



# From Inanimate Object to Agent: Impact of Pre-beginnings on the Emergence of Greetings with a Robot

**DAMIEN RUDAZ**, Dept. of Economics and Social Sciences, Telecom Paris and Institut Polytechnique de Paris, France

**KAREN TATARIAN**, Institute for Intelligent Systems and Robotics, Sorbonne University, France

**REBECCA STOWER**, Division of Robotics, Perception, and Learning, KTH Royal Institute of Technology, Sweden

**CHRISTIAN LICOPPE**, Dept. of Economics and Social Sciences, Telecom Paris and Institut Polytechnique de Paris, France

---

The very first moments of co-presence, during which a robot appears to a participant for the first time, are often “off-the-record” in the data collected from human-robot experiments (video recordings, motion tracking, methodology sections, etc.). Yet, this “pre-beginning” phase, well documented in the case of human-human interactions, is not an interactional vacuum: It is where interactional work from participants can take place so the production of a first speaking turn (like greeting the robot) becomes relevant and expected. We base our analysis on an experiment that replicated the interaction opening delays sometimes observed in laboratory or “in-the-wild” human-robot interaction studies—where robots can require time before springing to life after they are in co-presence with a human. Using an ethnomethodological and multimodal conversation analytic methodology (EMCA), we identify which properties of the robot’s behavior were oriented to by participants as creating the adequate conditions to produce a first greeting. Our findings highlight the importance of the state in which the robot originally appears to participants: as an immobile object or, instead, as an entity already involved in preexisting activity. Participants’ orientations to the very first behaviors manifested by the robot during this “pre-beginning” phase produced *a priori* unpredictable sequential trajectories, which configured the timing and the manner in which the robot emerged as a social agent. We suggest that these first instants of co-presence are not peripheral issues with respect to human-robot experiments but should be thought about and designed as an integral part of those.

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

Additional Key Words and Phrases: Pre-beginning, greetings, social agent, conversation analysis, ethnomethodology, robot latencies, computers are social actors, anthropomorphism

---

Authors’ addresses: D. Rudaz and C. Licoppe, Dept. of Economics and Social Sciences, Telecom Paris and Institut Polytechnique de Paris, France; emails: damien.rudaz@telecom-paris.fr, christian.licoppe@telecom-paris.fr; K. Tatarian, Institute for Intelligent Systems and Robotics, Sorbonne University, France; email: tatariankaren@gmail.com; R. Stower, Division of Robotics, Perception, and Learning, KTH Royal Institute of Technology, Sweden; email: stower@kth.se.



This work is licensed under a [Creative Commons Attribution International 4.0 License](https://creativecommons.org/licenses/by/4.0/).

© 2023 Copyright held by the owner/author(s).

2573-9522/2023/04-ART29

<https://doi.org/10.1145/3575806>

**ACM Reference format:**

Damien Rudaz, Karen Tatarián, Rebecca Stower, and Christian Licoppe. 2023. From Inanimate Object to Agent: Impact of Pre-beginnings on the Emergence of Greetings with a Robot. *ACM Trans. Hum.-Robot Interact.* 12, 3, Article 29 (April 2023), 31 pages. <https://doi.org/10.1145/3575806>

---

## 1 INTRODUCTION

A growing body of studies has recently approached the “social” status of a robot not as a “categorical property of the robot’s inside” [1], or as a “specifiable and implementable set of features” [51] but, instead, as an emergent phenomenon [66]. These works follow an ethnomethodological perspective for which, like many seemingly objective and trans-situational properties of social life, the “social” status of a “social robot” is not given but “locally produced, incrementally developed, and, by extension, [...] transformable at any moment” [32]. By closely studying the moment-to-moment interactions of humans with a robot, researchers using these approaches attempt to grasp through which local processes a robot can, momentarily, emerge as a “social partner” [91], a “social actor” [19, 62], a “subject” or “social being” [92], a “social agent” [1], an “artificial agent” [51], or as a new (and evolving over time) ontological category [66]. Without necessarily sharing the same theoretical background, they rely on a common assumption that, in the same way that “a situation becomes observable and is treated as the meeting of a jury when participants produce practices that others orient and respond to as practices of a jury” [22, 48], an entity becomes a “social robot” when it produces practices that humans orient and respond to as practices of a social robot.

Yet, this definition of sociality as an emergent feature “raises a question about the minimal conditions for a social interaction” [39]. Indeed, once a robot is demonstrably shown to be oriented to as a “social agent,” there remains the issue of what shifted this robot’s interactional status; especially when it was approached as a non-social object at first. What properties of the local situation were observably oriented to as relevant for participants before a “social” interaction could emerge—suddenly or incrementally, momentarily or durably? Our study extends on this ethnomethodological line of research by focusing on the initial emergence of a first, conditionally relevant, greeting. Based on an experiment from which were collected 80 recordings of dyadic interactions between a human and a robot, we try to identify (1) *if and when* a shift occurs from an inanimate artifact to an agent and (2) *how* this shift progressively emerges, when it does.

Of all the properties that could be constituted and oriented to as “relevant features” in the continuous flow of a local human-robot interaction, we wonder which of them were treated by participants as creating the adequate conditions to produce a first greeting move. That is, we attempt to describe the interactional work required before *behaviors* from the robot could be treated as *actions* that either established the adequate framework for the participant to initiate a first greeting sequence or, alternatively, produced a response slot that the participant was normatively pressured to complete with a return greeting. Overall, four typical paths to the emergence of a first greeting were identified, and one where the production of a greeting never became relevant. This typology is exemplified through the analysis of five fragments representative of our corpus.

## 2 PREVIOUS WORK

### 2.1 The Significance of Greetings for Human-robot Interactions

Human greetings, as a canonical part of the opening sequence in many interactional settings, are among the most documented practices in conversation analysis: from the first analysis of telephone calls [78, 79], to video calls [50, 55], to co-present encounters [12, 27, 42, 59, 64, 74]. Greeting sequences have been of special interest in the study of human-robot interactions, not only because

of the amount of data available, but also because of what they accomplish in human-human interactions. Indeed, they simultaneously reflect and construct the mutual status of the co-interactants by being tailored “to (display) their own understanding/appraisal of ‘who we are to one another right now’” [64], manifesting, in particular, that another is “recognized and categorized as a possible partner for future interaction” [58]. They also tend to be connected with observable changes in the structure of talk and in the physical configuration of participants, as they often form both “the end of a phase of incipient interaction” and “the first exchange of a conversation” [79]. In summary, greetings are critical for “organizational reasons (coordinated, well-tuned, reciprocal engagement), social reasons (recognition, display of the type of relationship the participants entertain), and normative reasons (mutual trust)” [58].

The previous approaches focus on what greetings *do* in an interaction. In parallel, several quantitatively oriented **human-robot interaction (HRI)** studies have highlighted the significance of greetings as an *indicator*. Humans who greeted a robot were commonly found to display specific behaviors during the rest of their interaction and/or to hold specific representations or perceptions of the robot. Notably, producing a greeting was observed to be a predictor of a “more social script” [46], of “patterns of discourse” [53], or of specific linguistic behaviors [18]. Greetings were also shown to correlate with the attribution “of higher linguistic, perceptual, and cognitive competence” to the robot [17] and, in particular, reacting to a robot’s greeting wave was observed by Baddoura and Gentiane [5] to be significantly correlated with evaluating this robot as sociable—suggesting that responding to a greeting documents more than a simple mimetic reflex action. However, Holthaus [34] demonstrated that the multimodal behavior of a humanoid robot can heavily impact the number of participants who greet it, and the timing of their greetings, highlighting that the first moments of the interaction can play a heavy role in constituting a framework in which a greeting sequence becomes relevant.

## 2.2 Conditional Relevance as a Breaking Point in the Robot’s Status as an Interactant

As “highly ritualized actions” [90], greetings are also typical cases involving conditional relevance. This concept refers to a property that binds together two turns at talk—from different speakers—in an interaction [44, 84]: sequences of question-answer, invitation-acceptance (or refusal), and so on. In each of these sequences, a first pair-part (e.g., a question) makes relevant for the recipient the production of “a second pair-part of the same sequence type” [44] (e.g., an answer). For example, a greeting creates expectations of a reciprocal greeting, whose absence can be accountably oriented to by participants as “a meaningful departure from the norm” [44]. Two turns united by conditional relevance (i.e., a “first pair part” and a “second pair part”) form an “adjacency pair.” Significantly, conditional relevance is not a mere statistical observation (that a first pair part tends to be followed by a second one of a certain type) but corresponds to the achievement of a “normative organization” [44] where a first action from a participant imposes constraints “on the type and form of action with which the recipient should respond” [44]. In this sense, Reference [37] argues that “when people respond to a social robot’s greetings, they do not merely respond to the robot, but orient to the moral obligation involved in the normative practice of greetings.”

Because of this documented property of greetings, the treatment of an action of a robot as initiating a first greeting pair part has been connected to the status of this robot as an interactant [1, 62]. A *response* to a robot’s greeting suggests that, at this specific moment, humans orient towards the robot as an entity that can impose normative constraints. Locally and momentarily, the robot stops being treated as an object performing an autonomous script—which is not inserted in an orderly sequence of conversational turns [78]—but is, rather, oriented to as an agent or partner [51, 91] whose actions can have (normative) consequences for the recipient [15, 37]. The emergence of a sequence of (mutual) greetings may therefore constitute an observable breaking point in the

interaction dynamic. When a participant initiates a first conditionally relevant greeting pair or responds to a robot's greeting action, this "enactment of the greeting ritual models the appropriate and expected way of acting and interacting that constitutes the addressee as a particular kind of entity" [1]. Relying on the distinction from Reference [37], a mutual greeting sequence with a robot can be said to display participants who are "talking" rather than "using speech": At this instant, the robot is treated as producing actions "discoverable within a normative order" [37] and, conversely, as having the capabilities to interpret the normativity of other participants' actions [37].

Crucially, mutual greetings emerge from a preexisting situation; they do not appear out of an interactional vacuum. Precisely because they enact the existence of a normative order over both the greeter and the greeted [37], the appearance of greetings supposes a "framework in which a greeting sequence is relevant and expectable" by the participants [59], or "a proper interaction frame" [50], usually established as part of the "pre-beginning" [79] or "pre-opening" [55]. Yet, co-participants accomplish various degrees of interactional "work" to establish such a framework. In particular, for humanoid robots, emerging as "agents" is not systematically granted by the sharing of a mutual space with other participants. Robots may require, even more than humans, to achieve "the type of self-affirming done through language [which] is of a different nature from mere physical presence" [15].

### 2.3 Pre-beginning Designs in Human-robot Interactions: "Coming into Sight" vs. "Coming into Existence"

2.3.1 *Pre-beginning Designs in HRI.* Focusing exclusively on the moment at which the robot appears<sup>1</sup> to participants for the first time, HRI studies and datasets collected in controlled or natural settings can, at first glance, be sorted into two general categories.

- (1) **Studies where the robot stands motionless when participants encounter it**, without displaying any preexisting idle behavior nor adjustments to the participants' approach or presence: the Wizard of Oz has to seat participants in front of the robot before going behind a divider to send commands to the robot (e.g., Reference [8]) or has to deal with a significant response time (e.g., References [77, 95]), the script is not launched yet (e.g., References [62, 93]—this study), the autonomous robot's reactions are delayed, and so on. In these situations, participants find themselves in physical copresence with the robot for a long period before reciprocal exchanges and mutual identification become possible: There is a non-accounted for delay "between entry into physical copresence and moves to enter into social copresence" [65].
- (2) **Studies where the robot, or virtual agent, is already observably involved in a preexisting activity** when it appears to participants (including idling behaviors such as simulated breathing, random head movements, and so on—e.g., References [4, 52, 73, 98]) and/or observably adjusts to the human's approach or physical co-presence (tracking their gaze, waving, producing a non-delayed greeting, approaching them, and so on—e.g., References [7, 30, 34, 40, 46, 67]). This includes any form of activity from the robot that may be witnessed by participants prior to their own interaction with it, similarly to human service-encounters where salespersons, doctors, help-desk staff, sushi chefs, and so on, are often already immersed in a (potentially competing) activity when they are sighted by the customer/patient/student [27, 59, 74, 97].

2.3.2 *"Coming into Sight" and "Coming into Existence."* The two types of HRI pre-beginnings described above make relevant an earlier distinction made by J. J. Gibson in his ecological

<sup>1</sup>I.e., when the robot enters the perceptual field of participants.

psychology, regarding the way humans may appear on the social scene, and gradually achieve participant status in the pre-beginnings of encounters. In co-present encounters in relatively uncluttered spaces, co-participants usually get into a greeting position progressively, relying on the way they move, their gaze and gestures to continuously coordinate their getting-together, and make relevant interactional moves such as distant greetings [42]. Gibson calls this type of appearance a “coming into sight” [24]. This is the most common configuration in co-present encounters. He opposes to this another type of appearance, in which the other person seems to materialize or come to life suddenly in the situation, as when someone hidden by features of the local environment suddenly becomes visible, which Gibson calls “coming into existence” [24] to allow for the “pop-up,” quasi-instantaneous character. Other examples of “coming into existence” would be the initial connection in a video call [50], or in a co-present encounter, someone sleeping who suddenly wakes up after being approached. Because of its suddenness, the exact moment of “comings into existence” can be difficult to anticipate, the causes and underlying processes for such a “coming into existence” are not apparent, and finally, the interactional status and competence of the potential co-participant at the moment of its “coming into existence” can be uncertain.

*2.3.3 “Off-the-record” Pre-beginnings in HRI.* Gibson’s distinction may be highly applicable to HRI, for it now appears that in the pre-beginnings of encounters of the first type of studies mentioned above, more or less prepared subjects have to deal with a robot that “comes into existence,” while in studies of the second type, the robot may seem to “come into sight” and allow for some form of embodied mutual co-ordination in the pre-beginning phase. However, in most cases, studies neither analyze nor clarify the state of the robot when participants see it [4] or enter in physical co-presence with it. Indeed, as “most experimental studies only start when the human is already placed in the appropriate starting position in front of the robot” [33], methodology sections rarely cover the observable behavior of the robot when participants encounter it. HRI experiments display an orientation to the “opening” phase of the interaction as the first relevant moment and tend to set aside the “pre-beginning” phase, although, depending on the scenario, it may be crucial to the way subjects and robots achieve some form of co-participation status. These very first seconds, during which the robot appears to participants, are often, so to speak, “off-the-record” in the data that ends up being collected.

Here, we will analyze a relatively common HRI experimental setup in which subjects are brought in the presence of a robot that suddenly animates and “comes into existence.” Thereby, it imitates the interaction opening delays regularly observed in laboratory or “in-the-wild” human-robot interaction studies, where robots can require time before springing to life after they stand in co-presence with a human. We will show how this is consequential with respect to the way in which openings unfold, in which some moves such as greeting and waving may become interactionally relevant, and, ultimately, in which the robot emerges as a social agent. We conclude that pre-beginnings are not anecdotal or peripheral issues with respect to HRI experiments but should be thought about and designed as an integral part of those.

## 3 METHOD

### 3.1 Analytic Methodology

*3.1.1 Ethnomethodology and Conversation Analysis.* **Ethnomethodological Conversation Analysis (EMCA)** [22, 58] is a micro-sociological approach that studies the temporal unfolding of events in an interaction to understand, on a moment-to-moment basis, how participants’ “actions—turns-at-talk, embodied actions, or complex multimodal moves—are produced and recognised as performing meaningful social actions” [6, 96]. EMCA offers methodological tools for understanding what is treated as publicly relevant by participants in a given situation—among the

potentially inexhaustible number of features of this situation describable from an external perspective—without imposing the researcher’s perspective on the data: It adopts an emic point of view rather than an etic point of view [96]. In sum, EMCA will consider “as phenomena only those practices of members which are used by them to produce, accomplish, sustain, reproduce, recognize, and give account of, to and for themselves, social order” [70].

This angle of analysis treats social order as a product of the local organization of participants: that is, as continuously maintained (or modified) through their local accomplishments [22]. EMCA multimodal approaches rely on the minute analysis of video recordings [29] and of their detailed transcriptions to identify the fine-tuned temporal unfolding [57] of participants’ actions with a much higher degree of precision than mere observation would allow [57]. The in-depth analysis of large collections of data can subsequently reveal recurring patterns in the procedures and methods through which social order is accomplished [28, 66].

**3.1.2 EMCA for HRI.** When applied to HRI more specifically, EMCA is suited to identify what, in a robot’s multimodal behavior (its talk, gestures, flashing LEDs, sounds signals, motor noise, etc.) [61] is treated by involved participants as *social actions*: that is, as actions that make relevant “a set of potential next actions” [96]. This becomes apparent when these behaviors from the robot are responded to (and in a certain way) by co-present participants, in the following turns-at-talk. In other terms, EMCA’s micro-analytic level of description can be leveraged to explore what emerges as pragmatically consequential when humans’ and robots’ actions intertwine and respond to each other (often in ways unforeseen by designers [60]), i.e., to retrospectively unpack how humans and robots co-constructed the interactions [68] that ended up being captured on camera. Therefore, researchers drawing on EMCA will focus on behaviors produced by a robot that are made publicly recognizable and accountable (as an action of a certain type, e.g., an offer, a question) by co-present humans, as these humans are immersed in a situated activity, where they face specific practical problems [31]—e.g., whether to respond to a robot producing a “waving gesture” at the start of an interaction and, if so, how to respond to it. In this sense, the methodological tools of EMCA can be mobilized by researchers faced with the issue of analyzing robots’ and humans’ micro-adjustments over time as something else than an inscrutable “black box” [31].

**3.1.3 Studying the “Sociality of Robots” Besides Mental Representations.** A result of this approach is to study the “sociality of robots” independently from mental representations. Our analytical focus will be exclusively limited to what is oriented to by participants as an “action” from the robot, i.e., the way in which some of its “behaviors” are responded to or, alternatively, what actions from the robot are accountably displayed as absent by humans. In particular, an ethnomethodological and conversation analytic methodology does not aim at establishing how the robot is otherwise mentally represented [26] by the participants (as a social agent, as an object, as a human, as an animal, as a new ontological category...) or if they *engage in pretense towards* it [87]. Unless it is made observably accountable, it makes no difference for the status occupied by the robot in the interaction that participants are “behaving as” or “behaving as if” this robot is a “subject with internal states and perceptual experiences” [87]. Similarly, this analytical perspective is independent from the set of questions related to whether participants’ actions involve a mental representation of the robot as they interact with it, or if these behaviors are produced as part of a non-representational “mindless coping” with the situation [14].<sup>2</sup> In other words, EMCA takes an agnostic stance regarding cognition [37, 54], as it is interested in participants’ (humans or robots) “observable and hearable conduct [...] at the interactional surface” [66].

<sup>2</sup>In particular, we do not intend to suggest that participants discretize and categorize the stream of conduct of the robot into a preexisting list of action types [16]: The robot can be positioned interactionally as an agent without any of its behaviors being mentally constituted as “actions” by participants as they react to them.

### 3.2 Participants

We base our analysis on 80 video recordings of dyadic interactions with an autonomous robot, which took place at the INSEAD-Sorbonne Université Behavioural Lab. Participants were all native French speakers aged between 18 and 30 years old. All participants were recruited by the INSEAD-Sorbonne University Behavioural Lab under ethics approval by the INSEAD Institutional Review Board. Separate consent was obtained for the use of video data. The experiment took on average 20 minutes to complete, and each participant received a compensation of 6 €.

### 3.3 Experimental Setup

A humanoid robot, “Pepper,” produced by Softbank Robotics, was positioned in the middle of a room, standing at a three-quarter angle from participants when they entered by the door (see Figure 2). The interaction was filmed with two cameras: one behind the robot, one on the left of the robot. An additional webcam was placed in a corner of the room. For a detailed description of our experimental setup and of the design of the autonomous robot, see Reference [93].

### 3.4 Instructions

Before entering the room, all participants were given the same verbal instructions:

- (1) “You are going to have an interaction with a social robot. This robot will try to help you plan your holidays, for this summer. Please answer as if you were really planning these holidays.”
- (2) “Speak loudly. If the robot does not respond, it is possible that it didn’t hear you. If you see a question mark displayed on the robot’s tablet, it means your utterance wasn’t understood: you can repeat or rephrase it.”
- (3) “The experiment should take 5 minutes to complete, then you will have to fill in a questionnaire in the next room.”

Then, as they entered to room, participants were informed that:

- (4) “The robot should start speaking to you in a few moments.”
- (5) “You can stand anywhere in the room.”

This characterization of the robot and of the task partially pre-configured the interaction. They created the expectation for an incoming, but delayed, first turn uttered by the robot. Doing so, they portrayed the robot as an entity that may not immediately be available as a co-interactant and potentially “come into existence” at some point. They also stated the robot may not hear the participant, depicting its perceptual abilities as imperfect. For these reasons, they should be treated as constitutive of pre-beginnings. In sum, the distribution of greetings (Figure 1) on which we will focus is not to be understood as a direct reflection of the strength of a transsituational (greeting) norm but, rather, as connected to a specific experimental configuration.

### 3.5 Scenario

The robot was designed as a “travel agent.” Once participants had entered the room, the experiment followed a “holiday planning scenario”: The Pepper robot “woke up” by going through several “activation steps,” introduced itself, produced a “how are you” question, offered to take water, and then asked participants several questions aimed at understanding their preferred destinations. When the scenario reached its end, participants moved to a different room and completed a questionnaire composed of several psychometric scales.

All participants studied in this article faced the same initial behavior from the Pepper robot. Because our focus is on the earliest moments of the interaction, the different conditions in which

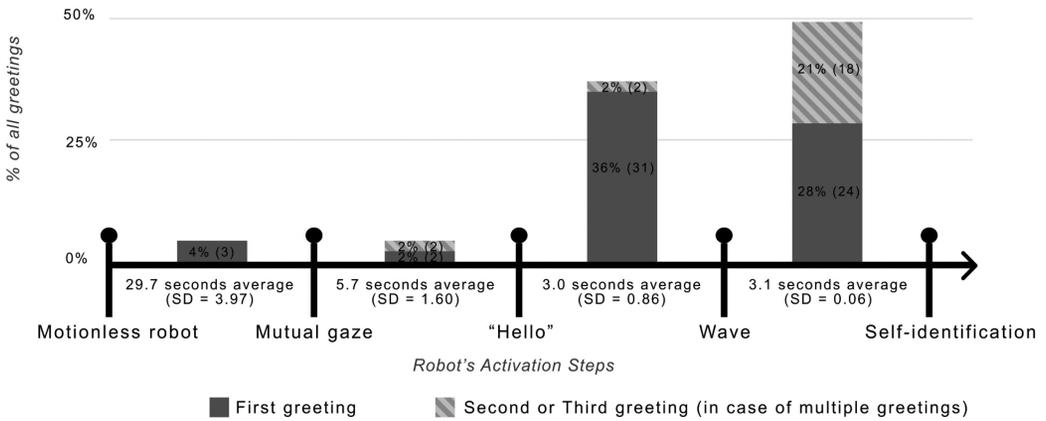


Fig. 1. Distribution of all greetings produced by participants between “activation steps,” including multiple greetings by the same participants. Total first greetings = 62, total greetings overall = 85.

these participants were placed did not impact the multimodal behavior of the robot yet. However, as part of a larger study [93], participants were distributed in five experimental conditions, each one featuring a different multimodal behavior from the robot later in the interaction (no social gaze, no approach, etc.). 101 valid participants took part in this experiment. In one of our experimental conditions, the robot did not wave during the opening of the interaction: These 21 participants were removed from our analysis (since they could not possibly react to the robot’s wave), leaving 80 remaining participants who all witnessed the same “activation steps” from the robot.

### 3.6 “Activation steps” Achieved by the Robot during the First Seconds of the Interaction

Immediately after each participant entered the room, the robot went through the same five “activation steps” (cf. Figure 1 for a detailed timeline):

- (1) *Physical co-presence*: When participants entered the room, the robot was motionless.
- (2) *Gaze tracking*: The robot started to track their gaze.<sup>3</sup>
- (3) *Greeting*: The robot uttered a “bonjour” (“hello”).
- (4) *Wave*: The robot produced a waving gesture.
- (5) *Self-identification*: The robot self-identified and introduced its role as a travel agent.

These steps exacerbated two features of the “coming into existence” often observed in natural or controlled human-robot interactions openings: The robot stood in physical co-presence with the participant for several seconds before a reciprocal interaction could start, and it displayed no preexisting activity when first appearing to this participant.

## 4 DISTRIBUTION OF FIRST GREETING OCCURRENCES DURING THE EXPERIMENT

Out of a total of 80 participants, 62 (78%) produced a greeting utterance (e.g., “Hi, Hello, Hey, Good morning” [65]) or gesture (e.g., “hand wave, palm display, head toss/bow, eyebrow flash” [65]). A micro-analysis reveals that most of these utterances and gestures corresponded to *greeting actions*,

<sup>3</sup>To do so, the robot had to produce a vertical head tilt. This was connected with two “consequential sounds” [94]: the sound of motors and gears (required to tilt the robot’s head up) and the squeaking plastic.

Table 1. Summary of First Greeting Occurrences

Activation step	First Greeting Occurrences	Description	Example
Physical co-presence	3 participants (5%)	Participants initiated a first greeting immediately after entering the room where the robot was placed	Fragment.1
Gaze tracking	2 participants (3%)	Participants' first greeting occurred after mutual gaze was established with the robot	Fragment.2
Greeting	31 participants (50%)	Participants' first greeting immediately followed the "hello" uttered by the robot	Fragment.3
Wave	24 participants (39%)	Participants produced a first greeting immediately after the wave achieved by the robot	Fragment.4
Post self-identification	2 participants (3%)	Participants' first greeting occurred at a later point during the interaction	
None	18 participants (22%)	Participants never produced any form of greeting	Fragment.5

based on the definition from Reference [65]: “discrete audible and visible (vocal, verbal/lexical, and embodied) actions that participants deploy to publicly mark the moment when they ratify another’s social copresence.” As the following fragments will illustrate, this does not imply the actions participants achieved through these utterances and gestures were limited to “greeting the robot.”

Among participants who produced a greeting utterance or gesture, the events or robot’s behaviors (“activation steps”) that immediately preceded the production of their *first greeting* were distributed as displayed in Table 1. The average delay<sup>4</sup> between these “activation steps” (cf. Figure 1) was long enough to prevent misattributing which “activation step” preceded participants’ greeting utterances or gestures.

Most *initial* greeting utterances or gestures, therefore, occurred after the robot’s own verbal greeting. However, overall, most greeting behaviors were produced after the robot’s wave, as many participants—among those who had previously greeted the robot—produced a *new greeting* after the robot achieved this gesture (cf. Figure 1).

## 5 FIVE PATHS TO THE PRODUCTION OF A FIRST GREETING

The five fragments below, ordered chronologically relative to the robot’s “activation steps,” are representative of this corpus.<sup>5</sup> They are analyzed using an **ethnomethodological and conversation analytic methodology (EMCA)** to reveal, on a moment-to-moment basis, *what interactional processes are aggregated in the statistical distribution presented above*. That is, which events or actions were typically oriented to by the human participant as constructing an appropriate

<sup>4</sup>The presence of a measurable standard deviation with this delay between “activation steps” is due to variations in the robot’s CPU load between participants.

<sup>5</sup>Each fragment displays the most common way in which a greeting emerged for each “activation step”—and a typical case in which greetings did not emerge for fragment 5. However, very few greetings occurred after the first activation step (Motionless robot) and the second activation step (Mutual gaze)—cf. Figure 1. In this sense, fragments 1 and 2 display rare occurrences in comparison with all 80 participants analyzed. Note that, even though these fragments were chosen as the most representative, some specificities of these participants’ behavior (i.e., the distance at which they were standing from the robot) are still idiosyncratic and different from the average.

framework for the production of greetings. Our transcription conventions are detailed in the Appendix.

## 5.1 Orient to Physical Co-presence as an Adequate Framework to Initiate a Greeting Sequence

### 5.1.1 Fragment 1.

1. \* (2.5) \*£ (3.0) \*(0.1)  
 hum \*>>places coat on chair\*moves in front of robot\*  
 hum £gazes at robot->
2. HUM hello peppeur?  
*hello Pepper*
3. (3.6)£# (3.8) £(2.4)£ (1.3) £+(1.2)  
 hum £tilts head on the side£ £gazes at camera£gazes at robot->  
 hum +smiles->>  
 img #img.1.1

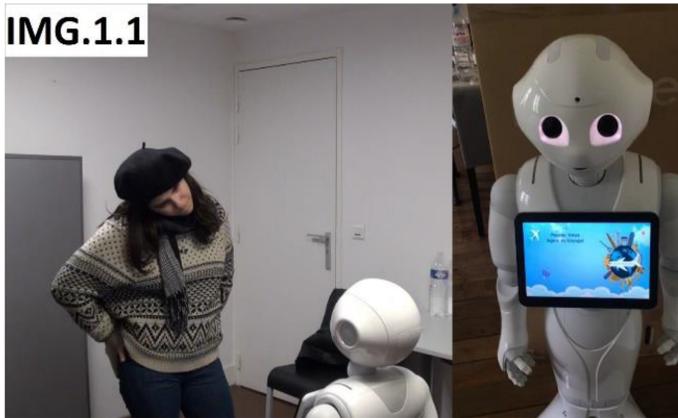


Fig. 2. Image 1.1 – Participant tilts her head after the robot does not reply to her greeting.

4. ROB ((motor and plastic sounds))%\$  
 rob %gazes at human->>  
 rob \$arms are shaking->>
5. (3.5)+ (0.5) +(1.0)  
 hum +raises one eyebrow+
6. HUM hello peppeur.  
*hello Pepper*
7. ROB =bon\$jour  
*hello*  
 rob \$opens its arms->

8. (1.1)\$(0.8)\$(0.3)\*(0.3)\*(0.6)\$(0.9)\*(0.5)  
 rob ->\$.....\$-----waves-----\$ , , , ->  
 hum \*.....\*---waves---\* , , , ->

9. **ROB je \$\* m'appelle peppour**  
**my name is Pepper**  
 rob ->\$  
 hum ->\*

5.1.2 *Description and Analysis.* After laying her coat on the chair, the participant gazes at the robot and positions herself in front of it, facing it (L.1). Standing at 0.8 meter from the robot, she is comparatively close to it in regard to other participants (average initial distance = 1.3 meter; SD = 0.26). She immediately initiates a first greeting pair and an address term (“Pepper,” L.2) and maintains her gaze, body orientation, and posture—slightly leaning towards the robot—during the next 3.6 seconds. After this standstill, she produces a lateral head tilt for 3.8 seconds (L.3), while maintaining her gaze towards the robot’s eyes. A few seconds later, the robot directs its face towards the human’s eyes, establishing mutual gaze. This quick adjustment of the robot’s head is associated with motor noise and plastic sounds (L.4). These so-called “consequential sounds” are known to be regularly oriented to by participants in human-robot interactions [20, 85, 94]. The robots’ arms also start shaking lightly (L.4), which will persist until the end of the experiment. After 3.5 seconds of this mutual gaze, the participant raises one eyebrow (L.5) and, after a second of silence, produces a new greeting (“hello Pepper,” L.6). This action appears to account for the robot’s continued silence and, in particular, to orient to the robot’s alignment with the human’s gaze as creating expectations for a next action on its part. However, the participant’s greeting is immediately followed by a greeting from the robot (L.7). This implies that, in this fragment,<sup>6</sup> the robot’s greeting turn is sequentially positioned as a second pair part *responding* to the participant greeting—who, indeed, does not achieve a new greeting in return (L.8), suggesting that she orients to the robot’s “*bonjour*” as a reaction to her own. After a short pause, the robot then starts to produce a greeting wave (L.8), with which the participant aligns by producing a similar wave. Once the robot starts to retract this waving gesture, the participant immediately begins to lower her own arm and finishes her retraction simultaneously with the robot.

The previous fragment constitutes one of only three openings in our corpus where a greeting is initiated by the participant before the robot was activated (from an etic perspective). All along the interaction, the robot is constructed as a co-present interlocutor, even when it is not moving yet. Besides the “*bonjour*” that the robot produces after her greeting, the participant orients to every “activation step” displayed by the robot (physical co-presence, mutual gaze, and wave) as opportunities to produce a first greeting. Thus, as soon as she is positioned in front of the robot, she treats the situation as a “framework in which a greeting sequence is relevant” [59] and the robot as able to react to the production of a greeting: by tilting her head<sup>7</sup> (L.3) and by raising

<sup>6</sup>That is, independently from the intended design for the script followed by the robot at the beginning of the interaction. Indeed, this first greeting was originally intended as a greeting initiation, not as a response. There is obviously no systematic overlap between the features of the local situation that are practically oriented to by participants as relevant in an interaction with a humanoid robot and those that were designed as *a priori* relevant during the design of the autonomous robot’s behavior [60].

<sup>7</sup>The lateral head-tilt she produces (L.3) may be understood as an embodied account of the non-response of the robot, in the same way they can be used as an embodied display of trouble in classroom interactions [2,86].

one eyebrow (L.5), she makes accountable the non-answer of the robot to the first pair parts she produced and she displays expectations for a reaction.

Remarkably, even though the participant's last verbal greeting (which took place immediately before the robot's vocal greeting) ultimately positioned her as having "initiated" the greeting sequence (L.6 & 7), she instantly re-positions the robot as the "anchor" [83] or, more generally, as the speaker initiating new sequences. Indeed, after the robot greets her back, the participant stays silent (L.8) and does not use her position as a first speaker to self-select [76] and to initiate a new sequence. This silence by the participant leads the robot to produce an "interlocked" turn [82] that combines both its response to her "hello" and its initiation of a greeting wave (L.8). By staying silent, the participant thus provides the robot with the adequate position to initiate subsequent sequences [82] and, later, to initiate the topic of the interaction (not in transcript).

## 5.2 Orient to Mutual Gaze as Projecting an Imminent Next Action from the Robot

### 5.2.1 Fragment 2.

1. \* (1.7) £\* (2.0) \*£# (0.7) £
- hum \*>>closes bag\*moves in front of robot\*adjusts clothes->
- hum £ gazes rob £gazes clothes£gazes rob->
- img #img.2.1



Fig. 3. Image 2.1 – Participant adjusts his clothes after positioning himself in front of the robot.

2. **ROB** ((motor and plastic sounds))%\$
- rob %gazes at human>>
- rob \$arms are shaking->>
3. (0.1)\*(0.6)
- hum ->\*
4. **HUM** .tsk .h
5. £(0.5)
- hum ->£gazes at own sleeve->



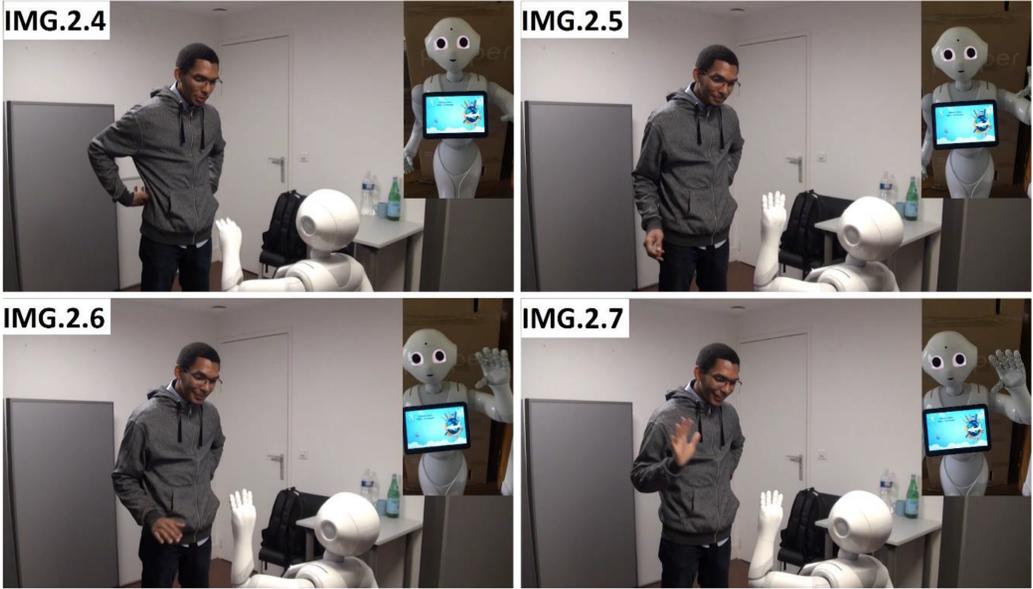


Fig. 5. Images 2.4 to 2.7 – Participant starts to extend his hand towards the robot, then modifies the trajectory of his hand to align with the robot’s waving gesture.

```

12.HUM      .tsk f*+ (.)#hello
              hello
hum         ->fgazes at robot’s face>>
hum         ->*waves->
hum         +smiles->>
img         #img.2.7
13.         (0.5)*(0.8)
hum         ->*
14.ROB      je m’appelle peppour
              my name is Pepper
rob         ->$

```

5.2.2 *Description and Analysis.* After positioning himself in front of the robot, the participant does some self-grooming [42] as he readjusts his clothes (L.1). Standing at 0.8 meter from the robot, he stands closer to it than most participants. After the robot orients its face towards his eyes (L.2), with the matching motor noise and squeaking plastic sounds, the participant gazes back at it and produces a tongue smack (“tsk,” L.4) followed by a first greeting (“hello,” L.6), interrupting his self-grooming (L.5). He then takes a step forward while staring at the robot and, after 2.1 seconds of silence, takes a step back to his original position (L.7). After 2.8 additional seconds of silent mutual gaze, the robot utters a greeting term (“bonjour,” L.8) as the participant just finished taking his arm outside of his pocket. The participant briefly extends his hand toward the robot (L.9), then retracts

it and produces a return greeting (L.10). The robot starts to raise its arm to prepare a waving gesture (L.10, image 2.6). When this gesture has not reached its apex yet, the participant starts to extend his own hand towards the robot (L.11, images 2.7 and 2.8); however, once the robot's arm becomes fully extended and starts the waving motion, the participant redirects his arm to produce a waving gesture (L.11, image 2.9). He simultaneously achieves a new verbal greeting and smiles (L.12).

In this fragment, physical co-presence is not treated as sufficient to initiate the interaction—unlike the previous example. The robot is not oriented to as a conversational partner from the very start. This is especially visible through the production of “self-grooming” by the participant, usually displayed during the approach between two interactants [30]. However, the status of the robot in the interaction shifts after the establishment of mutual gaze. The participant's interruption of his self-grooming (L.5), his greeting (L.6), and the step forwards he takes (L.6) accentuate a shared inner space [42] and display the expectation of an imminent action from the robot. This reconfiguration results from the “crucial analytic distinction” [59] made by the participant about what the gaze from the robot is projecting: It is not oriented to as a merely automatic “gaze tracking,” nor as a “mere look” [45, 59], but as a look projecting the initiation of an upcoming action. The participant's expectation is not met, however, as he goes back to his original spot. Mutual gaze, therefore, constitutes the first “breaking point” after which the robot becomes (momentarily) present as a potential interlocutor. As Reference [66] notes in the case of young children interacting with a toy robot, even “little sequential phenomena of the robot's timely conduct” in relation to participants' actions can have a critical impact on the “categorization and re-interpretation” [66] of this robot. In our fragment, the redirection of the robot's gaze towards the participant after a silence is treated as a *meaningful social action*.

Incidentally, two reconfigurations could be observed in this participant's gestures. First, L.9, he extends his hand towards the robot immediately after its first verbal greeting, before retracting this hand and producing a return “*bonjour*.” The cancellation of his tentative gesture and, instead, his production of a verbal greeting, appear to constitute alignments with the robot's (then verbal) mode of greeting. Later, as the robot starts to visibly raise its arm as part of its waving gesture, the participant's response gesture shifts from an apparent “handshake” gesture to a clearly observable wave (L.11 to L.12; images 2.4 to 2.7). These two episodes display quickly evolving interpretations of what action the robot is projecting during its first greeting and, later, during the preparation of its waving gesture. It highlights an online monitoring of the robot by the participant [68], which allows him to reconfigure his embodied course of action to align with the robots' co-occurring action.<sup>8</sup>

### 5.3 Multiple Greetings – Orient to the Robot's Wave as the Confirmation of an Ongoing Greeting Sequence

#### 5.3.1 Fragment 3.

#### 1. #f\*(4.1)

hum	f>>looks around the room->
hum	*>>swings from one leg to the other->
img	#img.3.1

<sup>8</sup>Of course, besides tracking the participant's gaze, our autonomous robot's behavior could not itself be reconfigured by co-occurring actions from the participant: From an etic point of view, there was no “loop of mutual adjustments” [69] to speak of.

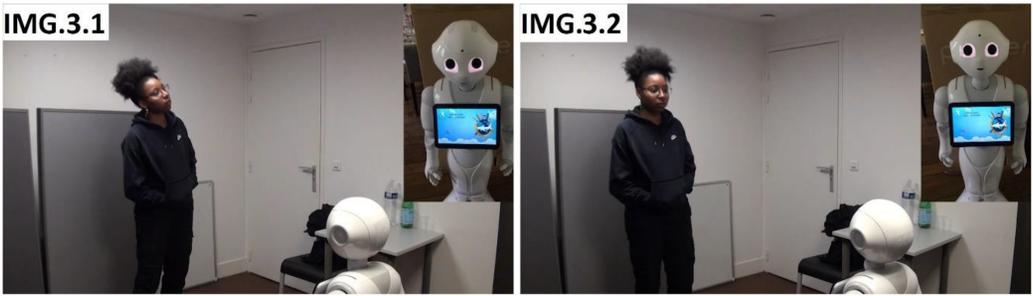


Fig. 6. Images 3.1 and 3.2 – The robot gazes at the participant, she stops her swinging movement and establishes mutual gaze.

```

2. ROB      ((motor and plastic sounds))%$
rob                                     ->%gazes at human->>
rob                                     $arms are shaking->>

3.          (0.1)£+ (1.5)*#(2.4) +
hum        ->£gazes at robot->
hum        +raises eyebrows+
hum        ->*
img        #img.3.2

4. ROB      bonjour
           hello

5.          (0.1)$ (3.5)
rob        $opens its arms->

6. HUM      °bonjour°?$(.)
           hello?
rob        ->$waves->

7.          (0.4)£ (1.0) £(0.4)
hum        ->£gazes at wave£gazes at robot's face->>

8. HUM      #bonjour,+# (.)
           hello
hum        +widens her smile->>
img        #img.3.3 #img.3.4

```

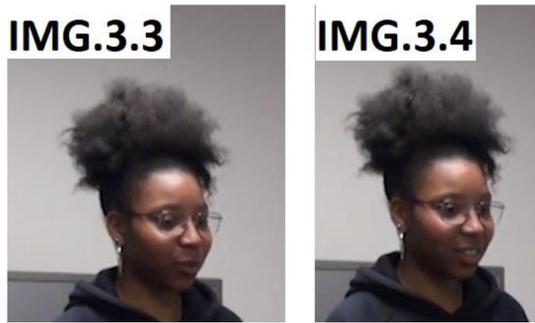


Fig. 7. Images 3.3 and 3.4 – Participant’s widened smile when uttering her second “hello.”

9. ROB            **je m’appelle\$ peppeur**  
                       *my name is Pepper*  
 rob                                ->\$

*5.3.2 Description and Analysis.* After entering the room, the participant positions herself in front of the robot, at a higher distance than most participants. She starts to swing from one leg to the other while looking around the room (L.1). After a few seconds, the robot gazes at the human’s face and, doing so, produces motor noise and squeaking plastic sounds (L.2). The participant instantly stops her swinging movement and initiates mutual gaze with the robot, while raising an eyebrow (L.3). She maintains this posture for the next 4 seconds of mutual silence, and even after the robot utters a first greeting (“*bonjour*,” L.4). Because of a momentary processor overload, the robot’s waving gesture takes 3.7 seconds to be triggered after its “*bonjour*,” unlike the rest of the corpus where it took on average 3 seconds. After 3.5 seconds of silence, possibly orienting to the silence and the stillness of the robot as offering a response slot, the participant softly whispers a first greeting (L.6) with a rising vocal pitch—right before the robot starts its wave. Once the robot initiates its waving gesture (L.6), the participant glances towards its waving arm (L.7) then utters a new greeting (“*bonjour*,” L.8). This new greeting is uttered out loud and with a final continuing intonation, while the participant widens her smile (L.8, image 3.4).

In this fragment, the participant orients to the first greeting of the robot as sequentially equivocal [35]. Her first greeting (whispered, delayed, with a rising pitch) displays uncertainty regarding what the robot’s greeting is projecting: The robot’s “*bonjour*” is not oriented to by the participant as clearly constituting the first part of an adjacency pair that should be completed by a return greeting. The design of her first greeting thus appears to question the status of the interaction—and even the existence of a “stepwise process of mutual adjustments” [67]. Conversely, the participant’s second greeting turn appears to confirm “what is going on” [6] as an “exchange of mutual greetings”: Uttered out loud and simultaneous with a widened smile, it orients positively<sup>9</sup> to the robot’s gesture as initiating a second greeting sequence.

This supports an interpretation where each greeting produced by the participant accomplishes a different task [55]. The first greeting, uttered 3.5 seconds after the robot’s own greeting, mainly checks the availability of the robot and its ability to perceive and react to the human’s relevant actions<sup>10</sup> —a form of “device testing” [71]—whereas the second greeting is a clear ratification of

<sup>9</sup>Smiling, which is “a principal way parties do ‘displaying a positive stance’ toward encountering recipients” [65] suggests the situation is now being treated as the beginning of a socially co-present encounter.

the start of the co-present interaction: It is a “sociability practice” [55]. Consequently, we observe a form of *inertia* in this fragment: The inanimate object that the robot is first oriented to requires interactional work (lasting over several seconds) to be replaced by a conversational agent. The first greeting term produced by the robot does not immediately nor automatically institute it as a conversational partner that can be greeted back.

#### 5.4 Orient to the Robot’s Waving Gesture as an Upgrade of Its First Greeting

##### 5.4.1 Fragment 4.

1.           £           (11.0)       £(0.2)  
hum       >>£looks around the room£gazes at robot->>
2. ROB       ((motor and plastic sounds))%\$  
rob                               %gazes at human->>  
rob                               \$arms are shaking->>
3.           (6.2)
4. ROB       \$bonjour  
              hello  
rob        \$opens its arms->

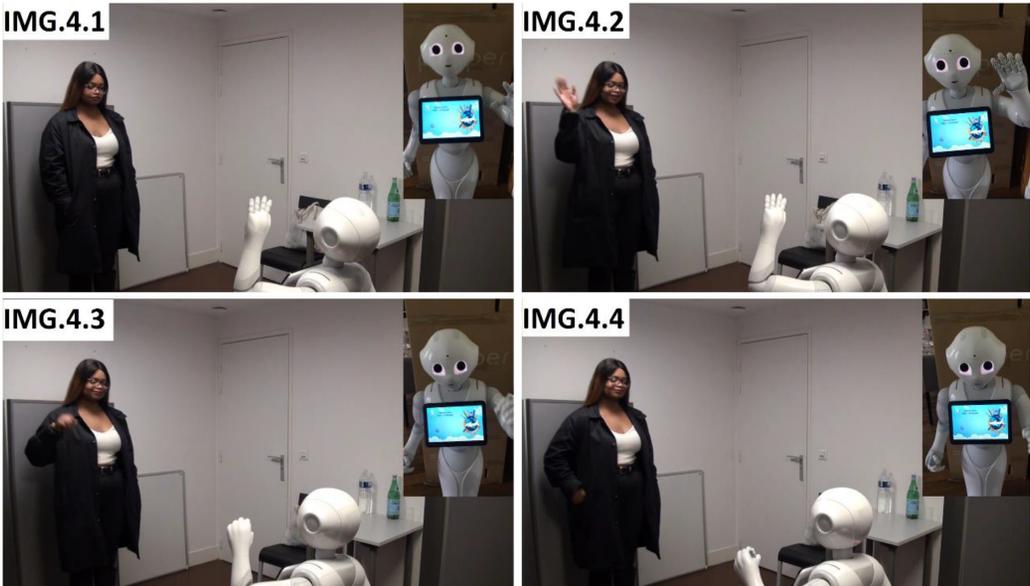


Fig. 8. Images 4.1 to 4.4 – The participant catches up with the robot’s waving gesture, then retracts her gesture immediately after the robot retracts its arm.

<sup>10</sup>This participant can be described as verifying whether the entity in front of her is “an animate object that is able to engage with her in re-occurring interactional patterns” [66].

5. (1.1)\$(0.8)\$(0.4)#  
 rob ->\$.....\$-waves->  
 img #img.4.1
6. HUM \*bon\*+jour:#  
*hello*  
 hum \*...\*-waves->  
 hum +smiles->>  
 img #img.4.2
7. \$(0.2) \*# (0.6) \*#(0.1)  
 rob ->\$,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,->  
 hum ->\*,,,,,,,,,,,,,,\*  
 img #img.4.3 #img.4.4
8. ROB je m'appelle\$ peppour  
*my name is Pepper*  
 rob ->\$

5.4.2 *Description and Analysis.* The participant positions herself in front of the robot, facing it. However, standing at 1.8 m from it, she is among the two participants who positioned the farthest. She looks around the room (L.1) for a few seconds, then switches her gaze to the robot. At the same time, the robot's head aligns with the participant's eyes, which triggers motor and plastic noises. This pose is silently maintained for the next six seconds by the participant and the robot (L.3). Then, the robot produces the greeting term "*bonjour*" L.4 while opening its arms, with its palms facing the ceiling. The participant produces no response, whether vocal or gestural. After 2.3 seconds of silence, the robot initiates a "wave" (L.5). When the wave reaches its apex (i.e., the robot has fully raised its hand), the participant produces a first verbal greeting ("*bonjour*," L.6) and a smile at the same time as she initiates a waving gesture—fast enough to catch up with the robot's own gesture. The participant's wave stays synchronized with the robot's wave, then stops immediately after the robot starts retracting it (L.7): 200 milliseconds after the robot's arm starts to lower, the participant also starts to retract her arm—as in fragment 1. Like the overwhelming majority of our corpus, this participant's gaze focuses on the robot as soon as it moves its head to track her gaze—but she does not immediately produce a speaking turn. The lasting silence, mutual gaze, and "consequential sounds" are not oriented to as initiating a "slot" where to self-select. Even after the robot utters a "*bonjour*," she returns no greeting and maintains her previous pose and gaze for the next few seconds.

However, once the robot starts a waving gesture, the participant silently observes its arm rise during the action's preparation. She then abruptly produces her own wave—which catches up with the robot's gesture—and simultaneously produces a smile and a verbal greeting (L.6). The speed of this return wave seems to indicate that the participant orients to the robot's gesture as, either, producing a normative obligation to achieve a return greeting, or, alternatively, as upgrading a normative obligation to respond that she would have previously failed to observe. In particular, based on the numerous occurrences of this situation in our corpus, we suggest that, in this fragment, the participant's hasty first greeting displays her alignment as normatively expected at an earlier point in the interaction. That is, she orients to the robot's wave as a





**11. ROB**      **je m'appelle\$£ peppeur**  
                   **my name is Pepper**

rob                    ->\$

hum                   ->£

*5.5.2 Description and Analysis.* After entering the room and dropping her bag on the ground, the participant gazes at the robot (L.1). As she starts approaching it, she produces a speaking turn involving a deictic reference to the robot as “this thing” (“*ce truc*,” L.2)—referring to it in the third person—and qualifies it as “weird.” This comment displays most of the typical properties of “self-talk” [41]: It is achieved while the participant is leaning forward, her body not oriented towards the robot, and part of it is uttered while looking at the ground, in a low voice. As a consequence, the participant does not manifest any expectation for an answer; her comment does not open a “conversational sequence” [41]. Once her approach is complete, the participant stands at 1 m from the robot, closer than the average initial distance of 1.3 meter for all participants. After a silence of 3.6 seconds, the robot shifts its gaze towards her eyes—doing so, it produces soft motor noises and squeaking plastic sounds (L.4). The participant reacts with a short laugh (L.6). After a new silence of 4 seconds, the robot initiates a first greeting (L.8) and opens its arms—palms towards the ceiling. The participant produces another laugh, more audible and longer than her previous one (L.10). This laugh is continued during most of the hand wave of the robot and is followed by an in-breath when the robot’s arm starts to retract (L.10).

The actions of this participant highlight a double orientation to the robot, as both normatively neutral and unable to react to human actions. First of all, she treats the robot as an autonomous script whose verbal utterances imply no normative obligation to be responded to, even after it produces a greeting term: No conditional relevance emerges from the behaviors produced by the robot in the course of the interaction. Yet, simultaneously, the participant orients to the robot as unable to respond to (or to perceive) her own actions. The “self-talk” (L.2) or laughs (L.6 and 10) she produces in front of the robot do not manifest any expectation for an answer, and the absence of reaction from the robot is not visibly made accountable (unlike fragment 1). Her turns are not “recipient designed” to be registered as “inputs” in response to which the robot would produce, reconfigure [25], or interrupt speaking turns. In other words, the robot is never “characterized as being able to perform reciprocal interactions” [91], which is a prerequisite for the existence of a “social encounter” [91]. Neither the robot nor the human imposes a normative order on the other: There is no observable “sequence organization” [82] that exerts a constraint on their actions. Using the previously mentioned terminology from Reference [37], this participant is merely “using speech” but not “talking” with the robot, in the sense where “talking” would imply to produce “actions that are discoverable within a normative order” [37] and to assume “other participants to be able to perceive them as actions within that order” [37].

Additionally, no facework is achieved by the participant as, in particular, her comments are not fully whispered (L.2) and her laughs are achieved audibly and visibly while standing right in front of the robot. Her first utterance (“this thing is really weird”) also explicitly characterizes the robot as an object and refers to it in the third person. As a whole, semantic content and sequential organization mutually reinforce to characterize and position the robot as a non-agent: The participant establishes herself as the spectator of a pre-recorded monologue, whose performer is not socially present with her in the room.

Last, we note that even though this participant approached the robot more than average (i.e., stood at less than 1.3 meters) this unusual proximity was part of a sequence where she scrutinized and commented on the robot, treating it as an inanimate object instead of an interactant.

This is unlike participants in fragments 1 and 2—even though they also stood unusually close to the robot—for whom this proximity displayed a treatment of the robot as an ongoing (fragment 1) or imminent (fragment 2) interactant. This connects with the general observation that, on an individual level, the distance at which a participant stood from the robot was meaningful only in connection with a sequential context.

## 6 DISCUSSION

### 6.1 Sequential Ambiguity

The previous fragments reveal the varied interactional work required before “behaviors” from the robot could be treated as “actions” that either (1) established the adequate framework for the participant to initiate a first greeting sequence or (2) produced a response slot that the participant was normatively pressured to complete with a return greeting. Even though, after they entered the room, all participants positioned themselves in front of the robot to form a vis-a-vis arrangement [36, 42], we see that the mere utterance of a greeting term from the robot (“hello”) did not automatically and immediately establish a reciprocal interaction. There was a regular delay in the shift from the robot as a normatively neutral artifact (which was discovered completely motionless at the very beginning of the interaction) to a conversational partner producing sequentially implicative turns. For many participants (exemplified in fragments 3, 4, and 5), the interactional status of the robot *persisted* after it produced a first greeting.

An explanation for this delayed emergence of the robot as an agent is that participants found themselves confronted (and prepared to be confronted by the instructions described in Section 3.4) by an inert robot that suddenly animates and “comes into existence” and were therefore engaged in sequentially ambiguous situations [35] as to what actions the robot was projecting (or if it was projecting anything): They had to “entertain the full range of possibilities momentarily, using the immediately following talk to find out what sort of sequence is in progress” [80]. During fragments 1, 2, and 3, participants’ initial turns can be considered as practical attempts to “probe” the current status of the interaction: By trying to elicit a response from the robot, these actions clarified whether a phase of mutual adjustments, or any form of turn-based coordinated activity, was either ongoing or technically feasible.

We argue that these face-to-face encounters with a humanoid robot “coming into existence” disrupted “background expectancies and methods at play in the accomplishment of commonplace activities, such as having a conversation” [88]. Participants had to achieve greetings in a situation marked by otherness, which, like it was observed in different contexts, “throw[s] the greeters and their practices of greeting into crisis” [58]. This experiment thus made especially apparent the constant “experiments in miniature” [23, 48] achieved by humans when interacting with robots (and, of course, with other humans), where each action “tests the hypothesis a participant has about a co-participant’s response to her/his action.” In particular, in these human-robot interactions, participants faced the challenge of (1) identifying if intersubjectivity [48] was even possible (i.e., if the entity in front of them possessed the required properties for mutually achieving a “reciprocity of perspectives” [48]) and, then, of (2) establishing this intersubjectivity—for example, by producing actions that displayed, and therefore tested, an orientation to the previous robot’s turns as opening a greeting sequence. We suggest that this double challenge is a common trait of first encounters with humanoid robots, which may require the use of different resources to be overcome, depending on the way in which the robot is first encountered by the human.

### 6.2 The Waving Gesture as a Threshold

The robot’s wave was often critical in clearing up this “sequential ambiguity” [35]. In several cases (although, not systematically) this gesture offered a practical answer to the practical issue

participants were encountering, namely, to document “what is going on” within a given spate of talk [6]. For example, in the specific sequential contexts presented in fragments 3 and 4 (i.e., not a wave “as such,” discretized and disconnected from local situations<sup>12</sup>), the wave functioned both as a clarification of the situation—as an ongoing greeting sequence—and, simultaneously, as a soft upgrade of the conditional relevance of the previous greeting turn produced by the robot (“hello”). In other terms, it manifested the normative obligation to produce a return greeting and retrospectively oriented to the participant’s non-response (or non-proper response) to the robot’s first vocal greeting. Therefore, in a similar way to the responses [60] observed after a two-part animation<sup>13</sup> of the robot Cozmo, the—etically designed—“two part greeting” achieved by our robot (vocal greeting, pause, wave) often led participants to reconsider their past actions. To paraphrase Reference [16], there was an observable evolution in the “multiple drafts” these participants produced of the situation, as the robot achieved a waving gesture.

In sum, more than any other of its behaviors, the robot’s wave was frequently responded to as a conditionally relevant first greeting pair. The instants following this wave constituted a frequent (and momentary) threshold between the robot oriented to as a “raw physical artifact” [10] (of plastic, sensors, etc.) and the robot treated as a socially co-present entity—whose actions could establish “a set of normative constraints on the type and form of action with which the recipient should respond” [44]. The wave regularly interrupted the persistence of the status of the robot as a non-agent: In these cases, the “self-affirming done through language” [15] emerged as a consequence of this gesture.<sup>14</sup>

### 6.3 Should a Robot Be Designed to Harness Conditional Relevance?

Antithetical design opportunities stem from the observation that behaviors from a robot can establish a normative pressure to produce an adequate social response (here, a return greeting action). Roboticians may use such behaviors—documented to produce alignment from humans—to enforce a robot as a “social agent” at the beginning of an interaction, or, conversely, design the robot to align with the way in which participants treat it from the very start. For example, one could imagine designing a robot that only produces a “reinforcement wave” when its verbal greeting is not answered with a greeting action after several seconds, to purposely pressure a human interlocutor into greeting it like a legitimate social agent. This raises the question of whether designers should leverage conditional relevance as a tool—i.e., harness the tacit normative order [49] of human-human sequence organization [82]—to induce social treatments of the robot by participants; no matter how the robot is perceived by these participants. And, if so, to which extent?

Indeed, existing ethical and usability debates [21] can be connected to the legitimacy and to the capability of a robot to impact the degree to which it is being responded to as a social agent. As Reference [19] notes, once a robot can identify that some of its actions (even a limited set

<sup>12</sup>“Actions are intrinsically meaningful because they unavoidably participate in an organization of activity, not because there is an abstract, decontexted meaning which they have independent of their occurrence. Action is intrinsically meaningful, not because it is meaningful outside of any concrete situation, but because it is always embedded in a concrete situation” [63].

<sup>13</sup>Focusing on the Cozmo robot’s “sad” animation, designed to unfold in two parts, Reference [60] observed that the second part of this animation could be treated by participants as the upgrade of an action projected by the first part, leading them to reconsider their prior actions.

<sup>14</sup>This is not to say that, from this moment on, the robot was consistently and exclusively treated as a social agent during the rest of the interaction. On the contrary, “quick changes of perspective” [19] have been observed in the responses of participants to a robot. Reference [1] also notes that “each facets of the robot can take center stage as the encounter develops.”

of ritualized actions [90] such as greetings or goodbyes) are not being treated as those of a social agent, this opens up new possibilities for the field of personalization in robotics [11]. Information about the ongoing interactional status of a robot enables to trigger different behaviors from the robot when humans appear to be construing it exclusively as a “raw physical artifact” [10]. In particular, information about the current status of a robot in an interaction offers the possibility for a robot to adapt to the user’s observable initial definition of the situation (for example, by aligning with a treatment of itself as a simple device by an “utilitarian” [47] or “non-player” [10] interactant who only uses keywords and does not greet the robot), or, on the contrary, to rely on different strategies to change the (e.g., non-social) way in which it is being treated.

The phenomenon of conditional relevance, therefore, constitutes another factor in the question of the degree of agency a robot should display—of which the interpretation of our robots’ wave as a “reinforcement of the normative pressure of the robot’s first greeting” is a striking example—and highlights how even very mundane and minute choices from designers have ethical ramifications. Human-human sequence organization in conversations has been argued to be the place for a form of “proto-morality” [9, 75]: The treatment of a robot as involved (or not) in a sequence organization is directly connected with the (non-)attribution of specific rights and responsibilities to this robot in a “micro-level moral order” [89]. Consequently, an attempt to “enforce” this treatment has normative implications.

## 7 CONCLUSION: ENTRY INTO PHYSICAL CO-PRESENCE AS A BLIND SPOT IN HRI

Based on the video data from 80 interactions, we observed that the sudden activation of our robot (its “coming into existence”) was pragmatically consequential for participants. The intertwining between participants’ actions and the “activation steps” displayed by the robot (including its original motionlessness, which was sometimes oriented to as meaningful by participants—see also Reference [85]) led to the emergence of various sequential trajectories: Some participants ended up orienting to the robot’s gaze shift, to its wave, or to its greeting as a *response* to a greeting they just produced; others ended up orienting to these behaviors as *initiating* a greeting sequence, as *reinforcing* a previous greeting, and more generally as a *slot for a next action*. Nevertheless, coming back to our original question of “how” changes first emerge in the status of a robot during an interaction, it is not possible for us to suggest to which degree some patterns we identified in Section 6 (e.g., participants’ treatment of the robot’s delayed waving gesture) might be generalizable or, instead, remain specific to the local configuration of our experiment. This point could be clarified by a systematic comparison of interactions with a humanoid robot that “comes into sight” versus a robot that “comes into existence,” whether in a natural context or in a controlled experimental setting.

Crucially, when our robot started to move or to greet the human, it did not do so in the middle of an interactional vacuum: Participants were already building courses of action with it. A sole focus on the opening phase—starting when the robot is “alive” and starts to greet the human—would abstract these greetings from the preexisting sequential trajectories from which they emerged and in relation to which they can be understood: Just because two participants greeted a robot at the same step in this robot’s script, they did not necessarily do the same thing. An EMCA approach allows us to analyze “greetings” as something else than “synchronic snapshots” [3] but, on the contrary, to get a diachronic understanding of how they emerged, as well as to clarify interactional phenomena (initiation, response, reinforcement, etc.) that were simultaneously taking place as these greetings unfolded. Even seemingly straightforward behaviors from the robot, such as this robot looking at the human, saying “hello,” or waving at a certain point in the interaction, could

“mould in different sequential trajectories” [72], in which what was projected by these (etically similar) behaviors was oriented to in a radically different manner by participants.<sup>15</sup>

The previous observations put into question the moment at which data collection should start (video recording, movement tracking, etc.) in human-robot experiments, especially those that deal with the topic of robots as agents or partners. These studies should pay close attention to the way in which their participants enter into physical co-presence with the robot and, in particular, whether the robot “comes into sight” or “comes into existence.” Participants’ orientations to the very first behaviors displayed by the robot can produce *a priori* unpredictable sequential trajectories, which are, in turn, susceptible to configuring the timing and the manner in which the robot emerges as a social agent, and possibly participants’ behavior during the rest of the scenario. A robot oriented to by participants as already “activated” is not the same kind of entity as a robot that first appears to these participants as an immobile object and then “wakes up.” Therefore, we suggest that, each time it is relevant, researchers should take into account and describe the conditions in which robot and human were put into physical co-presence and regard “pre-beginnings” as an integral part of the experiment. Depending on the studies’ methodology and hypothesis, the state of the robot when it appears to participants could impact the comparability, replicability, and explainability of the findings.

## APPENDIX

### Transcription Conventions for Verbal Interactions

Transcriptions of talk follow [38]’s transcription conventions:

=	Latching of utterances
(.)	Short pause in speech (<200 ms)
(0.6)	Timed pause to tenths of a second
:	Lengthening of the previous sound
.	Stopping fall in tone
,	Continuing intonation
?	Rising intonation
°uh°	Softer sound than the surrounding talk
.h	Aspiration
h	Out breath
heh	Laughter
((text))	Described phenomena
.tk	Lips parting or a lip smack

### Transcription Conventions for Embodied Conduct

Embodied actions were transcribed using [56]’s multimodal transcription conventions:<sup>16</sup>

* *	Gestures and descriptions of embodied actions are delimited between
+ +	two identical symbols (one symbol per participant)

<sup>15</sup>An EMCA analysis of these human-robot interactions therefore appears to offer a level of description that stands closer to the involved point of view [13] of participants (rather than, e.g., a superficial description or coding of these scenes as mere greetings). For this reason, an EMCA approach could constitute a promising preliminary micro-analytic step [43, 81] for HRI studies coding participants’ behaviors before processing them quantitatively. For example, in an endeavor to compare users’ perception of a robot (obtained as self-reports in a post hoc questionnaire) with their observable behavior when interacting with this robot, one appears likely to find more meaningful results by using emic coding categories based on a detailed analysis of the video data (e.g., “treating a robot’s action as a reinforcement greeting”) rather than using more generic and etic coding categories (e.g., “responding to the robot’s greeting”).

<sup>16</sup><https://www.lorenzamondada.net/multimodal-transcription>

△ △	and are synchronized with corresponding stretches of talk.
*->	The action described continues across subsequent lines
-->*	until the same symbol is reached.
>>	The action described begins before the excerpt's beginning.
-->>	The action described continues after the excerpt's end.
....	Action's preparation.
--	Action's apex is reached and maintained.
,,,	Action's retraction.
ric	Participant doing the embodied action is identified in small caps in the margin.

In the current article, symbols and abbreviations used in transcriptions referred to the following multimodal dimensions:

HUM	Turn at talk from the human
ROB	Turn at talk from the robot
hum	Multimodal action from the human
rob	Multimodal action from the robot
img	Screenshot of a transcribed event
£	Human's gaze
%	Robot's gaze
*	Human's body
\$	Robot's body
+	Human's face
#	Position of a screenshot in the turn at talk

## REFERENCES

- [1] Morana Alač. 2016. Social robots: Things or agents? *AI Soc.* 31, 4 (Nov. 2016), 519–535. DOI: <https://doi.org/10.1007/s00146-015-0631-6>
- [2] Marit Aldrup. 2019. “Well let me put it uhm the other way around maybe”: Managing students' trouble displays in the CLIL classroom. *Classr. Discourse* 10, 1 (2019), 46–70. DOI: <https://doi.org/10.1080/19463014.2019.1567360>
- [3] Charles Antaki, Susan Condor, and Mark Levine. 1996. Social identities in talk: Speakers' own orientations. *Br. J. Soc. Psychol.* 35, 4 (1996), 473–492. DOI: <https://doi.org/10.1111/j.2044-8309.1996.tb01109.x>
- [4] Kika Arias, Sooyeon Jeong, Hae Won Park, and Cynthia Breazeal. 2020. Toward designing user-centered idle behaviors for social robots in the home. In *Proceeding of the 1st international workshop on Designery HRI Knowledge. Held in conjunction with the 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN'20)*.
- [5] Ritta Baddoura and Gentiane Venture. 2015. This robot is sociable: Close-up on the gestures and measured motion of a human responding to a proactive robot. *Int. J. Soc. Robot.* 7, 4 (Aug. 2015), 489–496. DOI: <https://doi.org/10.1007/s12369-015-0279-x>
- [6] Wayne A. Beach and Stuart J. Sigman. 1995. Conversation analysis: “Okay” as a clue for understanding consequentiality. In *The Consequentiality of Communication*. Routledge, London. 121–162.
- [7] Atef Ben-Youssef, Chloé Clavel, Slim Essid, Miriam Bilac, Marine Chamoux, and Angelica Lim. 2017. UE-HRI: A new dataset for the study of user engagement in spontaneous human-robot interactions. In *Proceedings of the 19th ACM International Conference on Multimodal Interaction*, ACM, 464–472. DOI: <https://doi.org/10.1145/3136755.3136814>
- [8] Tanya N. Beran, Alejandro Ramirez-Serrano, Roman Kuzyk, Meghann Fior, and Sarah Nugent. 2011. Understanding how children understand robots: Perceived animism in child–robot interaction. *Int. J. Hum.-Comput. Stud.* 69, 7–8 (July 2011), 539–550. DOI: <https://doi.org/10.1016/j.ijhcs.2011.04.003>
- [9] Jörg R. Bergmann. 1998. Introduction: Morality in discourse. *Res. Lang. Soc. Interact.* 31, (1998). Retrieved from <http://www.tandfonline.com/doi/abs/10.1080/08351813.1998.9683594>.
- [10] Herbert Clark and Kerstin Fischer. 2022. Social robots as depictions of social agents. *Behav. Brain Sci.* 2022, (July 2022), 1–33.
- [11] K. Dautenhahn. 2004. Robots we like to live with?! – A developmental perspective on a personalized, life-long robot companion. In *Proceedings of the 13th IEEE International Workshop on Robot and Human Interactive Communication*. 17–22. DOI: <https://doi.org/10.1109/ROMAN.2004.1374720>

- [12] Elwys De Stefani and Lorenza Mondada. 2018. Encounters in public space: How acquainted versus unacquainted persons establish social and spatial arrangements. *Res. Lang. Soc. Interact.* 51, 3 (July 2018), 248–270. DOI : <https://doi.org/10.1080/08351813.2018.1485230>
- [13] Hubert L. Dreyfus. 2001. Phenomenological description versus rational reconstruction. *Rev. Int. Philos.* 216, 2 (2001), 181–196. DOI : <https://doi.org/10.3917/rip.216.0181>
- [14] Hubert L. Dreyfus. 2002. Intelligence without representation – Merleau-Ponty’s critique of mental representation: The relevance of phenomenology to scientific explanation. *Phenomenol. Cogn. Sci.* 1, 4 (Dec. 2002), 367–383. DOI : <https://doi.org/10.1023/A:1021351606209>
- [15] Alessandro Duranti. 2005. Agency in language. In *Proceeding of the a Companion to Linguistic Anthropology*. John Wiley & Sons, Ltd., 449–473. DOI : <https://doi.org/10.1002/9780470996522.ch20>
- [16] Nj Enfield and Jack Sidnell. 2017. On the concept of action in the study of interaction. *Discourse Stud.* 19, 5 (Oct. 2017), 515–535. DOI : <https://doi.org/10.1177/1461445617730235>
- [17] Kerstin Fischer. 2007. The role of users’ concepts of the robot in human-robot spatial instruction. In *Spatial Cognition V Reasoning, Action, Interaction*. Springer Berlin, 76–89. DOI : [https://doi.org/10.1007/978-3-540-75666-8\\_5](https://doi.org/10.1007/978-3-540-75666-8_5)
- [18] Kerstin Fischer. 2011. Interpersonal variation in understanding robots as social actors. In *Proceedings of the 6th International Conference on Human-Robot Interaction (HRI’11)*. Association for Computing Machinery, New York, NY, 53–60. DOI : <https://doi.org/10.1145/1957656.1957672>
- [19] Kerstin Fischer. 2021. Tracking anthropomorphizing behavior in human-robot interaction. *J. Hum.-Robot Interact.* 11, 1 (Oct. 2021). DOI : <https://doi.org/10.1145/3442677>
- [20] Emma Frid, Roberto Bresin, and Simon Alexanderson. 2018. Perception of mechanical sounds inherent to expressive gestures of a NAO robot – Implications for movement sonification of humanoids. In *Proceedings of the 15th Sound and Music Computing Conference*. DOI : <https://doi.org/10.5281/zenodo.1422499>
- [21] Nora Fronemann, Kathrin Pollmann, and Wulf Loh. 2022. Should my robot know what’s best for me? Human-robot interaction between user experience and ethical design. *AI Soc.* 37, 2 (June 2022), 517–533. DOI : <https://doi.org/10.1007/s00146-021-01210-3>
- [22] Harold Garfinkel. 1967. *Studies in Ethnomethodology*. Polity Press, Cambridge.
- [23] Harold Garfinkel. 2006. *Seeing Sociologically: The Routine Grounds of Social Action*. Paradigm, Boulder.
- [24] J. J. Gibson. 1986. *The Ecological Approach to Visual Perception*. Lawrence Erlbaum Associates, Inc., Hillsdale, NJ.
- [25] Charles Goodwin. 1981. *Conversational Organization: Interaction between Speakers and Hearers*. Irvington Publishers, New York.
- [26] Laura C. Hand and Thomas J. Catlaw. 2019. Accomplishing the public encounter: A case for ethnomethodology in public administration research. *Perspect. Public Manag. Gov.* 2, 2 (May 2019), 125–137. DOI : <https://doi.org/10.1093/ppmgov/gvz004>
- [27] Katariina Harjunpää, Lorenza Mondada, and Kimmo Svinhufvud. 2018. The coordinated entry into service encounters in food shops: Managing interactional space, availability, and service during openings. *Res. Lang. Soc. Interact.* 51, 3 (July 2018), 271–291. DOI : <https://doi.org/10.1080/08351813.2018.1485231>
- [28] Paul ten Have. 2007. *Doing Conversation Analysis: A Practical Guide*. Sage Publications. Retrieved from <http://www.uk.sagepub.com/booksProdDesc.nav?prodId=Book229124>.
- [29] Christian Heath, Jon Hindmarsh, and Paul Luff. 2022. *Video in Qualitative Research: Analysing Social Interaction in Everyday Life*. 55 City Road, London. DOI : <https://doi.org/10.4135/9781526435385>
- [30] Brandon Heenan, Saul Greenberg, Setareh Aghel-Manesh, and Ehud Sharlin. 2014. Designing social greetings in human robot interaction. In *Proceedings of the Conference on Designing Interactive Systems*. ACM, 855–864. DOI : <https://doi.org/10.1145/2598510.2598513>
- [31] John Heritage. 2001. Goffman, Garfinkel and Conversation Analysis. In *Discourse Theory and Practices*. SAGE.
- [32] John Heritage. 2005. Conversation analysis and institutional talk. In *Handbook of Language and Social Interaction*. Erlbaum, Mahwah, 103–146.
- [33] Patrick Holthaus, Karola Pitsch, and Sven Wachsmuth. 2011. How can I help?: Spatial attention strategies for a receptionist robot. *Int. J. Soc. Robot.* 3, 4 (Nov. 2011), 383–393. DOI : <https://doi.org/10.1007/s12369-011-0108-9>
- [34] Patrick Holthaus and Sven Wachsmuth. 2021. It was a pleasure meeting you: Towards a holistic model of human-robot encounters. *Int. J. Soc. Robot.* 13, 7 (Nov. 2021), 1729–1745. DOI : <https://doi.org/10.1007/s12369-021-00759-9>
- [35] Robert Hopper. 2005. A cognitive agnostic in conversation analysis: When do strategies affect spoken interaction? In *Conversation and Cognition* (1st ed.). Cambridge University Press, 134–158. DOI : <https://doi.org/10.1017/CBO9780511489990.007>
- [36] Helge Huettenrauch, Kerstin Eklundh, Anders Green, and Elin Topp. 2006. Investigating spatial relationships in human-robot interaction. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 5052–5059. DOI : <https://doi.org/10.1109/IROS.2006.282535>

- [37] Salla Jarske, Sanna Raudaskoski, and Kirsikka Kaipainen. 2020. The “social” of the socially interactive robot: Rethinking human-robot interaction through ethnomethodology. In *Culturally Sustainable Social Robotics: Proceedings of Robophilosophy 2020*. IOS Press, 194–203. DOI: <https://doi.org/10.3233/FAIA200915>
- [38] Gail Jefferson. 2004. Glossary of transcript symbols with an introduction. In *Conversation Analysis: Studies from the First Generation*. John Benjamins.
- [39] Raya A. Jones. 2017. What makes a robot “social”? *Soc. Stud. Sci.* 47, 4 (Aug. 2017), 556–579. DOI: <https://doi.org/10.1177/0306312717704722>
- [40] Yusuke Kato, Takayuki Kanda, and Hiroshi Ishiguro. 2015. May I help you? Design of human-like polite approaching behavior. In *Proceedings of the 10th Annual ACM/IEEE International Conference on Human-Robot Interaction (HRI’15)*. Association for Computing Machinery, New York, NY, 35–42. DOI: <https://doi.org/10.1145/2696454.2696463>
- [41] Leelo Keevallik. 2018. Sequence initiation or self-talk? Commenting on the surroundings while mucking out a sheep stable. *Res. Lang. Soc. Interact.* 51, 3 (July 2018), 313–328. DOI: <https://doi.org/10.1080/08351813.2018.1485233>
- [42] Adam Kendon. 1990. *Conducting Interaction: Patterns of Behavior in Focused Encounters*. Cambridge University Press, New York, NY.
- [43] Kobin H. Kendrick. 2017. Using conversation analysis in the lab. *Res. Lang. Soc. Interact.* 50, 1 (Jan. 2017), 1–11. DOI: <https://doi.org/10.1080/08351813.2017.1267911>
- [44] Kobin H. Kendrick, Penelope Brown, Mark Dingemanse, Simeon Floyd, Sonja Gipper, Kaoru Hayano, Elliott Hoey, Gertie Hoymann, Elizabeth Manrique, Giovanni Rossi, and Stephen C. Levinson. 2020. Sequence organization: A universal infrastructure for social action. *J. Pragmat.* 168, (Oct. 2020), 119–138. DOI: <https://doi.org/10.1016/j.pragma.2020.06.009>
- [45] Mardi Kidwell. 2005. Gaze as social control: How very young children differentiate “The Look” from a “Mere Look” by their adult caregivers. *Res. Lang. Soc. Interact.* 38, 4 (Oct. 2005), 417–449. DOI: [https://doi.org/10.1207/s15327973rlsi3804\\_2](https://doi.org/10.1207/s15327973rlsi3804_2)
- [46] Min Kyung Lee, Sara Kiesler, and Jodi Forlizzi. 2010. Receptionist or information kiosk: How do people talk with a robot? In *Proceedings of the ACM Conference on Computer Supported Cooperative Work – CSCW’10*. ACM Press. 31. DOI: <https://doi.org/10.1145/1718918.1718927>
- [47] Min Kyung Lee, Sara Kiesler, Jodi Forlizzi, Siddhartha Srinivasa, and Paul Rybski. 2010. Gracefully mitigating breakdowns in robotic services. In *Proceeding of the 2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 203–210. DOI: <https://doi.org/10.1109/HRI.2010.5453195>
- [48] Dirk vom Lehn. 2019. From Garfinkel’s “Experiments in Miniature” to the ethnomethodological analysis of interaction. *Hum. Stud.* 42, 2 (Sept. 2019), 305–326. DOI: <https://doi.org/10.1007/s10746-019-09496-5>
- [49] Michael Lempert. 2013. No ordinary ethics. *Anthropol. Theor.* 13, 4 (2013), 370–393. DOI: <https://doi.org/10.1177/1463499613505571>
- [50] Christian Licoppe. 2017. Skype appearances, multiple greetings and “coucou”: The sequential organization of video-mediated conversation openings. *Pragmat. Q. Publ. Int. Pragmat. Assoc. IPra* 27, 3 (Oct. 2017), 351–386. DOI: <https://doi.org/10.1075/prag.27.3.03lic>
- [51] Christian Licoppe and Nicolas Rollet. 2020. “«Je dois y aller».” Analyses de séquences de clôtures entre humains et robot. *Réseaux* N°220–221, 2 (2020), 151. DOI: <https://doi.org/10.3917/res.220.0151>
- [52] Phoebe Liu, Dylan F. Glas, Takayuki Kanda, and Hiroshi Ishiguro. 2016. Data-driven HRI: Learning social behaviors by example from human–human interaction. *IEEE Trans. Robot.* 32, 4 (Aug. 2016), 988–1008. DOI: <https://doi.org/10.1109/TRO.2016.2588880>
- [53] Maxim Makatchev, Min Kyung Lee, and Reid Simmons. 2009. Relating initial turns of human-robot dialogues to discourse. In *Proceedings of the 4th ACM/IEEE International Conference on Human Robot Interaction*. ACM Press. DOI: <https://doi.org/10.1145/1514095.1514196>
- [54] Douglas W. Maynard. 2006. Cognition on the ground. *Discourse Stud.* 8, (2006). Retrieved from <https://journals.sagepub.com/doi/abs/10.1177/1461445606059560>.
- [55] Lorenza Mondada. 2015. Ouverture et préouverture des réunions visiophoniques. *Réseaux* 194 (2015), 39–84. DOI: <https://doi.org/10.3917/res.194.0039>
- [56] Lorenza Mondada. 2016. Challenges of multimodality: Language and the body in social interaction. *J. Socioling.* 20, 3 (June 2016), 336–366. DOI: [https://doi.org/10.1111/josl.1\\_12177](https://doi.org/10.1111/josl.1_12177)
- [57] Lorenza Mondada. 2019. Transcribing silent actions: A multimodal approach of sequence organization. *Soc. Interact. Video-Based Stud. Hum. Sociality* 2, 1 (Mar. 2019). DOI: <https://doi.org/10.7146/si.v2i1.113150>
- [58] Lorenza Mondada, Julia Bänninger, Sofian A. Bouaouina, Laurent Camus, Guillaume Gauthier, Philipp Hänggi, Mizuki Koda, Hanna Svensson, and Burak S. Tekin. 2020. Human sociality in the times of the Covid-19 pandemic: A systematic examination of change in greetings. *J. Socioling.* 24, 4 (Sept. 2020), 441–468. DOI: <https://doi.org/10.1111/josl.12433>

- [59] Kristian Mortensen and Spencer Hazel. 2014. Moving into interaction—Social practices for initiating encounters at a help desk. *J. Pragmat.* 62, (Feb. 2014), 46–67. DOI : <https://doi.org/10.1016/j.pragma.2013.11.009>
- [60] Hannah R. M. Pelikan, Mathias Broth, and Leelo Keevallik. 2020. “Are you sad, Cozmo?”: How humans make sense of a home robot’s emotion displays. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction*. Association for Computing Machinery, New York, NY, 461–470. DOI : <https://doi.org/10.1145/3319502.3374814>
- [61] Hannah R. M. Pelikan. 2021. Why autonomous driving is so hard: The social dimension of traffic. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction*. 81–85.
- [62] Hannah R. M. Pelikan and Mathias Broth. 2016. Why that Nao?: How humans adapt to a conventional humanoid robot in taking turns-at-talk. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, 4921–4932. DOI : <https://doi.org/10.1145/2858036.2858478>
- [63] Mark Peyrot. 1982. Understanding ethnomethodology: A remedy for some common misconceptions. *Hum. Stud.* 5, 1 (Dec. 1982), 261–283. DOI : <https://doi.org/10.1007/BF02127681>
- [64] Danielle Pillet-Shore. 2012. Greeting: Displaying stance through prosodic recipient design. *Res. Lang. Soc. Interact.* 45, 4 (Oct. 2012), 375–398. DOI : <https://doi.org/10.1080/08351813.2012.724994>
- [65] Danielle Pillet-Shore. 2018. How to begin. *Res. Lang. Soc. Interact.* 51, 3 (July 2018), 213–231. DOI : <https://doi.org/10.1080/08351813.2018.1485224>
- [66] Karola Pitsch and Benjamin Koch. 2010. How infants perceive the toy robot Pleo. An exploratory case study on infant-robot-interaction. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction*.
- [67] Karola Pitsch, Hideaki Kuzuoka, Yuya Suzuki, Luise Sussenbach, Paul Luff, and Christian Heath. 2009. “The first five seconds”: Contingent stepwise entry into an interaction as a means to secure sustained engagement in HRI. In *Proceedings of the 18th IEEE International Symposium on Robot and Human Interactive Communication*. IEEE, 985–991. DOI : <https://doi.org/10.1109/ROMAN.2009.5326167>
- [68] Karola Pitsch, Anna-Lisa Vollmer, and Manuel Mühlhig. 2013. Robot feedback shapes the tutor’s presentation: How a robot’s online gaze strategies lead to micro-adaptation of the human’s conduct. *Interact. Stud. Soc. Behav. Commun. Biol. Artif. Syst.* 14, 2 (Aug. 2013), 268–296. DOI : <https://doi.org/10.1075/is.14.2.06pit>
- [69] Karola Pitsch, Anna-Lisa Vollmer, Katharina J. Rohlfing, Jannik Fritsch, and Britta Wrede. 2014. Tutoring in adult-child interaction: On the loop of the tutor’s action modification and the recipient’s gaze. *Interact. Stud. Soc. Behav. Commun. Biol. Artif. Syst.* 15, 1 (June 2014), 55–98. DOI : <https://doi.org/10.1075/is.15.1.03pit>
- [70] George Psathas. 1980. Approaches to the study of the world of everyday life. *Hum. Stud.* 3, 1 (1980), 3–17.
- [71] Marc Relieu. 2007. La téléprésence, ou l’autre visiophonie. *Réseaux* 144, 5 (2007), 183–223.
- [72] Marc Relieu, Merve Sahin, and Aurélien Francillon. 2020. Une approche configurationnelle des leurres conversationnels. *Réseaux* N°220-221, 2 (2020), 81. DOI : <https://doi.org/10.3917/res.220.0081>
- [73] Katie A. Riddoch and Emily. S. Cross. 2021. “Hit the robot on the head with this mallet” – Making a case for including more open questions in HRI research. *Front. Robot. AI* 8, (Feb. 2021), 603510. DOI : <https://doi.org/10.3389/frobt.2021.603510>
- [74] Jeffrey David Robinson. 1998. Getting down to business talk, gaze, and body orientation during openings of doctor-patient consultations. *Hum. Commun. Res.* 25, 1 (Sept. 1998), 97–123. DOI : <https://doi.org/10.1111/j.1468-2958.1998.tb00438.x>
- [75] Jessica S. Robles. 2015. Morality in discourse. *Int. Encyc. Lang. Soc. Interact.* 132–137.
- [76] Harvey Sacks, Emanuel A. Schegloff, and Gail Jefferson. 1974. A simplest systematics for the organization of turn-taking for conversation. *Language* 50, 4 (Dec. 1974), 696. DOI : <https://doi.org/10.2307/412243>
- [77] Joline Scheffler and Karola Pitsch. 2020. Pre-beginnings in human-robot encounters: Dealing with time delay. In *Proceedings of the European Conference on Computer-supported Cooperative Work*. European Society for Socially Embedded Technologies, Siegen. DOI : [https://doi.org/10.18420/ecscw2020\\_p02](https://doi.org/10.18420/ecscw2020_p02)
- [78] Emanuel A. Schegloff. 1968. Sequencing in conversational openings. *Amer. Anthropol.* 70, 6 (1968), 1075–1095. DOI : <https://doi.org/10.1525/aa.1968.70.6.02a00030>
- [79] Emanuel A. Schegloff. 1979. Identification and recognition in telephone conversation openings. In *Everyday Language: Studies in Ethnomethodology*. Irvington Publishers, Inc., New York, 23–78.
- [80] Emanuel A. Schegloff. 1980. Preliminaries to preliminaries: “Can I ask you a question?” *Sociol. Inq.* 50, 3–4 (July 1980), 104–152. DOI : <https://doi.org/10.1111/j.1475-682X.1980.tb00018.x>
- [81] Emanuel A. Schegloff. 1993. Reflections on quantification in the study of conversation. *Res. Lang. Soc. Interact.* 26, 1 (Jan. 1993), 99–128. DOI : [https://doi.org/10.1207/s15327973rlsi2601\\_5](https://doi.org/10.1207/s15327973rlsi2601_5)
- [82] Emanuel A. Schegloff. 2007. *Sequence Organization in Interaction: A Primer in Conversation Analysis*. Cambridge University Press, Cambridge. DOI : <https://doi.org/10.1017/CBO9780511791208>
- [83] Emanuel A. Schegloff. 2010. Some other “Uh(m)”s. *Discourse Process.* 47, 2 (Jan. 2010), 130–174. DOI : <https://doi.org/10.1080/01638530903223380>

- [84] Emanuel A. Schegloff and Harvey Sacks. 1973. Opening up closings. *Semiotica* 8, 4 (1973). DOI : <https://doi.org/10.1515/semi.1973.8.4.289>
- [85] Trenton Schulz, Rebekka Soma, and Patrick Holthaus. 2021. Movement acts in breakdown situations: How a robot's recovery procedure affects participants' opinions. *Paladyn J. Behav. Robot.* 12, 1 (Aug. 2021), 336–355. DOI : <https://doi.org/10.1515/pjbr-2021-0027>
- [86] Mi-Suk Seo and Irene Koshik. 2010. A conversation analytic study of gestures that engender repair in ESL conversational tutoring. *J. Pragmat.* 42, 8 (Aug. 2010), 2219–2239. DOI : <https://doi.org/10.1016/j.pragma.2010.01.021>
- [87] Rachel L. Severson and Stephanie M. Carlson. 2010. Behaving as or behaving as if? Children's conceptions of personified robots and the emergence of a new ontological category. *Neural Netw.* 23, 8–9 (2010), 1099–1103. DOI : <https://doi.org/10.1016/j.neunet.2010.08.014>
- [88] Steven Stanley, Robin James Smith, Eleanor Ford, and Joshua Jones. 2020. Making something out of nothing: Breaching everyday life by standing still in a public place. *Sociol. Rev.* 68, 6 (Nov. 2020), 1250–1272. DOI : <https://doi.org/10.1177/0038026120940616>
- [89] Tanya Stivers, Lorenza Mondada, and Jakob Steensig. 2011. Knowledge, morality and affiliation in social interaction. In *The Morality of Knowledge in Conversation*. Cambridge University Press, Cambridge. 3–24. DOI : <https://doi.org/10.1017/CBO9780511921674.002>
- [90] Tanya Stivers and Federico Rossano. 2010. Mobilizing response. *Res. Lang. Soc. Interact.* 43, 1 (Feb. 2010), 3–31. DOI : <https://doi.org/10.1080/08351810903471258>
- [91] Ilona Straub. 2016. “It looks like a human!” The interrelation of social presence, interaction and agency ascription: A case study about the effects of an android robot on social agency ascription. *AI Soc.* 31, 4 (Nov. 2016), 553–571. DOI : <https://doi.org/10.1007/s00146-015-0632-5>
- [92] Ilona Straub, Shuichi Nishio, and Hiroshi Ishiguro. 2012. From an object to a subject – Transitions of an android robot into a social being. In *Proceedings of the 21st IEEE International Symposium on Robot and Human Interactive Communication*. IEEE, 821–826. DOI : <https://doi.org/10.1109/ROMAN.2012.6343853>
- [93] Karen Tatarian, Rebecca Stower, Damien Rudaz, Marine Chamoux, Arvid Kappas, and Mohamed Chetouani. 2021. How does modality matter? Investigating the synthesis and effects of multi-modal robot behavior on social intelligence. *Int. J. Soc. Robot.* (Nov. 2021). DOI : <https://doi.org/10.1007/s12369-021-00839-w>
- [94] Hamish Tennent, Dylan Moore, Malte Jung, and Wendy Ju. 2017. Good vibrations: How consequential sounds affect perception of robotic arms. In *Proceedings of the 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 928–935. DOI : <https://doi.org/10.1109/ROMAN.2017.8172414>
- [95] Sam Thellman, Jacob Lundberg, Mattias Arvola, and Tom Ziemke. 2017. What is it like to be a bot?: Toward more immediate wizard-of-oz control in social human-robot interaction. In *Proceedings of the 5th International Conference on Human Agent Interaction*. ACM, 435–438. DOI : <https://doi.org/10.1145/3125739.3132580>
- [96] Sylvaine Tuncer, Sarah Gillet, and Iolanda Leite. 2022. Robot-mediated inclusive processes in groups of children: From gaze aversion to mutual smiling gaze. *Front. Robot. AI* 9, (2022). DOI : <https://doi.org/10.3389/frobt.2022.729146>
- [97] Yutaka Yamauchi and Takeshi Hiramoto. 2016. Reflexivity of routines: An ethnomethodological investigation of initial service encounters at sushi bars in Tokyo. *Organ. Stud.* 37, 10 (Oct. 2016), 1473–1499. DOI : <https://doi.org/10.1177/0170840616634125>
- [98] Fangkai Yang, Yuan Gao, Ruiyang Ma, Sahba Zojaji, Ginevra Castellano, and Christopher Peters. 2021. A dataset of human and robot approach behaviors into small free-standing conversational groups. *PLoS One* 16, 2 (Feb. 2021), e0247364. DOI : <https://doi.org/10.1371/journal.pone.0247364>

Received 24 February 2022; revised 29 August 2022; accepted 13 October 2022