Coming Into Sight or Coming Into Existence: Entry in Physical Co-presence as a Blind Spot in HRI

Damien Rudaz
Dept. of Economics and Social Sciences
Telecom Paris and Institut Polytechnique de Paris
Palaiseau, France
damien.rudaz@telecom-paris.fr

Christian Licoppe
Dept. of Economics and Social Sciences
Telecom Paris and Institut Polytechnique de Paris
Palaiseau, France
christian.licoppe@telecom-paris.fr

Karen Tatarian
Institute for Intelligent Systems and Robotics
Sorbonne University
Paris, France
tatarian.karen@sorbonne-universite.fr

Rebecca Stower
CHArt Laboratory
University Paris 8
Paris, France
rebecca.stower@univ-paris8.fr

***Abstract*—From an experiment which replicated the interaction opening delays often observed in laboratory or "in-the-wild" HRI studies, where robots often require several seconds before springing to life after they are in co-presence with a human, we suggest that the very first moments of physical co-presence between a participant and a robot are not anecdotal nor peripheral. We hold that a robot oriented to by participants as "alive" or "activated" is not the same kind of entity as a robot which first appears to these participants as an immobile object: it doesn't afford the same action possibilities. Using two examples from our corpus, we highlight that the intertwining between participants' actions and the very first behaviors or the motionlessness displayed by the robot produces *a priori* unpredictable sequential trajectories, which are susceptible to configuring the timing and the manner in which the robot emerges as a social agent during HRI experiments.**

Keywords—preopening, robot, social agent, motionless, delayed entry, greetings

I. Introduction

For humanoid robots as for human beings, emerging as "agents" is not systematically granted by the sharing of a mutual space [1] with other participants. Rather, this process requires interactional work, generally done as part of the "pre-opening" [2]. For example, the emergence of greetings, by which a participant would recognize "the Other as a possible agent" [1] (robot or human), implies the production of a "framework in which a greeting sequence is relevant and expectable" [3], or a "proper interaction frame" [4].

A small number of studies have explicitly studied the organisation of (pre-)opening sequences between humans and robots [5]–[9]. However, these very first seconds, during which the robot appears to participants, tend to be treated as peripheral or anecdotal in the ways many HRI experiments are designed and recorded: these moments where participants encounter the robot for the first time are often, so to speak, "off-the-record" in the data which end up being collected. Relying on an experiment during which a humanoid robot went through several "activation steps" once it was in physical co-presence with participants, we attempt to highlight that the robot's very first behaviors can be consequential with respect to the way in which it emerges as a conversation partner. That is, the early activation phase of the robot was oriented to by our participants in the way they designed and timed their first greeting turn.

II. Pre-opening desins in HRI

A. Ways of appearing for robots

Focusing exclusively on the moment at which the robot appears 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.

Studies where the robot stands motionless when participants encounter it, without displaying any preexisting idle behavior nor adjustments to the participants' approach or presence: e.g. the Wizard of Oz has to seat participants in front of the robot before going behind a divider to send commands to the robot [10] or has to deal with a significant response time [11], the script is not launched yet (e.g. [8], [12], this study), the autonomous robot's reactions are delayed [7], etc. In these situations, participants find themselves in physical co-presence with the robot for a long period, before reciprocal exchanges and mutual identification become possible: there is a delay "between entry into physical copresence and moves to enter into social copresence" [13].

Studies where the robot, or virtual agent, already displays idling behaviors when it appears to participants (simulated breathing, random head movements, etc. – e.g. [14], [15]) *and/or observably adjusts to the human's approach or physical co-presence* (e.g. [6], [16]–[18]). This includes any form of preexisting activity from the robot, similarly to human service-encounters where salespersons, help desk staff, etc. are often already immersed in an activity prior to the sighting of the customer/patient/student [3], [19].

B. *"Coming into sight" vs "coming into existence"*

These two categories make relevant an earlier distinction made by J.J. Gibson in his ecological 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 [20]. Gibson calls this type of appearance a "coming into sight" [21]. 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" [21] to allow for the "pop-up", quasi-instantaneous character. Another example of "coming into existence" would be the initial connection in a video call [4].

All this may be highly relevant to HRI, for it now appears clearly 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. A robot oriented to by participants as "alive" or "activated" is not the same kind of entity as a robot which first appears to these participants as an immobile object. It's likely to afford different action possibilities to these participants.

C. *Off-the-record pre-openings*

However, few studies mention the state of the robot when participants see it 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" [9], methodology sections rarely cover the observable behavior of the robot when participants encounter it. Most HRI experiments display an orientation to the "opening" phase of the interaction as the first relevant moment, and tend to neglect the "pre-opening" phase, although, depending on their experimental scenario, it may be crucial to the way participants and robots achieve some form of co-participation status.

III. Study Setup

The following fragments come from a corpus of 80 video recordings of dyadic interactions with an autonomous robot, which took place at the INSEAD-Sorbonne Université Behavioural Lab. 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 1.1). The interaction was filmed with two cameras: one behind the robot, one on the left of the robot.

A. *Scenario*

The robot was designed as a "travel agent". Once a participant 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. For a detailed description of our experimental setup and of the design of the autonomous robot, see [12].

B. *"Activation steps"*

Therefore, participants were brought in the presence of a robot which was designed to "come into existence" instead of displaying availability from the very start. After each participant entered the room, the robot went through the same 5 "activation steps".

1) Physical co-presence: When participants entered the room, the robot was motionless

2) Gaze tracking: The robot started to track their gaze and produced motor and plastic sounds as it shifted its head towards them

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

IV. Data

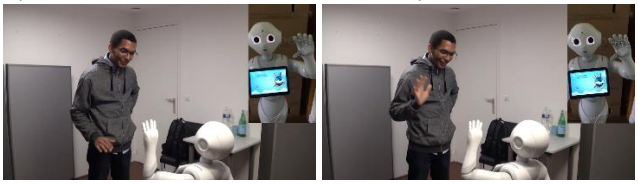
From this relatively common HRI experimental setup, we attempt to highlight that these "waking up" behaviors were relevant to participants as they designed and timed their first greeting turn. We illustrate this analysis by focusing on two fragments from this corpus.

The following transcriptions start immediately after the participants have entered the room.[1]

A. *Fragment 1: Mutual gaze as projecting an imminent next action from the robot*

1. *** (1.7) #* (2) * (0.7)**
hum *>>closes bag*moves to robot*adjusts clothes->
hum #gazes at robot->
2. ROB **((motor and plastic sounds))%$**
rob %gazes at human>>
rob $arms shaking>>
3. **(0.1)*(0.6)**
hum ->*
4. HUM **.tsk .h**
5. **#(0.5)**
hum ->#gazes at his sleeve->
6. HUM **°he#llow::°***
hello
hum ->#gazes at robot->
hum *takes a step forward->
7. **(1)*(2.1)* (0.8) *(2.8)**
hum ->* *takes a step backward*
8. ROB **$*bonjour**
hello
rob $opens its arms->
hum *takes hand out of pocket->
9. **(0.6)* (0.3) * (0.2) *(0.2)$(0.2)**
hum ->*extends arm*retracts arm*
rob ->$waves->
10. HUM **bonjour**
hello
11. **#(0.4)*(0.5)£**
hum ->#gazes at wave->
hum *extends hand toward robot->
fig £fig.1.1[2]

1.1 1.2



[1] Embodied actions were transcribed following Lorenza Mondada's multimodal transcription conventions: https://www.lorenzamondada.net/multimodal-transcription
Talk was transcribed following the ICOR conventions: http://icar.cnrs.fr/projets/corinte/documents/2013_Conv_ICOR_250313.pdf

[2] Participants' consent was obtained for the use of video data.

12.HUM .tsk #*+(.)£hello
hello
hum ->#gazes at robot's face>>
hum ->*waves->
hum +smiles->>
fig £fig.1.2
13. (0.5) * (0.8)
hum ->*
14.ROB je m'app$elle peppeur
my name is Pepper
rob ->$

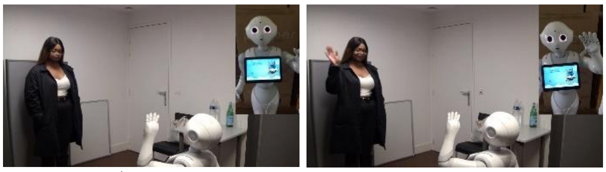
After entering the room, the participant doesn't treat physical co-presence as sufficient to initiate a greeting sequence. This is especially visible through his production of "self-grooming" (L.1), usually displayed during the approach between two interactants [17]. However, the status of the robot in the interaction shifts after the establishment of mutual gaze (L.2). The participant's interruption of his self-grooming (L.3), his greeting (L.6), and the step forwards he takes (L.6) accentuate a shared inner space [20] and display the expectation of an imminent action from the robot. This reconfiguration results from the "crucial analytic distinction" [3] made by the participant about what the gaze from the robot is projecting: it's not oriented to as a merely automatic "gaze tracking", nor as a "mere look" [3] 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.

Crucially, this participant orients to the first behaviors of the robot as sequentially equivocal [22]. His first actions display uncertainty regarding what the robot's gaze, greeting and wave are projecting (or if they are projecting anything). This is especially visible in the two reconfigurations in this participant's gestures. First, L.9, he extends his hand towards the robot 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; Fig.2.6 to 2.9). These two episodes display quickly evolving interpretations of what action the robot is projecting through its first behaviors after it was activated.

B. Fragment 2: The robot's waving gesture as an upgrade of its vocal greeting

1. # (11) # (0.2)
hum >>#looks around the room#gazes at robot->
2. ROB ((motor and plastic sounds))%$
rob %gazes at human>>
rob $arms shaking>>
3. (6.2)
4. ROB $bonjour
hello
rob $opens its arms->
5. (1.1)(0.8)(0.4)£
rob ->$........$-waves->
fig £fig.1.1
6. HUM *bon*+jour:£
hello
hum *..... *-waves->
hum +smiles>>
fig £fig.2.2

2.1 2.2



7. $(0.2) * (0.6) * (0.1)
rob ->$,,,,,,,,,,,,,,,,,,,,,,,,,,,,->
hum ->*,,,,,,,,,,,*
8. ROB je m'appelle$ peppeur
my name is Pepper
rob ->$

Significantly, this fragment displays a form of inertia: 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 doesn't immediately institute it as a conversational partner which can be greeted back.

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 (L.1) – however, she doesn't immediately produce a speaking turn. The lasting silence and mutual gaze (L.1) are not oriented to as initiating a "slot" where to self-select [23]. Even after the robot utters a "bonjour" (L.4), the participant returns no greeting and maintains her previous pose and gaze for the next few seconds.

Once the robot starts a waving gesture (L.5), 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 her verbal greeting (L.6). The speed of this return wave may 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 second greeting sequence which reinforces the conditional relevance attached to its first vocal first greeting ("bonjour"), to which she didn't answer.

V. Discussion

The previous fragments exemplify two forms of interactional work required before "behaviors" from the robot could be treated as "actions" which either 1) established the adequate framework to initiate a first greeting sequence or 2) produced a response slot that the participant was normatively pressured to complete with a return greeting. As the robot went through different "activation steps", participants founds themselves engaged in sequentially ambiguous situations [22] as to what actions it was projecting (or if it was projecting anything). When the robot suddenly animated and "came into existence", 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" [24]. This is a possible

explanation for the delayed emergence of the robot as an agent (fragment 2) – an extremely common situation in our corpus.

In sum, when the robot started to move or to greet the human, it didn't do so in the middle of an interactional vacuum: participants had already started to co-construct courses of action with it. The intertwining between participants' actions and the first behaviors displayed by the robot (including its immobility, which can be treated as meaningful by participants [25]) therefore led to the emergence of various sequential trajectories. Overall, in our corpus, 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, treating the robot as able to perceive their greeting and respond to it; others ended up orienting to these behaviors as initiating a greeting sequence, or as reinforcing a previous greeting.

VI. Methological implications

A sole focus on the opening phase – starting when the robot is "alive" and starts to greet the human – would abstract these participants' greetings from the pre-existing 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 didn't necessarily do the same thing. This puts into question the moment at which data collection should start (video recording, movement tracking, etc.) in human-robot experiments, especially those which deal with the topic of robots as agents or partners.

We suggest that, when it's relevant to their particular hypothesis, researchers should take in account and describe the conditions in which robot and human were put in physical co-presence; this may positively impact the comparability, replicability and explainability of their findings. This phase of the interaction is not anecdotal or peripheral with respect to many HRI experiments – in particular if they are interested in evaluating the perception or the treatment of a robot as a social agent – but should be thought about and designed as an integral part of those. *A priori* unpredictable sequential trajectories can emerge from these early moments and are, in turn, susceptible to configuring the timing and the manner in which the robot emerges as a social agent.

References

[1] A. Duranti, "Agency in Language," in *A Companion to Linguistic Anthropology*, 2007, pp. 449–473. doi: 10.1002/9780470996522.ch20.

[2] L. Mondada, "Ouverture et préouverture des réunions visiophoniques," *Réseaux*, no. 194, pp. 39–84, 2015, doi: 10.3917/res.194.0039.

[3] K. Mortensen and S. Hazel, "Moving into interaction—Social practices for initiating encounters at a help desk," *J. Pragmat.*, vol. 62, pp. 46–67, Feb. 2014, doi: 10.1016/j.pragma.2013.11.009.

[4] C. Licoppe, "Skype appearances, multiple greetings and 'coucou': The sequential organization of video-mediated conversation openings," *Pragmat. Q. Publ. Int. Pragmat. Assoc. IPrA*, vol. 27, no. 3, pp. 351–386, Oct. 2017, doi: 10.1075/prag.27.3.03lic.

[5] M. Alač, "Social robots: Things or agents?," *AI Soc.*, vol. 31, no. 4, pp. 519–535, Nov. 2016, doi: 10.1007/s00146-015-0631-6.

[6] K. Pitsch, H. Kuzuoka, Y. Suzuki, L. Sussenbach, P. Luff, and C. Heath, "'The first five seconds': Contingent stepwise entry into an interaction as a means to secure sustained engagement in HRI," in *RO-MAN 2009 - The 18th IEEE International Symposium on Robot and Human Interactive Communication*, Toyama, Japan, Sep. 2009, pp. 985–991. doi: 10.1109/ROMAN.2009.5326167.

[7] J. Scheffler and K. Pitsch, "Pre-beginnings in Human-Robot Encounters: Dealing with time delay," Siegen, 2020. doi: 10.18420/ecscw2020_p02.

[8] H. R. M. Pelikan and M. Broth, "Why That Nao?: How Humans Adapt to a Conventional Humanoid Robot in Taking Turns-at-Talk," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, San Jose California USA, May 2016, pp. 4921–4932. doi: 10.1145/2858036.2858478.

[9] P. Holthaus, K. Pitsch, and S. Wachsmuth, "How Can I Help?: Spatial Attention Strategies for a Receptionist Robot," *Int. J. Soc. Robot.*, vol. 3, no. 4, pp. 383–393, Nov. 2011, doi: 10.1007/s12369-011-0108-9.

[10] T. N. Beran, A. Ramirez-Serrano, R. Kuzyk, M. Fior, and S. Nugent, "Understanding how children understand robots: Perceived animism in child–robot interaction," *Int. J. Hum.-Comput. Stud.*, vol. 69, no. 7–8, pp. 539–550, Jul. 2011, doi: 10.1016/j.ijhcs.2011.04.003.

[11] S. Thellman, J. Lundberg, M. Arvola, and T. Ziemke, "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*, Bielefeld Germany, Oct. 2017, pp. 435–438. doi: 10.1145/3125739.3132580.

[12] K. Tatarian, R. Stower, D. Rudaz, M. Chamoux, A. Kappas, and M. Chetouani, "How does Modality Matter? Investigating the Synthesis and Effects of Multi-modal Robot Behavior on Social Intelligence," *Int. J. Soc. Robot.*, Nov. 2021, doi: 10.1007/s12369-021-00839-w.

[13] D. Pillet-Shore, "How to Begin," *Res. Lang. Soc. Interact.*, vol. 51, no. 3, pp. 213–231, Jul. 2018, doi: 10.1080/08351813.2018.1485224.

[14] K. A. Riddoch and Emily. S. Cross, "'Hit the Robot on the Head With This Mallet' – Making a Case for Including More Open Questions in HRI Research," *Front. Robot. AI*, vol. 8, p. 603510, Feb. 2021, doi: 10.3389/frobt.2021.603510.

[15] F. Yang, Y. Gao, R. Ma, S. Zojaji, G. Castellano, and C. Peters, "A dataset of human and robot approach behaviors into small free-standing conversational groups," *PLOS ONE*, vol. 16, no. 2, p. e0247364, Feb. 2021, doi: 10.1371/journal.pone.0247364.

[16] Y. Kato, T. Kanda, and H. Ishiguro, "May I Help You? Design of Human-like Polite Approaching Behavior," in *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, New York, NY, USA, 2015, pp. 35–42. doi: 10.1145/2696454.2696463.

[17] B. Heenan, S. Greenberg, S. Aghel-Manesh, and E. Sharlin, "Designing social greetings in human robot interaction," in *Proceedings of the 2014 conference on Designing interactive systems*, Vancouver BC Canada, Jun. 2014, pp. 855–864. doi: 10.1145/2598510.2598513.

[18] P. Holthaus and S. Wachsmuth, "It was a Pleasure Meeting You: Towards a Holistic Model of Human–Robot Encounters," *Int. J. Soc. Robot.*, vol. 13, no. 7, pp. 1729–1745, Nov. 2021, doi: 10.1007/s12369-021-00759-9.

[19] K. Harjunpää, L. Mondada, and K. Svinhufvud, "The Coordinated Entry into Service Encounters in Food Shops: Managing Interactional Space, Availability, and Service During Openings," *Res. Lang. Soc. Interact.*, vol. 51, no. 3, pp. 271–291, Jul. 2018, doi: 10.1080/08351813.2018.1485231.

[20] A. Kendon, *Conducting interaction: Patterns of behavior in focused encounters.* New York, NY, US: Cambridge University Press, 1990, pp. xi, 292.

[21] J. J. Gibson, *The Ecological Approach to Visual Perception*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc., 1986.

[22] R. Hopper, "A cognitive agnostic in conversation analysis: when do strategies affect spoken interaction?," in *Conversation and Cognition*, 1st ed., H. te Molder and J. Potter, Eds. Cambridge University Press, 2005, pp. 134–158. doi: 10.1017/CBO9780511489990.007.

[23] H. Sacks, E. A. Schegloff, and G. Jefferson, "A Simplest Systematics for the Organization of Turn-Taking for Conversation," *Language*, vol. 50, no. 4, p. 696, Dec. 1974, doi: 10.2307/412243.

[24] E. A. Schegloff, "Preliminaries to Preliminaries: 'Can I Ask You a Question?,'" *Sociol. Inq.*, vol. 50, no. 3–4, pp. 104–152, Jul. 1980, doi: 10.1111/j.1475-682X.1980.tb00018.x.

[25] T. Schulz, R. Soma, and P. Holthaus, "Movement acts in breakdown situations: How a robot's recovery procedure affects participants' opinions," *Paladyn J. Behav. Robot.*, vol. 12, no. 1, pp. 336–355, Aug. 2021, doi: 10.1515/pjbr-2021-0027.