response to a robot's actions in terms of displays of their understandings of its actions, following conversation analytic principles. The data analyzed were particularly well suited for that purpose since each robot behavior had been carefully crafted with all three levels in mind: The 'puppet interface' by means of which the robot's behaviors are defined [Pantofaru et al. 2013] allows the encoding of the proximal scene (e.g. robot lifting its arm) in terms of its physical properties, such as configurations of its motors and joints (base scene); the proximal scene in turn had been created with the distal scene in mind: having the robot ask for help non-verbally (see Table 3). The encoding of participants' behaviors then just needed to check for each of the robot's behaviors to which of the three scenes participants are oriented.

Fig. 3. Attention to base (1), proximal (2) and distal (3) scene by the 11 participants in the beep condition (represented by a line each) over the course of the interaction (the ten actions of the robot).

The results show high intra- and interpersonal variation in the response to each of the robot's actions, such that no two participants acted in similar ways over the course of these interactions. Thus, participants differ with respect to what they respond to at what level. This is in line with previous work showing interpersonal variation in treating robots as social actors [Epley et al. 2007; Fischer 2011, 2016; Lee et al. 2010]. Furthermore, not a single user responded to the distal scene only, which means that not a single participant treated the robot as a social actor over the course of the whole interaction. Moreover, some participants did not respond to the robot's behavior at the proximal or distal level at all and thus did not treat it in an anthropomorphic way.

To what extent do the results correspond to other methods of measuring anthropomorphism? In the original study [Fischer et al. 2014], we used a questionnaire to assess people's perceptions of the robot after the experiment. Table 6 reports the questionnaire results from the original study for the questions, how appealing, competent, intelligent, superior, safe, approachable, and confident the robot was perceived on a 7-point Likert scale in the two conditions. The results show mean ratings from 2.91 to 5.88, suggesting that the robot is attributed with human-like characteristics to a medium degree on average. This medium degree of ascription of human-like characteristics may correspond to the back and forth between the base, proximal and distal scenes observed in the current analysis. Similarly, the standard deviations between 1.13 and 2.66 for the two conditions may correspond to the interpersonal variation that we can see in Figures 3 and 4. A significant difference between the two conditions was found only for the attribute *friendly* ($F(1,17) = 7.42, p < .05, \eta^2 = .30$ [cf. Fischer et al. 2014]), which may correspond to the overall differences between the two conditions with respect to responses at the proximal and distal level apparent from Figures 3 and 4. However, since the questionnaire results present people's *post hoc* impressions, the statistical results do not replace the detailed analysis of people's responses to the robot's actions over the course of the interaction; the quantitative comparison of people's orientation to the base, proximal and distal scene in the two condition shows that the method proposed can not only track participants' anthropomorphizing behaviors over time, but also provide a more fine-grained analysis of the effects of particular design choices, such as adding a social framing utterance.

Table 7. Means (M) and Standard Deviations (sd) for the Questionnaire Data for the two Conditions

	appealing	intelligent	competent	superior	safe	approachable	confident	friendly*
beep: M	5.63	4.13	3.75	3.00	5.38	5.88	4.75	4.46
sd	1.19	2.23	2.66	1.85	1.77	1.13	2.25	1.42
speech: M	4.64	4.00	3.64	2.91	5.36	4.36	3.27	6.25
sd	2.01	1.78	1.50	1.58	2.16	2.11	1.85	1.42

A potential limitation of the method proposed could be that it relies on people's displays of their understandings of the robot's behavior second by second, thus being tailored to identifying anthropomorphizing behavior in response to robots' interactional behavior in the moment, thus not necessarily accounting for global effects, such as the effects of a robot's anthropomorphic design. However, as the differences between the two conditions in the case study illustrate, global effects, such as those due to social framing, do become apparent in global changes in participants' local responses, i.e. their responses in the moment. The method can thus be used to identify subtle effects of individual social cues, also in a quantifiable way, as the statistical analysis of the effects of social framing has shown, where the comparison between the conditions revealed that people oriented to the base, proximal and distal scenes significantly differently, depending on whether the robot used speech or a beep. Similarly, different user types may emerge in the same way; for instance, differences between introverts and extroverts could be identified using this methodology.

Another possible objection to the analysis presented may be that the robot chosen was not very human-like, and that the beep-series played in both conditions draws participants' attention to the machine-like properties of the robot; consequently, the reason for the fact that people switch back and forth between the different scenes depicted might be due to the fact that the robot provides conflicting cues. That is, on the one hand, the robot was greeting them politely (thus evoking a distal scene), on the other, it was trying to get the participants' attention by beeping like a machine. This criticism does not affect the method proposed, which serves to trace people's behavior irrespective of how consistently the robot's behavior is designed. Furthermore, the criticism only applies to the speech condition since here signals from the human domain (speech) were mixed with more basic robot behavior and appearance; in the beep condition, the robot only uses basic behaviors and a weak anthropomorphic design. The analysis of the beep condition still shows that people switch between depiction levels also in interactions with a consistently designed robot. Moreover, much indirect evidence suggests that people generally navigate between the different depiction levels; for instance, Fussell et al.'s [2008] findings, that different methods produce different results, indicate that people have no problems navigating between base, proximal and distal scenes. Similarly, Kory Westlund et al. [2016] find that even though children responded socially to the robots introduced, they did not object to the robot being put into a box against its will. We suggest that this is similar to applauding an actor (base scene) in his devil costume (proximal scene) after a performance for having been a brilliant Mephistopheles (distal scene), where the audience also navigates with ease between the different scenes [cf. Clark 2016].

Being a qualitative method, which relies on manual encoding, the method proposed is relatively tiresome and thus restricted in applicability because of the manual labor involved. However, for the specific points in the interactions analyzed and in response to certain robot behaviors, one

could easily imagine that a learning algorithm is applied to classify people's behaviors automatically. While we cannot simply classify a human action, say waving at the robot, as an instance of a behavior at any of the three levels in general because it may have different functions in different contexts, to what level waving at the robot is oriented in response to a specific action could be automatically assigned. For instance, when the robot is gesturing for help, waving can be automatically assigned to the proximal level (the robot is in the course of scanning its environment), whereas in other situations, such as in response to a greeting, it can be automatically classified as a response on the distal scene. Thus, while the method can be carried out manually, an automatic analysis is principally possible.

The fact that our student assistant could encode the data in highly similar ways without previous training and only on the basis of the current text furthermore indicates that the method is easily learnable and replicable; this particular dataset may seem to have been easier to analyze with the method proposed than others because the intentionally crafted robot behaviors provided natural ways of segmenting the interaction and because of the three levels attended to in the robot behavior design using the 'puppet' interface, which provided an intuitive anchor point for the analysis of the human participants' behaviors. Still, as my brief analysis of the three studies by Nakata et al. [1998], Tielmann et al. [2014] and Huang & Mutlu [2013] in Section 3.2 indicates, it is relatively straightforward to identify these three levels even *post hoc* in all kinds of robot behaviors. An additional tutorial for the functional encoding and segmentation of the interactions can be downloaded a link to the tutorial to be published with the main article.

6 IMPLICATIONS

In this study, we have analyzed participants' responses to a robot's behaviors based on whether they attend to the robot's mechanical properties, the proximal scene evoked or the distal scene. Thus, we have described people's behaviors, not their mental models or mental attributions of human-like properties to the robot; consequently, we cannot make assertions about what people believe about the robot.⁴ However, the results do show considerable intra- and interpersonal variation in the way people respond to identical robot behaviors, which has theoretical implications for our understanding of anthropomorphism. For instance, we have found that participants respond to the robot's actions with anthropomorphizing behaviors, such as reciprocating a greeting, offering help, asking what it wants etc., in one moment, but that in the next moment, they may also treat it like a machine. That is, even if some of the robot's actions, such as the polite verbal greeting, are often responded to in an anthropomorphizing manner at the distal level, this does not mean that the robot is responded to with anthropomorphism with regard to other actions. Instead, people shift between treating the robot like another human and treating it like a machine with no apparent difficulty. Consequently, one implication of this study is that people do not anthropomorphize robots in general, but respond anthropomorphically to individual aspects or behaviors of robots.

From this perspective, it is misleading to assume that robots *are* social actors: People rather seem to treat robots as social actors temporarily. These findings cast doubt on the assumption that people mindlessly attribute social characteristics to robots, as suggested in the media equation hypothesis [Reeves & Nass 1996], which was further developed into the Computers Are Social Actors paradigm [Nass & Moon 2000]. Instead, people do not seem to confuse robots with humans since they regularly return to treating the robot as a machine in between their anthropomorphizing behaviors.

⁴Similarly, because it relies on people's observable behavior, the method cannot safeguard against experimental bias, such that participants may behave in a way to please the experimenter, because it does not address motivations and intentions, just what can be observed.

Furthermore, people's anthropomorphizing behaviors do not seem to be 'triggered' by certain robot behaviors or its appearance, as suggested, for instance, by Nass [2004], since behaviors on the proximal and distal level occur at various places in the interactions and not directly in reaction to specific robot behaviors, as the concept of a 'trigger' would suggest.

Moreover, we have seen that even if people do not interpret a robot's behavior as mechanical movement, they still do not necessarily expect the robot to have its own agenda. Thus, anthropomorphism comes in different degrees, which is reflected in our distinction between proximal and distal scenes (Clark & Fischer under revision). An example is participant H101's response to the robot's gesturing towards the shelf; while recognizing the robot's behavior as attention to the shelf, he did not expect the robot to have its own intention for reaching up since he repeatedly provided the robot with the instruction to pick a book. (The failure to attribute intention into the robot in this situation does not keep him, however, from offering the robot a glass of water a few seconds later, even though robots are not particularly good with water and certainly don't need any.) The example illustrates that anthropomorphism is not an all-or-none phenomenon. This is in accordance with the suggestion by Ruijten et al. [2019] that some traits are more difficult to attribute to a robot than others. However, approaches that see anthropomorphizing responses to robots as based on a psychological disposition [e.g. Epley et al. 2007, 2008; Epley 2018; Ruijten et al. 2019] will also have problems explaining the intrapersonal variation over time. That is, if a general disposition caused anthropomorphizing behaviors, people should either anthropomorphize the robot and treat it as if it was a person, or they should not. In contrast, what we can observe is that people treat robots as if they had human-like characteristics in one second and as if they had not in the next.

Similarly, social constructivist approaches [e.g. Edwards 1994] do not necessarily explain the interpersonal differences and the changes in the attribution of human-like characteristics to robots, either. Edwards [1994, 2001], for instance, assumes intention attribution to be a common social practice, which is used to make sense of other people's behavior, as much as for making sense of animals' and robots' behaviors, since intentions and other psychological processes always need to be inferred. From this perspective, one might expect people to treat robots in a human-like manner throughout.

The method has furthermore some practical implications; for instance, it cannot only be used to track subtle effects of certain social cues in the robot's behavior and thus support the analysis of the effects of certain design choices in the robot behavior development, it could also serve to check at specific points in interaction to which depiction level the respective user is oriented and to decide on that basis between different courses of future action, as suggested by Simmons et al. [2011]. It can thus be used to adapt the robot's behavior in real time.

To sum up, I have presented a method to describe the dynamics of anthropomorphizing behavior in HRI. This method complements the set of methods already available to shed light on the different facets of anthropomorphism. The application of the method proposed here in a case study revealed that people respond to different social 'triggers' differently and move between anthropomorphizing and non-anthropomorphizing responses with apparent ease. These findings have consequences for explanatory accounts of anthropomorphism (cf. Clark & Fischer under revision); future work will have to determine the relationship between people's anthropomorphizing behavior in response to the robot and their concepts of these robots.

REFERENCES

- S. Andrist, E. Spannan, and B. Mutlu. 2013. Rhetorical robots: Making robots more effective speakers using linguistic cues of expertise. In *Proceedings of the 2013 Human-Robot Interaction Conference* pp. 341–348.

- S. Andrist, X. Z. Tan, M. Gleicher, and B. Mutlu. 2014. Conversational gaze aversion for humanlike robots. In *Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interaction* (pp. 25–32). ACM.
- S. Andrist, M. Ziadde, H. Boukaram, B. Mutlu, and M. Sakr. 2015. Effects of culture on the credibility of robot speech: A comparison between English and Arabic. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction (HRI'15)*. ACM, New York, NY, USA. 157–164.
- T. Arnold and M. Scheutz. 2017. Beyond moral dilemmas: Exploring the ethical landscape in HRI. *Proceedings of the 2017 International Conference on Human-Robot Interaction* 445–452.
- A. Bangerter and E. Major. 2013. Interactional theories of communication. In Cobley, P. & Schulz, P.J. (Eds.). *Theories and Models of Communication. Handbook of Communication Science*. Berlin: De Gruyter. 257–271.
- C. Bartneck, E. Croft, D. Kulic, and S. Zoghbi. 2009. Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International Journal of Social Robotics* 1, 1 (2009) 71–81.
- J. B. Bavelas, L. Coates, and T. Johnson. 2000. Listeners as co-narrators. *Journal of Personality and Social Psychology* 79, 6 (2000), 941–952.
- J. M. Beer, A. D. Fisk, and W. A. Rogers. 2014. Toward a framework for levels of robot autonomy in human-robot interaction. *Journal of Human-Robot Interaction* 3, 2 (2014), 74–99.
- J. M. Bradshaw. 2013. From knowledge science to symbiosis science. *International Journal of Human-Computer Studies* 71 (2013), 171–176.
- H. P. Branigan, M. J. Pickering, J. Pearson, and J. McLean. 2010. Linguistic alignment between people and computers. *Journal of Pragmatics* 42, 2355–2368.
- H. Branigan and J. Pearson. 2006. Alignment in human-computer interaction. In K. Fischer (Ed.), *Proceedings of the Workshop on 'How People Talk to Computers, Robots, and Other Artificial Communication Partners'*, Hansewissenschaftskolleg, Delmenhorst, April 21–23, 2006. SFB/TR8 Report 010-09 2006.
- D. Brščić, H. Kidokoro, Y. Suehiro, and T. Kanda. 2015. Escaping from children's abuse of social robots. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction (HRI'15)*. Association for Computing Machinery, New York, NY, USA, 59–66. DOI: <https://doi.org/10.1145/2696454.2696468>
- D. A. Bryant, J. Borenstein, and A. Howard. 2020. Why should we gender? The effect of robot gendering and occupational stereotypes on human trust and perceived competency. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction* (pp. 13–21).
- C. M. Carpinella, A. B. Wyman, M. A. Perez, and S. J. Stroessner. 2017. The robotic social attributes scale (RoSAS): Development and validation. *Proceedings of HRI'17*, Vienna.
- T. Chaminade, M. Zecca, S.-J. Blakemore, A. Takanishi, C. D. Frith, S. Micera, P. Dario, G. Rizzolatti, V. Gallese, and M. A. Umiltà. 2010. Brain response to a humanoid robot in areas implicated in the perception of human emotional gestures. *PLoS ONE* 5, 7 (2010), e11577.
- H. H. Clark. 1996. *Using Language*. Cambridge: Cambridge University Press.
- H. H. Clark and E. F. Schaefer. 1989. Contributing to discourse. *Cognitive Science* 13, 259–294.
- H. H. Clark. 2016. Depicting as a method of communication. *Psychological Review* 123, 3 (2016), 324–347.
- H.H. Clark and K. Fischer. (under revision): Robots as Dynamic Depictions.
- M. F. Damholdt C. Vestergaard, and J. Seibt. 2020. Testing for 'Anthropomorphization': A case for mixed methods in human-robot interaction. In Jost C. et al. (eds) *Human-Robot Interaction. Springer Series on Bio- and Neurosystems*, vol 12. Springer, Cham
- N. Epley, A. Waytz, and J. T. Cacioppo. 2007. On seeing human: A three-factor theory of anthropomorphism. *Psychological Review* 114, 4 (2007), 864.
- N. Epley, A. Waytz, S. Akalis, and J. T. Cacioppo. 2008. When we need a human: Motivational determinants of anthropomorphism. *Social Cognition* 26, 2, pp. 143–155.
- N. Epley. 2018. A mind like mine: The exceptionally ordinary underpinnings of anthropomorphism. *Journal of the Association for Consumer Research* 3, 4 (2018), 591–598.
- F. A. Eyseland and D. Kuchenbrandt. 2012. Social categorization of social robots: Anthropomorphism as a function of robot group membership. *British Journal of Social Psychology* 51 (2012), 724–731.
- R. Falcone and C. Castelfranchi. 2001. The human in the loop of a delegated agent: The theory of adjustable social autonomy. *IEEE Transactions of Systems, Man and Cybernetics – Part A: Systems and Humans* 31 (2001), 5.
- J. Fink. 2012. Anthropomorphism and human likeness in the design of robots and human-robot interaction. *ICSR 2012*, pp. 199–208.
- K. Fischer. 2006. What computer talk is and isn't. *Human-Computer Conversation as Intercultural Communication. Linguistics – Computational Linguistics* 17, Saarbrücken: AQ.
- K. Fischer. 2011. Interpersonal variation in understanding robots as social actors. In *Proceedings of HRI'11*, March 6-9th, 2011. Lausanne, Switzerland, pp. 53–60.

- K. Fischer, K. Lohan, and K. Foth. 2012. Levels of embodiment: Linguistic analyses of factors influencing HRI. In *2012 7th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 463–470). IEEE.
- K. Fischer, B. Soto, C. Pontafaru, and L. Takayama. 2014. The effects of social framing on people's responses to robots' requests for help. *Proceedings of the IEEE Conference on Robot-Human Interactive Communication – Ro-man'14*, Edinburgh.
- K. Fischer, S. Yang, B. Mok, R. Maheshwari, D. Sirkin, and W. Ju. 2015. Initiating interactions and negotiating approach: A robotic trash can in the field. *AAAI Symposium on Turn-taking and Coordination in Human-Machine Interaction*, Stanford, March 2015.
- Fogg and C. I. Nass. 1997. Silicon sycophants: Effects of computers that flatter. *International Journal of Human-Computer Studies* 46, 5 (1997), 551–561.
- J. Forlizzi. 2007. How robotic products become social products: An ethnographic study of cleaning in the home. In *2007 2nd ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 129–136). IEEE.
- S. R. Fussell, S. Kiessler, L. D. Setlock, and V. Yew. 2008. How people anthropomorphize robots. *Proceedings of HRI'08*, Amsterdam, p. 145–152.
- H. Garfinkel. 1972. Remarks on ethnomethodology. In J. J. Gumperz and D. Hymes (Eds.), *Directions in Sociolinguistics. The Ethnography of Communication*. New York : Holt, Rinehart and Winston.
- F. Hegel. 2012. Effects of a robot's aesthetic design on the attribution of social capabilities. *Proceedings of the 21st Ro-Man Conference*, Paris, France.
- Chien-Ming Huang and Bilge Mutlu. 2013. Modeling and evaluating narrative gestures for humanlike robots. In *Robotics: Science and Systems*, pp. 57–64. 2013.
- J. Ham, M. van Esch, Y. Limpens, J. de Pee, J. J. Cabibihan, and S. S. Ge. 2012. The automaticity of social behavior towards robots: The influence of cognitive load on interpersonal distance to approachable versus less approachable robots. In *Social Robotics. ICSR 2012*, S. S. Ge, O. Khatib, J. J. Cabibihan, R. Simmons, and M.-A. Williams (Eds.). Lecture Notes in Computer Science, vol 7621. Springer, Berlin, Heidelberg.
- I. Hutchby. 2001. *Conversation and Technology: From the Telephone to the Internet*. Cambridge: Polity.
- G. Jefferson. 2004. Glossary of transcript symbols with an introduction. *Pragmatics and Beyond New Series*, 125 (2004), 13–34.
- P. H. Kahn, Takayuki Kanda, Hiroshi Ishiguro, Brian T. Gill, Solace Shen, Heather E. Gary, Jolina H. Ruckert 2015. Will people keep the secret of a humanoid robot?: Psychological intimacy in HRI. *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, pp. 173–180.
- A. Klapper, R. Ramsey, D. Wigboldus, and E. S. Cross. 2014. The control of automatic imitation based on bottom-up and top-down cues to animacy: Insights from brain and behavior. *Journal of Cognitive Neuroscience* 26, 11 (2014), 2503–2513.
- S. Kopp. 2006. How people talk to a virtual human - conversations from a real-world application. In K. Fischer (Ed.), *Proceedings of the Workshop on 'How People Talk to Computers, Robots, and Other Artificial Communication Partners'*, Hansewissenschaftskolleg, Delmenhorst, April 21–23, 2006, pp. 131–139. SFB/TR8 Report 010-09 2006.
- S. Krach, F. Hegel, B. Wrede, G. Sagerer, F. Binkofski, and T. Kircher. 2008. Can machines think? Interaction and Perspective Taking with Robots Investigated via fMRI. *PloS ONE* 2008.
- S. Kuz, H. Petruck, M. Heisterüber, H. Patel, B. Schumann, C. M. Schlick, and F. Binkofski. 2015. Mirror neurons and human-robot interaction in assembly cells. *Procedia Manufacturing* 3 (2015), 402–408.
- T. Law, M. Chita-Tegmark, and M. Scheutz. 2020. The interplay between emotional intelligence, trust, and gender in human-robot interaction. *International Journal of Social Robotics* 1–13.
- K. M. Lee, Y. Jung, J. Kim, and S. R. Kim. 2006. Are physically embodied social agents better than disembodied social agents? The effects of physical embodiment, tactile interaction and people's loneliness in human-robot interaction. *International Journal of Human-Computer Studies* 64 (2006), 962–973.
- M. K. Lee, S. Kiesler, and J. Forlizzi. 2010. Receptionist or information kiosk: How do people talk with a robot? In *Computer Supported Cooperative Work*, Savannah, Georgia, pp. 31–40.
- Sau-lai Lee, Sara Kiesler, Ivy Yee-man Lau, and Chi-Yue Chiu. 2005. Human mental models of humanoid robots. *Proceedings of ICRA 2005*, Barcelona, Spain.
- A. Leeper, K. Hsiao, M. Ciocarlie, L. Takayama, and D. Gossow. 2012. Strategies for human-in-the-loop robotic grasping. *Proc. of Human-Robot Interaction (HRI'12)*, Boston, MA.
- S. Lemaignan, J. Fink, and P. Dillenbourg. 2014. The dynamics of anthropomorphism in robotics. *Proceedings of HRI'14*, p. 226–227.
- C. Stephen Levinson. 1983. *Pragmatics*. Cambridge University Press.
- C. Stephen Levinson. 2006. On the human 'interaction engine'. In N. C. Enfield and S. C. Levinson (Eds.), *Roots of Human Sociality: Culture, Cognition and Interaction*, pp. 39–69. Oxford/New York: Berg.
- J. J. Li, W. Ju, and B. Reeves. 2017. Touching a mechanical body: Tactile contact with body parts of a humanoid robot is physiologically arousing. *Journal of Human-Robot Interaction* 6, 3 (2017), 118–130.

- H. Mieczkowski, S. Xun Liu, J. Hancock, and B. Reeves. 2019. Helping Not Hurting: Applying the Stereotype Content Model and BIAS Map to Social Robotics. *Proceedings of the 14th International Conference on Human-Robot Interaction*. Daegu, Korea.
- Toru Nakata, Tomomasa Sato, Taketoshi Mori, and Hiroshi Mizoguchi. 1998. Expression of emotion and intention by robot body movement. In *Proceedings of the 5th International Conference on Autonomous Systems*. 1998.
- C. Nass and S. Sundar. 1996. Are programmers psychologically relevant to human-computer interaction? *Paper presented at the Annual Meeting of the International Communication Association*.
- C. Nass and Y. Moon. 2000. Machines and mindlessness: Social responses to computers. *Journal of Social Issues* 56, 1 (2000), 81–103.
- C. Nass. 2004. Etiquette equality: exhibitions and expectations of computer politeness. *Communications of the ACM* 47, 4 (2004), 35–37.
- C. Nass. 2010. *The Man Who Lied to his Laptop: What Machines Teach us about Human Relationships*. New York: Penguin.
- C. Nass and S. Brave. 2005. *Wired for Speech. How Voice Activates and Advances the Human- Computer Relationship*. Cambridge, MA. London: MIT Press.
- Caroline Pantofaru, Austin Hendrix, Andreas Paepcke, Dirk Thomas, Sharon Marzouk, and Sarah Elliott. 2013. Programming robots at the museum. *ACM International Conference Proceeding Series*. 305–308. DOI: <https://doi.org/10.1145/2485760.2485797>.
- M. Paetzel, G. Perugia, and G. Castellano. 2020. The persistence of first impressions: The effect of repeated interactions on the perception of a social robot. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*. 73–82.
- L. A. Penner. 2002. The causes of sustained volunteerism: An interactionist perspective. *Journal of Social Issues* 58 (2002), 447–468.
- A. Powers and S. Kiesler. 2006. The advisor robot: Tracing people’s mental model from a robot’s physical attributes. *Proceedings of the Human-Robot Interaction Conference HRI’06*, Salt Lake City, Utah, USA.
- H. Preece, J. Rogers, and Y. Sharp. 2015. *Interaction Design: Beyond Human Computer Interaction* (4th Edition). Wiley.
- R. Read and T. Belpaeme. 2013. People interpret robotic non-linguistic utterances categorically. In *Proceedings of the 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. ACM, New York, pp. 209–210.
- R. Read and T. Belpaeme. 2014. Non-linguistic utterances should be used alongside language, rather than on their own or as a replacement. In *Proceedings of the 9th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. ACM, New York, pp. 276–277.
- B. Reeves and C. Nass. 1996. *The Media Equation*. Stanford: CSLI and Cambridge: Cambridge University Press.
- B. Reeves and J. Hancock. 2020. Social robots are like real people: First impressions, attributes, and stereotyping of social robots. *Issue 1*, 1 (2020), 1.
- L. D. Riek. 2017. Healthcare robotics. *Communications of the ACM* 60, 11, pp. 68–78.
- A. Rosenthal-von der Pütten. and N. Krämer. 2014. How design characteristics of robots determine evaluation and uncanny valley related responses. *Computers in Human Behavior* 36 (2014), 422–439.
- P. A. M. Ruijten. 2015. Responses to human-like artificial agents: Effects of user and agent characteristics. Technische Universiteit Eindhoven.
- P. A. Ruijten, A. Haans, J. Ham, and C. J. Midden. 2019. Perceived human-likeness of social robots: testing the Rasch model as a method for measuring anthropomorphism. *International Journal of Social Robotics* 11, 3 (2019), 477–494.
- H. Sacks. 1984. Notes on methodology. In J. Atkinson and J. Heritage (Eds.), *Structure of Social Action: Studies in Conversation Analysis*, pp. 21–27. Cambridge: Cambridge University Press.
- H. Sacks. 1996. *Lectures on conversation, Volume I & II*. Oxford: Blackwell.
- H. Sacks, E. A. Schegloff, and G. Jefferson. 1974. A simplest systematics for the organization of turn-taking for conversation. *Language* 50, 4 (1974), 696–735.
- E. A. Schegloff and H. Sacks. 1973. Opening up closings. *Semiotica* 8, 4 (1973), 289–327.
- J. Seibt. 2017. Towards an ontology of simulated social interaction: Varieties of the “as if” for robots and humans. In R. Hakli, J. Seibt (eds.), *Sociality and Normativity for Robots, Studies in the Philosophy of Sociality* 9. Springer.
- S. Seo, D. Geiskovitch, Masayuki Nakane, Corey King, and James E. Young. 2015. Poor thing! Would you feel sorry for a simulated robot? A comparison of empathy toward a physical and a simulated robot, In *Proceedings of the 10th ACM/IEEE International Conference on Human-Robot Interaction, HRI’2015*, Portland, Oregon, USA.
- R. Simmons, M. Makatchev, R. Kirby, M.K. Lee, I. Fanaswala, B. Browning, J. Forlizzi, and M. Sakr. 2011. Believable robot characters. *AI Magazine* 32, 4 (2011), 39–52.
- V. K. Sims, M. G. Chin, D. J. Sushil, D. J. Barber, T. Ballion, B. R. Clark, K. A. Garfield, M. J. Dolezal, R. Shumaker, and N. Finkelstein. 2005. Anthropomorphism of robotic forms: A response to affordances? *Proceedings of the 49th Annual Meeting of the Human Factors and Ergonomics Society*.

- A. Stenzel, E. Chinellato, B. Tiradou, Maria A. del Pobil, Angel P., Lappe, Markus and Roman Liepelt. 2012. When humanoid robots become human-like interaction partners: Corepresentation of robotic actions. *Journal of Experimental Psychology* 38, 5 (2012), 1073–1077.
- T. Stivers, N. J. Enfield, P. Brown, C. Englert, M. Hayashi, T. Heinemann, G. Hoymann, F. Rossano, J. P. De Ruiter, K. E. Yoon, and S. C. Levinson. 2009. Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences* 106, 26 (2009), 10587–10592.
- Myrthe Tielman, Mark Neerinx, John-Jules Meyer, and Rosemarijn Looije. 2014. Adaptive emotional expression in robot-child interaction. In *Proceedings of the 2014 International Conference on Human-Robot Interaction*, pp. 407–414. ACM, 2014.
- K. M. Tsui, D.-J. Kim, A. Behal, D. Kontak, and H. A. Yanco. 2011. I want that: Human-in-the-loop control of a wheelchair-mounted robotic arm. *Journal of Applied Bionics and Biomechanics* 8, 1 (2011), 127–147.
- S. Turkle, C. Breazeal, O. Dasté, and B. Scassellati. 2006. Encounters with kismet and cog: Children respond to relational artifacts. *Digital Media: Transformations in Human Communication*. 120.
- K. Walton. 1993. Metaphor and prop oriented make-believe. *European Journal of Philosophy* 1, 1 (1993), 39–57.
- A. Weiss, R. Bernhaupt, D. Schwaiger, M. Altmaninger, R. Buchner, and M. Tscheligi. 2009. User experience evaluation with a Wizard of Oz approach: Technical and methodological considerations. In *9th IEEE-RAS International Conference on Humanoid Robots, Humanoids 2009*, Paris, France, December 7–10, 2009, pp. 303–308.
- Kory Westlund, M. Jacqueline, Marayna Martinez, Maryam Archie, Madhurima Das, and Cynthia Breazeal. 2016. Effects of framing a robot as a social agent or as a machine on children’s social behavior. *Proceedings of the HRI’16 Conference*, Christchurch NZ.
- S. A. Wensveen, J. P. Djajadiningrat, and C. J. Overbeeke. 2004. Interaction frogger: A design framework to couple action and function through feedback and feedforward. In *Proceedings of the 5th Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques* (pp. 177–184). ACM.
- A. Wykowska, T. Chaminade, and G. Cheng. 2016. Embodied artificial agents for understanding human social cognition. *Philosophical Transactions of the Royal Society B: Biological Sciences* 371, 1693 (2016), 20150375.
- J. Zlotowski, H. Sumioka, F. Eyssel, S. Nishio, C. Bartneck, and H. Ishiguro. 2018. Model of dual anthropomorphism: The relationship between the media equation effect and implicit anthropomorphism. *International Journal of Social Robotics* 10, 5 (2018), 701–714.

Received November 2019; revised June 2020; accepted September 2020