Empathy Display Influence on Human-Robot Interactions: a Pilot Study

Laurianne Charrier, Alexandre Galdeano, Amélie Cordier, and Mathieu Lefort

***Abstract*—Social robots are designed to interact and communicate with humans. We have conducted a pilot study to explore how an artificial empathy module can affect Human-Robot Interactions. For that pilot study, we chose to evaluate the effects of a module we developed called "attention-based empathic module" and we set up an experiment within two conditions, "Empathy" (i.e., with this module), and "No-Empathy" (i.e., without this module). In order to define what aspects of HRI are affected, we used several metrics found in HRI literature including self-reported questionnaires—e.g., perceived empathy test—physiological measures—e.g., number of attentional disengagements—and objective measures—e.g., time of interaction, and performance. Dividing 36 participants into two groups and controlling the main biases inherent in subjects selection, we found that the "attention-based empathic module" seems to have affected 9 metrics: the interaction duration, how trustworthy the robot was perceived, the number of disengagements, how empathic the robot was perceived, how much participants felt they knew the robot, how the robot's intelligence was perceived, how comfortable the interaction was perceived, how much the robot was perceived as knowledgeable, and how engaging the interaction was perceived. Due to the exploratory approach of this study, these results have to be confirmed.**

***Keywords*—Human-Robot Interaction, Social Robots, Empathy, Empathic Display, Interaction Involvement, Attention.**

I. Introduction

With small and affordable robots like Cozmo or Jibo, social robots arrive on the mass market, democratizing technologies that were up to now reserved to research institutions and companies. A social robot must be able to perceive its environment through senses [14], act on it following social norms [3], and more generally be able to perform social interactions in a Human-like way [8]. One way to improve social interactions with robots and which is greatly explored in HRI studies is to give robots empathic capabilities. Empathy can be defined as a complex process whereby one understands and/or shares an entity's frame of reference [41], and/or react appropriately [45] without one having doubts on which frame of reference belongs to whom [44]. Depending on the definitions, this frame of reference ranges from only one's emotional state to all one's mental states. Empathic capabilities can be analyzed under two main categories, the ones that are more emotional and the ones that are more cognitive. Emotional empathy can be defined as the ability to experience and understand another entity's affective experience by sharing the same feelings [33], [1], it also includes mimicking behaviors. Cognitive empathy refers to the ability to represent and understand the internal mental states of someone and to be intellectually able of perspective taking [7], [31]. This kind of empathy is generally associated with the notion of theory of mind [33], which is the ability of an observer to attribute mental states—e.g., knowledge, beliefs, intentions, thoughts, emotions, desires—to themselves and to others [7], [34] in order to predict, adapt to, and explain their behaviors [27]. It's important to notice that, with these definitions, being unemphatic is to be unable to understand and predict the mental states of others whereas doing incongruent empathy is about misunderstanding others' mental states and displaying an inappropriate use of empathy. As in several studies—e.g., [23], [24], [13], [20]—we chose to focus our study on cognitive empathy only, more precisely on the robot's relevance in adapting its behaviors to the user's interaction involvement. Interaction involvement can be considered as a mental state and its understanding is part of cognitive empathy.

According to literature definitions, interaction involvement can be conceptualized as how much an individual partakes in a social environment [9] and, more precisely, to the extent participants are immersed and engaged in an ongoing social interaction [10], [12]. It is a characteristic way of processing information and respond to messages during face-to-face communication that can be decomposed into three factors [17]:

Responsiveness: the need to respond to the situation in order to lead to a meaningful conversation.

Perceptiveness: an individual's ability to assign meaning to others' behaviors and interpret the meanings others assign to one's own behaviors.

Attentiveness: one's degree of cognitive involvement in an interaction, as attentional commitment.

In face-to-face settings, someone being involved in the interaction is attentive to the other and responsive to the evolving circumstances of conversation. In this pilot study, we used a robotic module called "attention-based empathic module" that measures the user's attentiveness using a custom deep-learning model which has images from the robot camera as inputs. When the module detects a loss of attention, the robot reacts in order to get the user's attention back, i.e., it triggers an animation, and it makes the robot say sentences such as "You must be attentive".

The goal of this pilot study is to identify which measures are the most relevant when studying the influence of the empathic display on the interaction. This will help to give insights on how to setup further experiments, and on which tools and

Laurianne Charrier and Alexandre Galdeano are with Univ Lyon, Université Lyon 1, CNRS, LIRIS, F-69621, Villeurbanne, France and with Hoomano, 4 rue du Professeur Charles Appleton, 69007, Lyon, France, e-mail: laurianne.charrier@hoomano.com; alexandre.galdeano@liris.cnrs.fr.

Amélie Cordier is with Hoomano, 4 rue du Professeur Charles Appleton, 69007, Lyon, France, e-mail: amelie.cordier@hoomano.com.

Mathieu Lefort is with Univ Lyon, Université Lyon 1, CNRS, LIRIS, F-69621, Villeurbanne, e-mail: mathieu.lefort@liris.cnrs.fr.

measures to use, rather than drawing solid conclusions about the algorithm's effects on the interaction.

We first detail our methodology in Section II including the material used (Section II-A), the participants' selection and evaluation (Section II-B), the experiment procedure (Section II-C), the measures used (Section II-D), and how we analyzed the data (Section II-E). In Section III, we present our results regarding the self-reported (Section III-A), objective (Section III-B), and physiological measures (Section III-C). We then discuss our results in Section IV.

II. Methodology

A. Material

We chose to use Pepper[1]—a 1.2meter high humanoid robot made by SoftBank Robotics. This mass-market robot has the ability to communicate with Humans through speech, gestures, and its tablet.

A quiz in English has been implemented with 32 questions in various knowledge fields, 4 answers per question are given (Fig 1), and each question is followed by an anecdote. It is very much like the "Who Wants to Be a Millionaire?" TV show. The quiz aims at entertaining the user as long as possible, but the user is asked every four questions to continue or not. The first condition is the quiz as-is—the No-Empathy condition—and the second one is the quiz with the attention-based empathic module—the Empathy condition. In the two conditions, Pepper was animated to make the interaction more natural, with a face tracking and gestures when speaking.

In the Empathy condition, when the user is not attentive to what the robot is saying, Pepper makes large and exaggerated gestures and calls back for the attention of the user using sentences like "You must be attentive" as could be done by a professor in front of an inattentive student. This simple behavior makes it easier to understand the causes behind its effects on the user's perception of both the robot and the interaction. Pepper was placed in a calm and closed room with good lighting conditions, and always in the same spot to avoid biases. A NAO robot[2] was put on a cabinet next to Pepper to record the experimentation without participants knowing they were filmed, and all its lights were turned off for this purpose. GStreamer[3] was used to get the video stream from the NAO's camera and VLC player[4] was used to record it.

B. Participants

All the questionnaires were in English and we wanted to avoid biases with translations So we chose to also do the quiz in English for coherence. We also evaluated the English understanding level of each participant. to keep the questionnaires and the quiz's questions in English to avoid biases with translations, we evaluated the English understanding level of each participant, i.e., only people that could at least make simple sentences and understand the main points of a conversation in English were included in the experiment. Then, robot experience, acceptability, and personality were tested to ensure that each group was balanced as much as possible in term of bias sources: each bias could indeed lead to a different way to interact. Participants proceed to the mini IPIP test [16] to measure five sides of their personality, including their extroversion, and a homemade questionnaire inspired by the Eurobarometer 382 [42] to measure their robot experience. To complete these questionnaires, acceptability was measured with the Negative Attitudes towards Robots Scale (NARS) [40] and the Robot Anxiety Scale (RAS) [26]. Personal empathy level was tested with the Interpersonal Reactivity Index (IRI) [15].

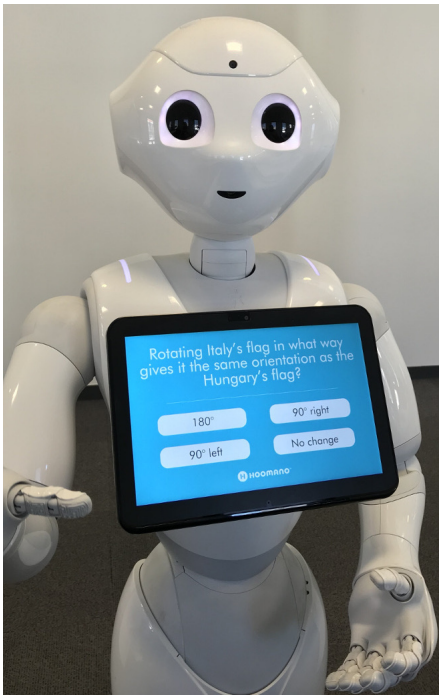

Figure 1. Example of a question in our quiz. The robot could give clues to the user after a fixed time. And, after giving the answer, tell anecdotes about each answer.

The 36 participants were subdivided randomly into two groups to perform one of the two experimental conditions. Each group was composed of 18 persons—9 males and 9 females—with most of them being in the 18–27 years-old range. In addition to these participants, 8 more participants in the Empathy condition initially achieved the experiment but did not trigger the interaction involvement detector and consequently did not experience the robot's empathic behaviors. This is due to two factors: 1) the attention-based empathic module did not detect any loss of attention, or; 2) the participant was attentive during the whole experiment. These subjects' results were removed from the study because we wanted to only measure the potential effects of the empathic displays on how the interaction and the robot are perceived.

C. Procedure

First, participants responded to an online questionnaire about demography, personality, empathy, robot experience

[1]See https://www.softbankrobotics.com/emea/en/robots/pepper
[2]See https://www.softbankrobotics.com/emea/en/robots/nao
[3]See https://gstreamer.freedesktop.org/
[4]See https://www.videolan.org/vlc/

and acceptability. The results from these questionnaire were used to make the two groups while limiting biases as much as possible. They were told that the experiment was about studying the effects of personality on how the social robots is used to limit response bias. They then came to test our application at Hoomano's office for a session of about 30 minutes. After giving their authorization to use video and signing consent forms, the experiment began without them knowing it. Participants were told that they had to train first with the quiz on Pepper, during as much time as they wanted, and that they will then perform the experiment. We did that to limit experimental bias with people knowing they were looked at and not acting naturally [18], [29]. After being sure that the subject well understood the instructions, the experimenter left the subject alone with the robot, launching the data gathering when coming out of the room. After the participants came back from the quiz task, we told them that they actually performed the experiment and that they had to fulfill a post-experiment questionnaire. This questionnaire was used to evaluate how the participants felt about the interaction and the robot.

D. Measures

We wanted to test different kinds of measures so we decided to mix self-reported, physiological, and objective measures. In total, we chose to test eight metrics to evaluate the effect of the attention-based empathic module, each of these have already been used in HRI studies and, more generally, in social studies. The self-reported metrics were:

- The Godspeed test [4]: a robot acceptance questionnaire.
- Bickmore's test items [6]: a robot and an interaction evaluation questionnaire.
- The Barrett-Lennard Reactivity Index (BLRI) [2]: an empathy questionnaire used to evaluate the robot's empathy.

As a physiological measure, we used the number of disengagements of the subject calculated with the video. At last, we tested four objective measures of the interaction: the distance between the robot and the subject, the number of questions answered, the number of good answers given, and the duration of the interaction. We evaluated the distance using the front sonar of the robot, while we recovered the duration of usage and number of disengagements on the robot after every play. For the distance, we removed the last 15 percents of the data to avoid noise deriving from the experiment's end. We then calculated the mean distance and its standard deviation.

E. Statistical analysis

Since we are using an exploratory experimental method, we decided to set the significance level α to 20% for the metrics analysis, i.e., we considered a metric as being potentially affected by the attention-based empathic module if the difference in this metric's results between the groups leads to a p-value below α. We did that to select measures of interest for future experiments, rather than drawing solid conclusions. To assess the difference in a measure between the two groups, we used a T-test [39] if and only if the data from both groups were close enough to a normal distribution—i.e., the Shapiro-Wilk tests [32] were not significant—and they had fairly equal variances—i.e., the F-test of equality of variances was not significant either. If a measure did not follow these conditions, we used a Mann-Whitney test [25]. We measured correlations using Spearman's rank correlation coefficient [36]. All statistical analysis were processed using the R software[5].

Test	Metric	p
Bi	Not deceitful – Deceitful	.20
Go	Perceived safety	.21
Bi	Careful – Careless	.22
Bi	Reliable	.23
Bi	Natural	.26
Bi	Smooth	.30
Bi	Respectful – Disrespectful	.33
Bi	Confidential – Divulging	.34
Bi	Warm	.36
Bi	How did Pepper understand you?	.38
In	Measured distance (mean)	.41
Go	Anthropomorphism	.44
Bi	Successful	.45
Bi	Lifelike	.45
Ba	Unconditionnality	.49
Bi	Expert	.49
Bi	How do you characterize your relationship with Pepper?	.52
Ba	Level of regard	.54
Bi	How well do you feel Pepper knows you and your needs?	.55
In	Number of good answers given	.56
Bi	Safe – Dangerous	.56
In	Number of questions asked	.56
Bi	Straightforward – Tricky	.57
Bi	Intelligent	.58
Bi	Would you enjoy playing with Pepper again?	.61
In	Disengagements per minute	.61
In	Measured distance (st. dev)	.61
Bi	Sincere – Insincere	.61
Bi	Likable	.62
Bi	Satisfying	.63
In	Good answer ratio	.65
Bi	Candid – Deceptive	.67
Bi	Honest – Dishonest	.73
Bi	Awkward	.77
Go	Animacy	.75
Bi	Fun	.76
Go	Likability	.76
Bi	Faithful – Unfaithful	.78
Bi	How much did you like Pepper?	.78
Bi	Trustful of Pepper – Distrustful of Pepper	.81
Bi	Credible	.81
Bi	Benevolent – Exploitive	.81
Bi	Involving	.82
Ba	Congruence	.84
Bi	Enjoyable	.85
Bi	Reliable – Unreliable	.85
Bi	Informed	.85
Bi	Friendly	.86
Bi	Tedious	.89
Bi	Competent	.90
Bi	Interesting	.91
Bi	Familiar	.91
Bi	Efficient	.92
Bi	Pleasant	.94
Bi	Considerate – Inconsiderate	.99

Table I
LIST OF ALL THE METRICS WHICH HAD P-VALUES ABOVE α ORDERED BY INCREASING P-VALUE. "BA" IS FOR THE BARRETT-LENNARD REACTIVITY INDEX [2], "BI" IS FOR BICKMORE'S TEST ITEMS [6], "GO" IS FOR THE GODSPEED TEST [4], AND "IN" IS FOR THE INTERACTION MEASURES.

III. RESULTS

A. Self-reported measures

We subdivided the self-reported measures into three sections: perceived empathy, robot acceptance, and interaction evaluation. Due to the important number of tested items, we just reported significant items here. A list of all non-significant items is available in Table I.

For the interaction evaluation, there were only two significant differences between Empathy and No-Empathy groups. First, a significant difference has been found between the two groups in term of empathy ($U = 213.5, p = .106$) in the BLRI (see Fig 2).

Indeed, the robot has been perceived as more empathic in the Empathy condition with a mean of ($M = -0.89, SD = 6.88$) whereas the No-Empathy condition had a mean of ($M = -3.56, SD = 6.44$).

For the Godspeed test [4], the perceived intelligence in the No-Empathy condition and the Empathy condition may be considered as normally distributed—($S = .965, p = .707$)

[5]See https://www.r-project.org/

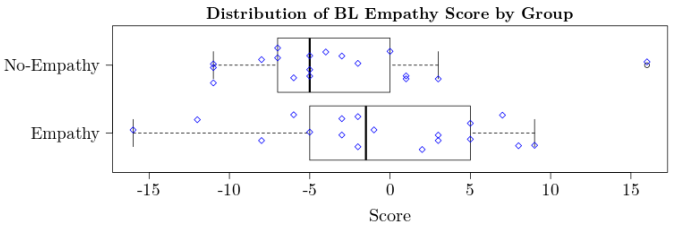

Figure 2. Distribution of the perceived empathy scores measured by the BLRI between the two conditions. The average score for the No-Empathy condition is ($M = -3.56, SD = 6.44$) and ($M = -.89, SD = 6.88$) for the Empathy condition. The Mann-Whitney score is 213.5 which accounts for a p-value of .106.

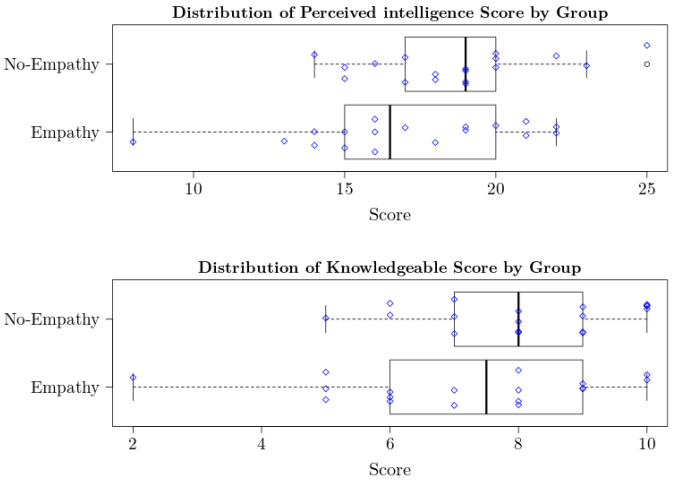

Figure 3. The robot was perceived as less intelligent (top) and less knowledgeable (bottom) in the Empathy condition: ($t(35) = -1.528, p = .136$) and ($U = 116.5, p = .149$) respectively.

and ($S = .945, p = .349$) respectively—and having equal variances ($F = 1.638, p = .319$). This was the only scale which was significantly different between the Empathy and the No-Empathy groups ($t(35) = -1.528, p = .136$) (see Fig 3). Pepper with the empathy module has been considered less intelligent than without empathy with respective means of ($M = 17.00, SD = 3.65$) and ($M = 18.67, SD = 2.85$), and less knowledgeable with respective means of ($M = 7.11, SD = 2.08$) and ($M = 8.11, SD = 1.53$). These two scales have a moderate positive correlation: ($r_S = .52, p = .027$) for the Empathy condition, and ($r_S = .731, p = .001$) for the No-Empathy condition. It is interesting to note that this difference in perceived intelligence did not affect likability nor anthropomorphism results.

Bickmore's test items [6] were analyzed separately. Some items reported interaction evaluation and others robot evaluation. First, the comfortability ($U = 116.5, p = .142$), with an interaction perceived as less comfortable with a mean of ($M = 6.67, SD = 1.91$) in the Empathy condition whereas in the No-Empathy condition it had a mean of ($M = 7.50, SD = 1.20$) (see Fig 4). Second, its engaging aspect ($U = 203.5, p = .188$), with an interaction perceived as more engaging with the empathy algorithm ($M = 7.22, SD = 1.56$) than without

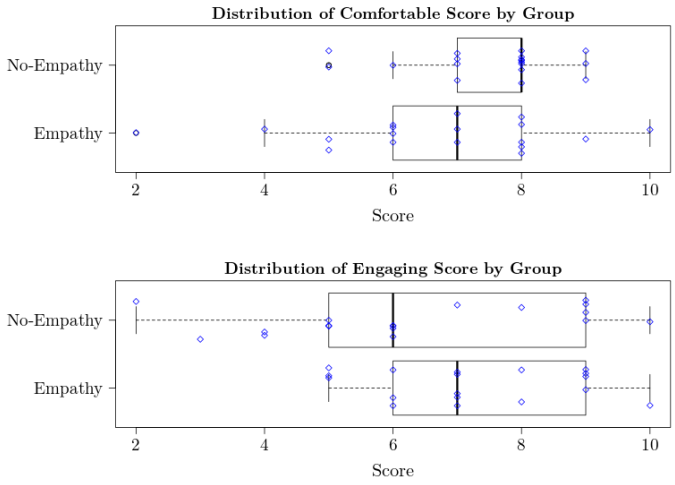

Figure 4. While the average comfortable score (top) for the No-Empathy condition is greater than in the Empathy condition ($U = 116.5, p = .142$), the interaction was considered more engaging (bottom) in the Empathy condition ($U = 203.5, p = .188$).

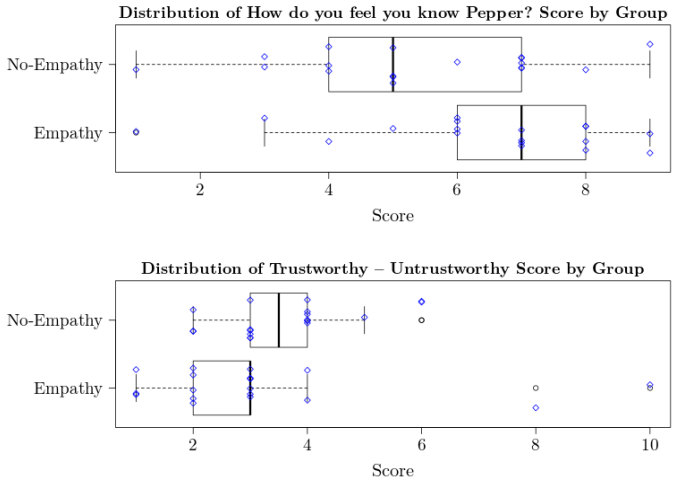

Figure 5. Participants in the Empathy condition felt they knew better Pepper (top) and that it was more trustworthy (bottom) with scores ($U = 212, p = .113$) and ($U = 103.5, p = .059$) respectively.

($M = 6.28, SD = 2.32$) (see figure 5). In term of how the robot was perceived, the participants significantly felt that they knew better the robot ($U = 212, p = .113$) in the Empathy condition ($M = 6.39, SD = 2.09$) than in the No-Empathy condition ($M = 5.39, SD = 2.03$). They also significantly perceived ($U = 103.5, p = .059$) the robot as less untrustworthy in the Empathy condition ($M = 3.17, SD = 2.33$) than in the No-Empathy condition ($M = 3.61, SD = 1.20$). Finally, there is a significant difference ($U = 212, p = .113$) between the two groups in term of perceived knowledge of the robot. The robot has been perceived as less knowledgeable with the empathy module ($M = 7.11, SD = 2.08$) than without ($M = 8.11, SD = 1.53$).

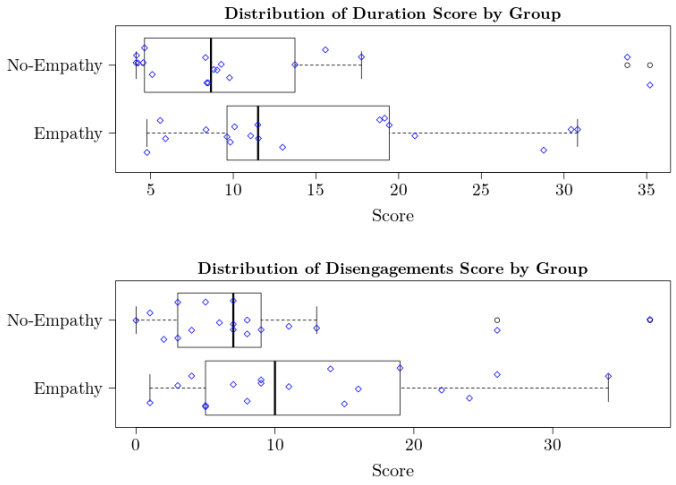

Figure 6. The interaction duration (top) and disengagements number (bottom) were higher in the Empathy condition: ($U = 226, p = .044$) and ($U = 216.5, p = .087$) respectively. However, there was no difference in the disengagements per minute, this can be explained by the weak-to-moderate positive correlation between the duration and the disengagements: ($r_S = .39, p = .109$) for the Empathy condition, and ($r_S = .512, p = .03$) for the No-Empathy condition.

B. Objective measures

Distance between the robot and the user, number of questions asked, number of good answers were not significantly different between the Empathy and No-Empathy groups. On the other hand, interaction duration was significantly longer ($U = 226, p = .044$) in the Empathy condition ($14.98m \pm 8.44m$) than in the No-Empathy condition ($11.39m \pm 9.29m$). The participants played more than 3 minutes longer with the quiz in the Empathy condition.

C. Physiological measure

The physiological measure that we used—the number of disengagements—was significantly different ($U = 216.5, p = .087$) between the Empathy group and the No-Empathy group but the difference in number of disengagements by minute was not significant ($U = 178.5, p = .613$). This can be explained by the positive correlation between the number of disengagements and duration in both conditions: ($r_S = .39, p = .109$) for the Empathy condition, and ($r_S = .512, p = .03$) for the No-Empathy condition.

IV. Discussion

We observed that the "attention-based empathic module" seemed to have affected 9 metrics—7 if we remove the two correlations:

1) the interaction duration;
2) how trustworthy the robot was perceived;
3) the number of disengagements (positively correlated to the interaction duration);
4) how empathic the robot was perceived;
5) how much participants felt they knew the robot;
6) how the robot's intelligence was perceived;
7) how comfortable the interaction was perceived;
8) how much the robot was perceived as knowledgeable (positively correlated to the perceived intelligence); and
9) how engaging the interaction was perceived.

We observed that even if there was a slight increase in perceived empathy, we also observed a reduction of interaction quality: users perceived the interaction as less comfortable with the empathy module. This might be a sign of an inappropriate use of empathy. Indeed, previous results shown that a robot is perceived as less safe and credible than a neutral robot when displaying incongruent empathy [13]. And since an empathic display not adapted to the situation may lead to a worse interaction perception, there is a need to think about the way our robots can show effectively that they understand Human behaviors in order to avoid increasing frustration facing actual robots. However, [22] has shown that empathy can improve a lot the Human-Computer interactions even in case of dysfunction of the robot.

Another line of thought is that maybe we cannot just apply Human-Human empathy to a social robot because we may not accept the same behavior coming from a Machine than from another person. Moreover, [5] has shown that interacting at a personal distance, with 46cm to 122cm between Human and robot, only small and medium gestures were appropriate. And in our study the users stood at an average distance of 50cm, this could explain why the users may have wrongly interpreted the exaggerated movements that Pepper did to get back their attention. So the empathy display might have to be more subtle to do not disturb the user. One way to achieve that would be for the robot to mimic the user's posture and gestures. Many studies demonstrated the relationship between mimicry and liking in Human-Human interaction. Indeed, [28] demonstrated a better satisfaction with the interaction if a robot mimics the upper body gestures of the user than if it does not. [37] shown an increased empathy towards the mimicked while being intentionally or spontaneously mimicked. [38] also compared liking after mimicking an a priori disliked or liked person. They concluded among others that when a person mimics a disliked person, liking for them was not improved. However, mimicking a liked person improved liking. If there is no information about the liking or disliking toward the person, mimicry also improves the liking for the mimicker [11]. This conclusion shows the importance for a robot to be at least positively seen to lead to a beneficial effect of mimicking. Mimicking could allow better emotional contagion [38] and make robots more lively. Other forms of empathy displays could be interesting to use like face emotion displays [19], use of light effects on the robot [35], and semantic adaptations to the context [24].

On the other hand, [43] manipulated the description of a NAO robot, giving it more of less cognitive abilities. The robot was always the same and always acted the same way, but it was perceived differently due to the initial description. The more the robot is said to have greater cognitive abilities, the less it is perceived as smart, trustful or as having true emotions. Participants also felt less understood.

When the user thinks that the robot has higher cognitive behaviors, this leads to a disappointment because the user expects the robot to act like a Human.

It might explain why our robot is perceived as less intelligent and less knowledgeable because, noticing that the robot is interpreting their behaviors, subjects may have attributed more cognitive abilities to the robot and, doing so, had more expectations. According to [8], user's expectations are one of the main issues in HRI. They "can mitigate the person's disappointment or frustration when interacting with the robot", and they "can also gently steer the person to interact with the robot in the way it was intended". We have to think about the effect of empathy displays on users expectations to avoid depreciation of interaction quality due to higher expectations. However, the "Intelligent" item from Bickmore's test did not show a significant difference between the two groups, so maybe the difference in Godspeed's item was a false positive.

The fact that the robot has been perceived as more trustworthy in the Empathy condition can be explained by giving trust to Pepper knowing it better as mechanical and as not willing to try to manipulate their mind because of a lack of cleverness. Attention has to be paid to the features that create expectations and the way the user sees the robot, as a tool or as a sociable partner distinguished on the base of the mental model a Human has of the robot when interacting with it [8].

An interesting phenomenon is the interaction duration that is on average about 3 minutes and a half longer in the Empathy group—about +32%—while responding to the same number of questions. It can partly be explained by the time taken by the robot to ask for attention. Moreover, in the Empathy condition, the robot asked back the attention of the users when they did not focus anymore on the game, so the users may be more focused and could spend more time to think about the question because of the feeling to be called to order. Having the robot asking for attention would also explain why the interaction felt more engaging for them.

In addition to the low number of subjects, this experiment suffers from limitations such as metrics fetching that can be improved. Videos were bad in recording quality leading to a slight loss of interaction details that could be interpreted. A better quality could help to evaluate more precisely our metrics. Another limitation is the use of English language based questionnaires and quiz questions on a French-speaking population. We chose to keep all the questionnaires in English, but it has been proved that word perception and meaning, such as emotion meaning, could vary with culture and the language in which the word is displayed [30], [21]. Misunderstanding of the meaning of some English words in the questionnaire could lead to biases.

V. Conclusion and Future Work

In this 36-person pilot study, we explored how the HRI may be affected by the robot displaying its empathic understanding, and we selected nine measures that seem to be good candidates for such studies. In the Empathy condition, the participants perceived the robot as more empathic and the interaction as more engaging, the participants also they felt they knew the robot better and interacted with for about 3.5 minutes more. Furthermore, in the No-Empathy condition, the robot was perceived as more intelligent and more knowledgeable, and the interaction felt more comfortable. Moreover, participants made three kinds of comments about the study after completion:

a) The way Pepper gets back the attention is exaggerated: A need to carefully analyze Human behavior to adapt empathy displays from the robot is one of the main conclusions of this study.

b) The questionnaires are too long: Because the pre-questionnaires are used to evaluate the demography of our sample and to limit biases while assigning participants to groups, they cannot really be changed. However, with this pilot study, we found that some of the measures we used measures—self-reported, objective, or physiological—are either not important or—due to correlations—redundant in our case. They will be removed in future studies.

c) The Barrett-Lennard Reactivity Index is not adapted: While this measure was influenced as expected in the Empathy condition, some questions—such as "Pepper doesn't avoid or go round anything that matters between us"—seemed weird to the participants. This kind of questions should probably be removed because changing them may affect the final score in ways we do not expect.

We removed 8 participants for whom the experiment did not go as planned: they did not experience the empathy display. This means that they had the same experience as if they were in the No-Empathy condition. So we also analyzed the data by considering them as being in the No-Empathy condition, and the measures that were significant before still are after this change, with—for most of them—a lower p-value. Moreover, new measures appeared below the 20% threshold: "How do you characterize your relationship with Pepper?" in favor of the Empathy condition, and "Perceived safety" and "Respectful – Disrespectful" in favor of the No-Empathy condition.

For a next study, we should change the empathic display, translate material in the native language of the observed population, select fewer hypotheses, decrease α to 10 or 5%, apply multiple testing corrections, and increase the number of participants. Based on our results and our definitions, we should make new metrics to evaluate perceived empathy with two scales—Empathic Understanding and Empathic Response. To evaluate interaction quality and perception of the robot, we should work on two questionnaires with different scales to avoid the multiplication of hypotheses. What follows from our study is that interaction quality should be evaluated in term of comfortableness and engagingness whereas the perception of the robot should be assessed regarding its perceived intelligence and trustworthiness.

Acknowledgements

This research was supported by the by the French *Association Nationale de la Recherche et de la Technologie* (ANRT) contract 2017/0226 and the joint laboratory BEHAVIORS.AI[6] funded by the French *Agence Nationale de la Recherche* (ANR) contract ANR-16-LCV2-0003-01.

[6]Website: https://behaviors.ai

REFERENCES

[1] M. Asada, "Development of artificial empathy," *Neuroscience research*, vol. 90, pp. 41–50, 2015.

[2] G. Barrett-Lennard, "The relationship inventory: A technique for measuring therapeutic dimensions of an interpersonal relationship," in *Mimeographed paper read at the Annual Conference of the Southwestern Psychological Association in St. Augustine, Florida*, 1959.

[3] C. Bartneck and J. Forlizzi, "A design-centred framework for social human-robot interaction," in *Robot and Human Interactive Communication, 2004. ROMAN 2004. 13th IEEE International Workshop on*. IEEE, 2004, pp. 591–594.

[4] C. Bartneck, D. Kulić, E. Croft, and S. Zoghbi, "Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots," *International journal of social robotics*, vol. 1, no. 1, pp. 71–81, 2009.

[5] C. L. Bethel and R. R. Murphy, "Survey of non-facial/non-verbal affective expressions for appearance-constrained robots," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 38, no. 1, pp. 83–92, 2008.

[6] T. W. Bickmore, "Relational agents: Effecting change through human-computer relationships," Ph.D. dissertation, Massachusetts Institute of Technology, 2003.

[7] R. J. R. Blair, "Responding to the emotions of others: dissociating forms of empathy through the study of typical and psychiatric populations," *Consciousness and cognition*, vol. 14, no. 4, pp. 698–718, 2005.

[8] C. Breazeal, "Social interactions in hri: the robot view," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 34, no. 2, pp. 181–186, 2004.

[9] D. J. Cegala, "Interaction involvement: A cognitive dimension of communicative competence," *Communication Education*, vol. 30, no. 2, pp. 109–121, 1981.

[10] D. J. Cegala, G. T. Savage, C. C. Brunner, and A. B. Conrad, "An elaboration of the meaning of interaction involvement: Toward the development of a theoretical concept," *Communications Monographs*, vol. 49, no. 4, pp. 229–248, 1982.

[11] T. L. Chartrand and J. A. Bargh, "The chameleon effect: the perception–behavior link and social interaction." *Journal of personality and social psychology*, vol. 76, no. 6, p. 893, 1999.

[12] D. A. Coker and J. Burgoon, "The nature of conversational involvement and nonverbal encoding patterns," *Human Communication Research*, vol. 13, no. 4, pp. 463–494, 1987.

[13] H. Cramer, J. Goddijn, B. Wielinga, and V. Evers, "Effects of (in) accurate empathy and situational valence on attitudes towards robots," in *Human-Robot Interaction (HRI), 2010 5th ACM/IEEE International Conference on*. IEEE, 2010, pp. 141–142.

[14] K. Dautenhahn and A. Billard, "Bringing up robots or—the psychology of socially intelligent robots: From theory to implementation," in *Proceedings of the third annual conference on Autonomous Agents*. ACM, 1999, pp. 366–367.

[15] M. H. Davis, *Interpersonal reactivity index*. Edwin Mellen Press, 1980.

[16] M. B. Donnellan, F. L. Oswald, B. M. Baird, and R. E. Lucas, "The mini-ipip scales: tiny-yet-effective measures of the big five factors of personality." *Psychological assessment*, vol. 18, no. 2, p. 192, 2006.

[17] R. L. Duran and B. H. Spitzberg, "Toward the development and validation of a measure of cognitive communication competence," *Communication Quarterly*, vol. 43, no. 3, pp. 259–275, 1995.

[18] K. L. Fode, "The effect of non-visual and non-verbal interaction on experimental bias," Ph.D. dissertation, University of North Dakota, 1960.

[19] B. Gonsior, S. Sosnowski, C. Mayer, J. Blume, B. Radig, D. Wollherr, and K. Kühnlenz, "Improving aspects of empathy subjective performance for hri through mirroring emotions," in *Proc. IEEE Intern. Symposium on Robot and Human Interactive Communication, RO-MAN 2011, Atlanta, USA*, 2011.

[20] F. Hegel, T. Spexard, B. Wrede, G. Horstmann, and T. Vogt, "Playing a different imitation game: Interaction with an empathic android robot," in *Humanoid Robots, 2006 6th IEEE-RAS International Conference on*. IEEE, 2006, pp. 56–61.

[21] G. J. Hofstede, "Mental activity and culture: The elusive real world," in *Advances in Culturally-Aware Intelligent Systems and in Cross-Cultural Psychological Studies*. Springer, 2018, pp. 143–164.

[22] K. Hone, "Empathic agents to reduce user frustration: The effects of varying agent characteristics," *Interacting with computers*, vol. 18, no. 2, pp. 227–245, 2006.

[23] I. Leite, G. Castellano, A. Pereira, C. Martinho, and A. Paiva, "Empathic robots for long-term interaction," *International Journal of Social Robotics*, vol. 6, no. 3, pp. 329–341, 2014.

[24] I. Leite, A. Pereira, S. Mascarenhas, C. Martinho, R. Prada, and A. Paiva, "The influence of empathy in human–robot relations," *International journal of human-computer studies*, vol. 71, no. 3, pp. 250–260, 2013.

[25] H. B. Mann and D. R. Whitney, "On a test of whether one of two random variables is stochastically larger than the other," *The annals of mathematical statistics*, pp. 50–60, 1947.

[26] T. Nomura, T. Suzuki, T. Kanda, and K. Kato, "Altered attitudes of people toward robots: Investigation through the negative attitudes toward robots scale," in *Proc. AAAI-06 workshop on human implications of human-robot interaction*, vol. 2006, 2006, pp. 29–35.

[27] D. Premack and G. Woodruff, "Does the chimpanzee have a theory of mind?" *Behavioral and brain sciences*, vol. 1, no. 4, pp. 515–526, 1978.

[28] L. D. Riek and P. Robinson, "Real-time empathy: Facial mimicry on a robot," in *Workshop on Affective Interaction in Natural Environments (AFFINE) at the International ACM Conference on Multimodal Interfaces (ICMI 08). ACM*. Citeseer, 2008.

[29] R. Rosenthal and K. L. Fode, "The problem of experimenter outcome-bias," *Series research in social psychology. Washington, DC: National Institute of Social and Behavioral Science*, 1961.

[30] J. A. Russell, "Culture and the categorization of emotions." *Psychological bulletin*, vol. 110, no. 3, p. 426, 1991.

[31] M. Sarlo, L. Lotto, R. Rumiati, and D. Palomba, "If it makes you feel bad, don't do it! egoistic rather than altruistic empathy modulates neural and behavioral responses in moral dilemmas," *Physiology & behavior*, vol. 130, pp. 127–134, 2014.

[32] S. S. Shapiro and M. B. Wilk, "An analysis of variance test for normality (complete samples)," *Biometrika*, vol. 52, no. 3/4, pp. 591–611, 1965.

[33] A. Smith, "Cognitive empathy and emotional empathy in human behavior and evolution," *The Psychological Record*, vol. 56, no. 1, pp. 3–21, 2006.

[34] B. Sodian and S. Kristen, "Theory of mind," in *Towards a theory of thinking*. Springer, 2010, pp. 189–201.

[35] S. Song and S. Yamada, "Bioluminescence-inspired human-robot interaction: Designing expressive lights that affect human's willingness to interact with a robot," in *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*. ACM, 2018, pp. 224–232.

[36] C. Spearman, "The proof and measurement of association between two things," *The American journal of psychology*, vol. 15, no. 1, pp. 72–101, 1904.

[37] M. Stel, R. B. Van Baaren, and R. Vonk, "Effects of mimicking: Acting prosocially by being emotionally moved," *European Journal of Social Psychology*, vol. 38, no. 6, pp. 965–976, 2008.

[38] M. Stel and R. Vonk, "Mimicry in social interaction: Benefits for mimickers, mimickees, and their interaction," *British Journal of Psychology*, vol. 101, no. 2, pp. 311–323, 2010.

[39] Student, "The probable error of a mean," *Biometrika*, pp. 1–25, 1908.

[40] D. S. Syrdal, K. Dautenhahn, K. L. Koay, and M. L. Walters, "The negative attitudes towards robots scale and reactions to robot behaviour in a live human-robot interaction study," *Adaptive and Emergent Behaviour and Complex Systems*, 2009.

[41] S. Tisseron, *Le jour où mon robot m'aimera : vers l'empathie artificielle [The day my robot will love me: towards artificial intelligence]*. Albin Michel, 2015.

[42] TNS Opinion & Social, "Special Eurobarometer 382: public attitudes towards robots," Tech. Rep., 2012.

[43] J. Vallverdú, T. Nishida, Y. Ohmoto, S. Moran, and S. Lázare, "Fake empathy and human-robot interaction (hri): A preliminary study," *International Journal of Technology and Human Interaction (IJTHI)*, vol. 14, no. 1, pp. 44–59, 2018.

[44] P. A. Van Lange *et al.*, *Bridging social psychology: Benefits of transdisciplinary approaches*. Psychology Press, 2006.

[45] A. van Ruiten, D. Haitas, P. Bingley, H. Hoonhout, B. Meerbeek, and J. Terken, "Attitude of elderly towards a robotic game-and-train-buddy: evaluation of empathy and objective control," in *Proceedings of the Doctoral consortium, in the scope of ACII2007 Conference*, 2007.