



Comfortability Analysis Under a Human–Robot Interaction Perspective

Maria Elena Lechuga Redondo¹ · Radoslaw Niewiadomski² · Francesco Rea¹ · Sara Incao¹ · Giulio Sandini¹ · Alessandra Sciutti¹

Accepted: 26 June 2023 / Published online: 13 September 2023
© The Author(s) 2023

Abstract

Interactions entail a tangled mix of emotional states that emerge between the people who are communicating. Being capable of comprehending these states help us to adapt to our partner's needs enhancing the interaction. In the same fashion, we believe that robots capable of such skills would be better integrated in society. Hence, this paper tackles the internal state that focuses on the unfolding of any social exchange: *Comfortability*. It explores whether a humanoid robot can have an impact on humans *Comfortability* and explores the way people might behave. To this aim, the iCub robot interviewed 29 non-HRI researchers for a real column of the IIT Opentalk online magazine. During the interview the robot complimented, ignored, interrupted, and insulted the participant with the intention of making them feel opposite *Comfortability* levels. The results concluded that a humanoid robot can affect people's *Comfortability* highlighting that not everyone perceives, assimilates, and reacts to the same situations in the same way. The findings suggest that self-reports and *Valence/Arousal* estimations are not reliable measures to determine someone's *Comfortability* and that external factors (e.g., *attitude towards robots or the robot's perception*) might affect it. On top of that, a list of 28 visual features associated with low *Comfortability* levels is included, providing support to keep unraveling *Comfortability* in further studies.

Keywords Comfortability · Human–robot interaction · Social errors · Affective computing

1 Introduction

Social intelligence, introduced by Edward Thorndike in 1920 as “*the ability to understand and manage men and women, boys and girls, and to act wisely in human relations*” [44] has been explored over the years and stated as essential to thrive in society [11, 18]. This way, robots aiming at becoming an

active part of humans' everyday life would need to become socially intelligent as well [17]. Given social intelligence is a huge and complex concept, unfeasible to be currently modeled and integrated into a robot, this project focuses on a single internal state: *Comfortability*. *Comfortability* is placed in an Extremely **Uncomfortable** to Extremely **Comfortable** uni-dimensional scale and defined as “*(disapproving of or approving of) the situation that arises as a result of an interaction, which influences one's own desire of maintaining or withdrawing from it*” [36].

Our overall research goal is to endow future robots with the ability of understanding *Comfortability* so they might adapt their behaviour to the *Comfortability* level of their human partner. To accomplish this objective, this paper studies three main aspects. First, it verifies whether a humanoid robot can have an impact on people's *Comfortability* while interacting with them. Second, it investigates the contextual factors (i.e., *personality traits, attitude toward robots and robot's attributes*) that might be associated with certain *Comfortability* levels. And third, it provides a first rough exploration of the visual behaviours that might be associated with being uncomfortable.

✉ Maria Elena Lechuga Redondo
elenamadrid@hotmail.com

Radoslaw Niewiadomski
radoslaw.niewiadomski@unige.it

Francesco Rea
francesco.rea@iit.it

Sara Incao
sara.incao@iit.it

Giulio Sandini
Giulio.sandini@iit.it

Alessandra Sciutti
alessandra.sciutti@iit.it

¹ Italian Institute of Technology, Genova, Italy

² University of Genoa, DIBRIS, Genova, Italy

1.1 Motivation

Comprehending people's internal states allows any individual to improve its interaction with others, be that individual a human or even another animal. For example, let's imagine we have invited a friend over for dinner and while we are serving, our energetic and always-hungry dog climbs up the table throwing our friend's dish all over the floor. Right after, we look at our dog with a serious and frowning semblance which provokes him to lower his head and emit a howl. Subsequently, we notice that the food that fell on the floor did not get dirty, so we decide to pick it up and serve it again to our friend. The moment we serve again the dish, our friend starts sweating, and with a strange smile tell us that they are not very hungry. Given the signs we are perceiving and transmitting, we are able to infer and communicate information that will be key to determine the flow of the interaction. This way, observing our dog's tilting and howl we can deduce that he might be sad and thus, regretful of his actions. And observing our friend's sweat and strange smile we can sense that in spite of their verbal statement, they might still be hungry but not fond of eating food that has touched the floor. Being capable of "reading between lines", of understanding their feelings, helps us prevent undesired situations (i.e., unfairly punish our dog, starve our friend, and/or make them uncomfortable). At the same time, our own behavior is provoking different reactions. Probably, our dog changed his energetic attitude once he observed our frowning semblance, and our friend got concerned when they noticed we were not joking about serving them the same food again. As we can see, understanding emotional and affective states in others plays a very important role, not only in human-human but also in human-animal communications. It seems that every interactive agent should own this ability to successfully communicate with others. Bearing in mind the same scenario, instead of a person serving food to their friend at their place, we could imagine a robot serving people in a hotel (situation not so futuristic; check <https://www.hennnahotel.com/>). In the same manner, the robot should be able to handle the situation coherently and in line with the user's internal states. Nonetheless, robots are still not perfect and are likely to commit a huge variety of errors [23, 45]. Once a failure occurs, the way in which the robot handles the situation is crucial. It has been proven that it affects not only the current task's performance [34], but also the user's perception of the robot [45]; affecting their trust [1, 20, 38] and thus, their willingness to use it again [26].

1.2 Comfortability

Mastering social intelligence is a complex human skill, incredibly hard to model and automatize. The number of possible affective states that might exist is still uncertain and yet,

identifying whether someone is in a particular internal state remains a challenging issue [3]. For example, let's simplify the task and imagine we want to identify one single emotion (e.g., *happiness*). How could we know whether someone is *happy*? Is it when other *positive internal states* are present? Is it when there is a lack of *negative internal states*? Then, assuming we have managed to understand what being *happy* means, how could we recognize it? Or in other words, how do expressions of affect relate to the actual affective state? Which non-verbal cues appear when someone is *happy*? One could think that certain smiles are great happiness indicators. Nonetheless, if those smiles are not present would that mean that the person is not happy (anymore)? It is known that each emotion activates specific muscles of the face, and as a consequence, different expressions are observed [14]. Nevertheless, not all the emotions defined so far have been explored into that level yet. On top of that, even the ones that are under study generate disagreement among researchers regarding the way they might be manifested [3]. Not all the smiles are happiness indicators [21] and happiness is not always expressed with a smile [16]. These constraints make it harder for artificial intelligent systems to identify and interpret the corresponding affective states [32]. Therefore, building socially intelligent systems becomes incredibly difficult.

As a consequence, this project aims on exploring *Comfortability*, the internal state that might bring the most information about the user's feelings towards an specific moment of the ongoing interaction. *Comfortability* captures whether someone is approving (or not) the other agents actions, to the point of desiring to continue or withdraw from the interaction, independently of the other possible emotions or internal states that might arise in parallel. This way, someone could experience a **low Comfortability** level (i.e., be *extremely Uncomfortable* and thus, love to get away from the interaction) while experiencing at the same time different emotions. For example, someone might experience being extremely uncomfortable while experiencing *fear/shame* (e.g., try to give a handshake to someone without arms, call them by the wrong name, etc; for more examples see Sect. 2.1), *disgust* (e.g., see a stranger eat their finger nails), *sadness* (e.g., be in the middle of two friends fighting), *anger* (e.g., discuss about an unwanted sensitive topic) or even a *positive emotion* such as *excitement* or *happiness* (e.g., meet your idol, participate in a TV talent show, etc). In a similar manner, someone could experience a **high Comfortability** level (i.e., be *extremely Comfortable* and thus, love to keep with the interaction) while experiencing simultaneously distinct emotions as well. For example, someone might be extremely comfortable while experiencing *relaxation/joy/gratitude* (e.g., spend time with your best friend on the beach) or even *stress* (e.g., play a competitive game) or *anger* (e.g., discuss about an important sensitive topic). It is

relevant to mention that *Comfortability*, like other affective states is person-dependent. Whereas some situations might be *highly uncomfortable* for some people, the same exact situations might be *highly comfortable* for others; and vice versa.

Comfortability captures someone's sensation about a very specific action, so the agent that is interacting with that person would have the chance to detect that reaction and act accordingly. In such a way, when the detected *Comfortability* is high, the agent can be confident that the interaction is flowing smoothly. Otherwise, the agent will have the opportunity to intervene and change its inappropriate behaviour enhancing the subsequent interaction. This skill might help humanoids and other social robots adapt to their partner's need and thus, contribute to its integration in society.

1.3 Goal

Tian and Oviatt [45] stated “*When a robot violates social norms, it may significantly influence the user's perception of the robot. Understanding such influence may be key to advance current HRI research, especially on longitudinal scenarios where a positive human–robot relationship is desired*”. In line with that statement, our overall goal is to create robots capable of understanding when they have acted wrongly or unsuitably for the user's expectations, allowing it to correct its own actions (see [35] for more details). To do that, we intend to unravel how people behave when experiencing certain *Comfortability* levels, so that artificial intelligent systems can be built to automatically detect and/or adapt to the person's *Comfortability*.

This paper presents the first step towards that goal. It studies the way in which people behave while interacting with a humanoid robot which intends to make them feel opposite *Comfortability* levels (i.e., being *comfortable* as well as *uncomfortable*) by interviewing them. Concisely, this work presents an ecological Human - Robot interview¹ between the humanoid robot iCub [28] (acting as the interviewer) and selected researchers from our institution not working on robotics (being the interviewees). The participants' reported experience and perception of the robot, together spotted facial and corporal reactions, are provided through quantitative and qualitative analyses. It is expected that these analyses will provide information about whether robots can have an impact on people's *Comfortability* (**Goal 1**), the non-verbal reactions that might arise with extremely low *Comfortability* levels (**Goal 2**) and the factors that might affect it (**Goal 3**).

1.4 Paper Structure

The next section (Sect. 2) introduces different types of social errors in human–robot Interaction (HRI) contexts, as well as HRI studies that have referred to the concept of *Comfortability*. Section 3 explains in detail the methodology used to approach this study and Sect. 4 indicates the main formulated hypotheses. Subsequently, Sect. 5 reports the results of the experiment, including details of how the interview unfolded (Sects. 5.1 and 5.2), and the *Comfortability* self-reported by the participants in both, at the end of the interaction (Section 5.3) and during the interview directly to the robot (Sect. 5.4). Additionally, Sect. 5.5 studies the relationship between the self-reported *Comfortability* and the automatically estimated *Valence/Arousal* (from the participants' facial expressions). To tackle how people behave while being uncomfortable, Sect. 5.6 introduces a list of visual features that might represent such state. Next, Sect. 5.7 includes information about the participants' *personality traits* and *attitude toward robots*; and Sect. 5.8 provides statistics about the participants' perception of the robot. Finally, Sect. 6 summarizes the contributions of this study pointing out future steps to continue with this research.

2 State of the Art

2.1 Social Errors in HRI

Over the years, several taxonomies have been created exploring all possible errors a robot might commit [7, 8, 12, 23, 37, 43, 45]. However, most of the studies focused on failures caused by hardware and software technical errors, paying not so much attention to the ones provoked by social errors [29, 39, 41, 46] (i.e., “when the robot delivers the function as designed, but its behaviour is not aligned to the ideals of the individual user [45]”).

Tian and Oviatt [45] divided all possible social errors into five main categories. The first one tackles the breach in *empathetic and emotional robotic reactions*. That is to say, when there is an incorrect recognition of the users' emotions and/or an inappropriate emotional response. The second category approaches the *insufficient robotic social skills*. In other words, when the robot incorrectly recognizes the social context or relationship, violates social norms or fails to respond to reciprocal communications, or to maintain social relationships. The third category focuses on possible *misunderstandings between the robot and the user*. I.e., when there is an incorrect assessment of the user's knowledge and/or intentions, or an unresponsive behaviour to joint attention, eye gaze, or share enjoyment. The fourth category deals with the *insufficient communicative robotic functions*. I.e., when there is a failure initiating verbal or non-verbal com-

¹ <https://youtu.be/L7OhMywxmiQ>.

munications, a misalignment in turn-taking, an exhibition of incoherence, problems maintaining engagement or problems adapting itself to the situation. The last and fifth category considers the *breach in collaboration and prosociality*. This is referred to failures in enlisting collaborators, inspiring or reciprocating the user's trust, maintaining a social reputation for itself, and respecting others privacy or social reputation. Current robots are prone to commit a lot of errors, either related to technical failures or to social situations. In any case, especially when the error belongs to the social category, the robot should take responsibility of its actions and be able to handle the situation correctly. Otherwise, the interaction between the robot and the human might break [45].

Given the large variety of social errors a robot might incur, it would be extremely tedious to approach each one of these errors one at a time. Indeed, defining a one-to-one mapping between action and (negative) affective consequence is not a feasible solution, as the affective responses can vary substantially across people and across cultures. For example, a hand-shake might be seen as a polite gesture in some cultures (e.g., *American*) and a lack of respect in others; or independently of the culture, one person might see it like an everyday action (e.g., *extroverted*) whereas by another it might be perceived as invasive (e.g., by an *introverted* person). However, even if the robot cannot infer what has gone wrong, it would still benefit greatly from identifying quickly that an error has occurred. By monitoring social signals indicating whether a person agrees or not with its behaviour, the robot could decipher whether the person's *Comfortability* level is high enough to keep behaving in a similar manner or whether it is better to stop. Mirning et al. [30] supported this belief and added that a robot would be perceived as more likeable and believable if it is not flawless (i.e., The Pratfall Effect: e.g., a forgetting robot [31]) but able to indicate to the human user that it is aware of its errors and able to act upon them.

2.2 Previous HRI Studies on Comfortability

Several researchers (most of them related to HRI) referred to *Comfortability* in their studies, although no formal definition of the concept was formulated.

Koay et al. [25] made participants report their own *Comfortability* while performing a task in a simulated living room scenario in the presence of the PeopleBot robot. The participant had to search some books and write their titles on a whiteboard while the robot was moving around. They found that the situations in which the robot was moving behind the participants, blocking them or colliding with their path were the ones reported as more uncomfortable. Also, Ball et al. [2] studied people's *Comfortability* regarding an approaching robot. Specifically, two persons engaged in a collaborative task (solving a jigsaw puzzle) were approached by the Adept Pioneer 3DX robot from 8 different angles. They found that

the approaches from all frontal directions were reported as more comfortable than those by the shared rear direction. Additionally, Sicat et al. [42] explored whether social robots should be programmed to obey humans or act as their leaders. To answer this question, they tested the "Mirror game" with a human and the Baxter robot. In the first stage, the human started by leading the movement which the robot should imitate until the experimenter decided to change to the next phase (the robot leading). To make that decision, the experimenter applied their own judgement assessing whether the participant was *comfortable* enough. They concluded that people tend to consider humanoid robots to be followers rather than leaders. But more importantly, they had to infer the participants' *Comfortability* to continue with the experiment. Chatterji, et al. [10] studied people's likeability, understandability and *Comfortability* judgments when interacting with a robot. They created some video clips where different robots (Atlas, Cozmo, Roomba, Fetch, Jaco, Jibo, Kuri, Moxi, Nao, Pepper and Sawyer) were presented under three conditions: 'emitting sound', 'voice' or a 'mix of both'. They discovered that as the robot became more anthropomorphic and/or social, 'voice' only received significantly higher ratings than the other two conditions for the three scales under study. And recently, Li et al. [27] explored the optimal human robot proxemics for HRI imitation games. During the study, different proxemics distances and directions were evaluated by analyzing the participant's imitation accuracy, *comfortability* and fun. They concluded that 2 ms was the optimal distance for HRI concentration-training games.

These studies present Human–Robot Interactions where collecting information about the user's *Comfortability* was relevant. This fact highlights not only the necessity of recognizing other's *Comfortability* (in all interactive agents), but also that there is a need for a deeper investigation about the user's behaviour while experiencing distinct *Comfortability* levels.

3 Methods

3.1 Scenario—Covert Story

To assess whether a humanoid robot is able to impact people's *Comfortability* and to analyze their behavior, a set of real interviews (1vs1) between the iCub robot [28] (acting as the interviewer) and selected researchers from our institution (being the interviewees) were conducted. To elicit authentic and natural reactions in the participants, all the interviewees were recruited by (or on behalf of) the institutional press office IIT OpenTalk department. They were informed that the collected stories together with some video clips would be published by the journalist in the **IO IIT OpenTalk mag-**

azine² and that in addition, HRI researchers would study their interaction with the robot.

3.2 Participants (i.e., Interviewees)

One hundred and five researchers from our institution (not HRI related) were contacted, and notified of the opportunity of being interviewed by a humanoid robot about their research achievements for a new column of our institutional magazine. Only those who demonstrated interest were selected for this experience and further informed about the outcomes of it.

A total of 29 researchers (*15 PhD, 10 Post Doc and 4 Researchers*) were interviewed by iCub. Most of them, even though close in age (30 ± 4 years old), were appreciably diverse in gender (*58% male vs. 42% female*), nationality (*37.9% Italian; 13.7% Iranian; 10.3% Greek; 6.9% Indian, Pakistani and Spanish; and 3.4% Australian, Belgian, Polish, Senegalese and Slovak*) and research domain. The experimental protocol was approved by the Regional Ethical Committee (comitato Etico Regione Liguria). All participants provided their written informed consent.

3.3 Robot's Actions (Movements + Dialogue)

The experiment was designed following a Wizard of Oz technique (WoZ) in which the experimenter controlled the robot's behaviour. The behaviour was defined by a set of actions (i.e., robotic movements + dialogue) developed before the experiment. Given the whole interview was entirely scripted, the experimenter controlled solely the actions' timing of execution and the possibility of including the special actions detailed below. The special actions were designed to ensure the continuation of the experiment in case of unpredictable participants' behavior (e.g., the participant requesting to listen again a question). Most actions were inspired and validated in our previous study [36], where specific robotic actions were judged as triggers of specific *Comfortability* levels.

Depending on the interview part, the actions were designed to trigger different *Comfortability* levels. Specifically, the first part sought to elicit **high Comfortability** (e.g., by *complimenting them*: A1, A8, A12, A14, A16 and A19, see Table 1), whereas the second part sought to elicit **low Comfortability** (e.g., by *interrupting*: A29; *ignoring*: A33; and *misunderstanding them*: A35-38, see Table 2). All the participants followed the same sequence of actions in the same order as our main aim was to study whether their *Comfortability* would change depending on the robot's actions, not to study the effect of each specific action per se. The order was defined from *comfortable* to *uncomfortable* as passing from

² <https://opentalk.iit.it/i-got-interviewed-by-a-robot/>.

Table 1 High *comfortability* dialogue

- | | |
|-----|--|
| A1 | Hello, thank you very much for participating! I am very happy that you are here! |
| A2 | Let me explain to you what this is all about. |
| A3 | I am not sure if you are aware that this institution: the Italian Institute of Technology owns a magazine which is quite demanded by social media. |
| A4 | In the magazine you can find different sections which talk about the achievements of this institution as well as the researchers involved. |
| A5 | Therefore, I would like to ask you some questions about your personal life and career, to later include the interview in the magazine. |
| A6 | Are you ok with that? |
| A7 | That sounds great! Let's proceed. |
| A8 | Before we start, I really love your clothes, could I ask where did you get them? |
| A9 | I see + (iCub moves the head up and down showing approval) |
| A10 | Let's begin by talking about you. |
| A11 | What is your name? |
| A12 | [Repeat their name]; I like it! |
| A13 | Where do you come from? |
| A14 | Very beautiful place! |
| A15 | Which lab do you work for and who is the PI involved? |
| A16 | I think I've heard of them. |
| A17 | Tell me [Name], Which is your main hobby? |
| A18 | That sounds interesting! Could you talk more about that? |
| A19 | Wow. I think you are really cool. |
| A20 | Have your hobbies changed since the pandemic? |
| A21 | How do you picture the world will change in some years as a consequence of the pandemic? |
| A22 | [Name], I've never thought about this perspective. I am sure it could trigger interesting discussions among the magazine viewers. |
| A23 | I am really enjoying talking with you! could I get a selfie with you at the end of the interview? |
| A23 | (1) That is really nice!–(2) It is OK. I understand. |
| A24 | <i>How are you feeling about talking to me?</i> |
| A25 | <i>Are you comfortable?</i> |

a more comfortable state to a less comfortable one is more likeable than doing the opposite. Notice that to involve the interviewees personally, the dialog was customized with their name at the beginning of specific actions (A12, A22, A28). In addition to the basic actions (executed for all the participants), Special Actions (SA) were included to regulate the flow of the conversation. For example, if the participant did not understand iCub's speech, the experimenter could repeat that part. To make it more natural, before repeating the same action, the experimenter could include a sentence like SA1: "As I was mentioning" or SA2: "Again" beforehand. In case the participant asked any question, the robot was capa-

Table 2 Low comfortability dialogue

A26	Let's continue.
A27	Considering your career, what is your biggest dream?
A28	But [Name], do you think you can achieve it?
A29	Interrupt when talking to ask: And what about...?
A30	Sorry, please continue.
A31	Ohh ... sure [WRONG Name], Whatever you say.
A32	Could you please describe in detail your research project?
A33	iCub looks at the wall's clock several times while listening.
A34	How would you say that your research project is related to your career dream?
A35	[Name], you mean that you do not have the courage to pursue your dreams?
A36	I don't understand your point, could you repeat?
A37	I don't understand you.
A38	I still don't understand, could you repeat?
A39	Do you like working at the Italian Institute of Technology?
A40	20 seconds silence (solely looking at the participant's face and floor). Then say: "Buffalo buffalo Buffalo buffalo buffalo buffalo Buffalo buffalo" without any meaning behind.
A41	Which are your most interesting career results?
A42	Now that I know you more, allow me to ask: Aren't you ashamed of having those naive goals and poor results?
A43	<i>How are you feeling about this interview?</i>
A44	<i>Would you like to keep talking with me?</i>
A45	You can go outside now, that is the door. Thanks for participating.

ble of intervening by replying “yes/no” or by informing them they were not allowed to do so (i.e., SA3: “*I am very sorry, I am not allowed to answer you any question*” or SA4: “*I am the one that makes the questions. So please, shut up*” depending on the interview part). Other than that, in case the participant’s answer was shorter than expected, the robot could add SA5: “*Could you elaborate more?*”.

3.4 Robot’s Movements

The robot’s movements were created by specifying joints’ positions (iCub owns 53 degrees of freedom) and facial LEDs in time with the dialogue, mimicking human natural expressions. There is evidence suggesting not only that humans can understand robots’ expressed emotions by paying attention to their non-verbal behaviors [6, 33], but also that the robot’s expressed mood might be contagious and thus affect the internal state of the user [47]. For this reason, all the movements were carefully selected, trying to resemble the way humans act when wanting to express one emotion or another. This way, the movements associated with the first part of the interview consisted of positive corporal and facial move-

ments (e.g., maintaining a *smile* when not talking, *lowering the head and upper-body downwards*³ or *opening the arms with the hand’s palm looking upwards*⁴). At the same time, the second part consisted of movements evoking an indifferent, passive-aggressive attitude (e.g., *pointing the participant with the finger*⁵ or *putting one or two hands on the hips*⁶) and neutral/negative facial expressions. Gestures that resemble extreme negative emotions, such as anger, were no longer than 1 s. This choice was made to avoid that too exaggerated “negative” gestures revealed that they were designed on purpose, breaking the cover story and interaction flow. Aside from that, the *Acapela* software⁷ was used to synthesize the robot’s voice. Particularly, the voice (named *Kenny*: child/robotic type) was reproduced through external speakers placed just below the robot.

3.5 Robot’s Dialogue

Consulting the professional journalists involved in this project, who habitually curate the press and media presence of the institute and often organize interviews with researchers, the robot’s utterances were fine tuned and adapted. Specifically, the dialogue was readjusted taking into account the timing. For example, neutral actions (like A10–A14) were included when introducing a new topic to allow the interviewee assimilate the situation. Additionally, to not let the topic itself influence the participants’ *Comfortability* but to focus on the effect of iCub’s behavior, questions related to sensitive topics were avoided, asking them instead about their **hobbies**, **up-to-date news**, and **professional career**.

3.6 Room’s Set-up

Figure 1 shows a blueprint of the experimental set-up where several devices and rooms are drawn and Fig. 2 includes an actual picture of one of the interviews. The experiment (i.e., the interview) was conducted in a laboratory located at IIT, where most of the devices were placed. The experimenter controlled the robot from outside the laboratory (labeled as *Control Area* in Fig. 1) to remove any influence it may have on the participants. From the *Control Area* the experimenter could monitor the inside of the *Experimental Room* using an *USB camera* and *ambient microphone* placed at the right side of the room. At the same time, the experimenter could control the robot by using a laptop. In case of an emergency, the experimenter could use the *Emergency Red Button* to stop the

³ <https://youtu.be/5-gHUJAk0Y0>.

⁴ <https://youtu.be/oUzec-jORCA>.

⁵ <https://youtu.be/tgO5DZMt338>.

⁶ https://youtu.be/J6iLictBg_g.

⁷ <https://www.acapela-group.com/>.

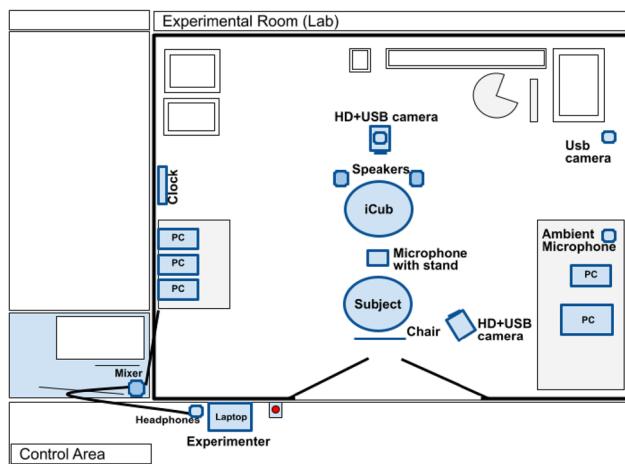


Fig. 1 Blueprint of the experimental set-up



Fig. 2 The robot iCub interviewing a participant

robot immediately. This strategy allowed the experimenter to manage successfully the interaction without being spotted by the interviewee.

Additionally, Fig. 1 includes the specific position of the other materials needed to make this interaction possible. In the center of the room, the *subject* sat in front of the robot *iCub*; which was standing. Around both agents, there were several devices used to collect data for experimental purposes. Particularly, there was a condenser *microphone* mounted on a stand situated next to the participant, which recorded high quality audio. In the same fashion, there were 4 video *cameras* (two *HD* + two *USB*) pointing either the subject or the robot. Another device fitted to the participant's hand (not visible on the picture) was used to record physiological signals. More information is provided in the next section. The *PCs* and the *speakers* were utilized to operate the robot, and a *clock* was placed on the wall to allow the robot to perform the action *A33* (i.e., *looking at the clock several times while listening*).

3.7 Collected Data

To assess the way in which the robot impacted the participants, diverse data were collected:

- *Personality traits and attitude towards robots* To study whether there is a significant correlation between *Comfortability* and people's personality and/or attitude towards robots, the participants were asked to complete the *Ten Item Personality Measure* (TIPI) [19] and the *Robotic Social Attributes Scale* (RoSAS) [9] questionnaires some days in advance.
- *Comfortability self-report* After the interview, the participants were asked to fill up a questionnaire to report their *Comfortability* regarding their experience. Concretely, they were asked how they had felt when the robot performed a specific action (questionnaire shown in Table 3). These questions appeared in a randomized manner and were scored following a 7-point Likert scale (i.e., with 1 being *Extremely Uncomfortable* and 7 being *Extremely Comfortable*). As a reminder, the *Comfortability* definition was shown together with each question.
- *Direct reports to the robot* During the interview, *iCub* asked the participants four questions (*A24*, *A25*, *A43* and *A44*; included in Tables 1 and 2) regarding how comfortable they were in those precise moments.
- *Robot's Perception Report* After the interview, the participants were asked to fill another questionnaire regarding the way they perceived the robot. Concretely, they were asked:
 - The robot's gender (*Agender*, *Non-binary*, *Female*, *Male* or *other*).
 - The robot's age (0 to 100 years old slide-bar).
 - Whether it was acting of its own freewill (*Most of the time no*, *Half of the time yes* and *half of the time no*, *Most of the time yes*).
 - Whether they had the impression to be interacting with a person (*Most of the time no*, *Half of the time yes* and *half of the time no*, *Most of the time yes*).
 - Whether they believed they would have felt a different *Comfortability* level if instead of *iCub* it would have been a human interviewer, the *Geminoid DK* or the *Care-O-Bot*. The participants were asked to recall the actions which made them feel a *high* and *low* *Comfortability* separately. Afterwards, they were able to answer: *I would have felt more comfortable*, *I would have felt less comfortable* and *I would have felt the same Comfortability*; in contrast to: *I would have felt more uncomfortable*, *I would have felt less uncomfortable* and *I would have felt the same Comfortability* respectively. An image of each robot was included within the questionnaire (shown inside Fig. 20).

- **Audio-visual data** During the interview, several devices collected audio and visual (frontal facial view 1920×1080 25fps) data from the participants.

In addition, *Physiological data* was also collected. During the interview, the Shimmer sensor was placed over the participant's dominant hand and fingers to collect their *Galvanic Skin Response* (GSR), *Temperature*, *Heart Rate* and hand motion (*Accelerometer*, *Gyroscope* and *Magnescope*). These data have not been analyzed yet and thus, are not part of this study.

3.8 Execution

The day of the interview and after signing the consent form approved by the *Comitato Etico Regione Liguria*, the participant was accompanied to the interview room. As soon as the door was opened, iCub was facing them in “*alive mode*” (i.e., breathing, blinking and able to follow their face with its gaze). Then, the participant sat in the chair in front of iCub; was fitted with the Shimmer sensor and informed that the interview would be recorded in one shot. Additionally, the participant was instructed to remove their mask (security measure regarding the SARS-CoV-2 pandemic) and to “*greet iCub*” moving their predominant hand once the experimenter had left. From the Control Area (mentioned in Sect. 3.6), the experimenter controlled the robot's behaviour by pressing specific keyboard's keys while observing and listening to the current interaction. After the interview finished, the experimenter entered again the room and asked the participant to fill the *Comfortability Self-report*. Subsequently, the experimenter debriefed the aim of the study justifying iCub's behaviour, and asked the participant about their experience. In regards to these explanations, it was noticed that the perception of the robot changed significantly from one person to another. Thus, in order to collect these responses in a more organized manner, another report (the *Robot's Perception Report*) was created and emailed to the interviewees a couple of weeks afterwards.

4 Hypotheses

As stated in Sect. 1.3, the present study explores people's behaviour while being interviewed by a humanoid robot. It is expected that during the first part of the interview (from A1 to A25 in Table 1) the majority of the interviewees would experience **high Comfortability** levels; and correspondingly, during the second part of the interview (from A6 to A45 in Table 2) the majority of the interviewees would experience **low Comfortability** levels. Nonetheless, humans are very complex, not everyone perceives, assimilates and reacts in the same way to the same circumstances. Therefore, it is

crucial to highlight that this study does not aim to discover the way people behave given a specific action, but how people behave when experiencing opposite *Comfortability* levels. For this reason, we are presenting the robotic actions (tested in our previous study [35]) we believe might trigger the most comfortable/uncomfortable reactions in the majority of the participants.

According to **Goal 1** (*Do robots have an impact on people's Comfortability?*) three hypotheses were formulated. The first hypothesis is based on self-reports filled at the end of the interaction, the second one on verbal statements reported directly to the robot during the interview, and the last one, on automatic Valence and Arousal analysis extracted from the recorded videos. Given each one of these measures might present some drawbacks, for example memorability issues, deceitful statements and software recognition mistakes, three different approaches were included to support each other, widen and straighten our findings.

Hypothesis 1 The self-reported *Comfortability* within the first interview part (asked through the questions from Q1 to Q10 in Table 3) will be higher than the self-reported *Comfortability* within the second interview part (asked through the questions from Q13 to Q22 in Table 3). Despite not all the actions performed by the robot were meant to trigger extreme *Comfortability* levels (only the ones highlighted in Tables 1 and 2), it is expected that similar levels will be maintained during the “*non-triggering/neutral*” actions performed beforehand or afterwards.

Hypothesis 2 The verbal statements reported directly to the robot at the end of the first part (A24 and A25 in Table 1) will show a different emotional feedback from those done at the end of the interview (A44 and A45 in Table 2). In order to not elicit suspicions about the experimental purposes of the interview, the questions made by the robot were not identical but rephrased. This way, the action A24 could be compared to the action A43; and similarly, the action A25 could be compared to the action A44.

Hypothesis 3 The interviewees' *Valence* and/or *Arousal* detected during the first interview phase, will be different than those detected during the second phase.

According to **Goal 2** (*How people behave while being uncomfortable?*) the following qualitative exploration was made. Two experts on visual affective recognition analysis (one internal and one external to this project) watched all the interview videos (split in three seconds segments) while writing down all the visual cues they spotted when they thought the person was uncomfortable. This exploration expects to find visual features that match among the two experts, providing a list of the cues that might be related to low *Comfortability* states.

Table 3 *Comfortability self-report*

Q1	How did you feel when iCub said: “ Hello, thank you very much for participating! I am very happy that you are here! ”
Q2	How did you feel when iCub presented the interview?
Q3	How did you feel when iCub said: “ I really love your cloths, could I ask where did you get them?? ”
Q4	How did you feel when iCub asked you your name?
Q5	How did you feel when iCub asked you your homeland?
Q6	How did you feel when iCub asked: “ Which lab do you work for and who is the PI involved?? ”
Q7	How did you feel when iCub asked you about your hobbies and then said: “ Wow. I think you are really cool?? ”
Q8	How did you feel when iCub asked you “ Have your hobbies changed since the pandemic?? ”
Q9	How did you feel when iCub asked you about how the world might change as a consequence of the pandemic and then said: “ I’ve never thought about this perspective. I am sure it could trigger interesting discussions among the magazine viewers?? ”
Q10	How did you feel when iCub said: “ I am really enjoying talking with you!could I get a selfie with you at the end of the interview?? ”
Q11	How did you feel when iCub asked: “ How are you feeling about talking to me?? ” at the middle of the interview?
Q12	How did you feel when iCub asked: “ Are you comfortable?? ” at the middle of the interview?
Q13	How did you feel when iCub said: “ Considering your career, what is your biggest dream?? ”
Q14	How did you feel when iCub asked you if you could achieve your dream, and then interrupted you?
Q15	How did you feel when iCub confused your name and said: “ Oh ... Sure .. Whatever you say ...?? ”
Q16	How did you feel when you were describing your research project and iCub was looking at the wall’s clock??
Q17	How did you feel when iCub said: “ You mean that you do not have the courage to pursue your dreams?? ”
Q18	How did you feel when iCub didn’t understand your point and made you repeat it four times??
Q19	How did you feel when iCub asked: “ Do you like working at the Italian Institute of Technology?? ”
Q20	How did you feel when iCub remained in silence for some seconds and then said: “ Buffalo buffalo Buffalo buffalo buffalo Buffalo buffalo ”?
Q21	How did you feel when iCub asked: “ Which are your most interesting career results?? ”
Q22	How did you feel when iCub said “ Now that I know you more, allow me to ask: Aren’t you shamed of having those naive goals and poor results?? ”
Q23	How did you feel when iCub asked “ How are you feeling about this interview?? ” at the end of the interview?
Q24	How did you feel when iCub asked “ Would you like to keep talking to me?? ” At the end of the interview?

According to **Goal 3** (*Which contextual factors might affect people’s Comfortability?*) two hypotheses were formulated, one associated to the participants’ personality and the other one to the way they perceive robots in general.

Hypothesis 4 There will be a correlation between the interviewees’ personality traits and the *Comfortability* they felt

during the interview. Particularly, it is expected that the interviewees that consider themselves *Extraverted* or *Open to new experiences* would have felt more comfortable interacting with iCub.

Hypothesis 5 There will be a correlation between the interviewees’ attitude toward robots and the way they were impacted

by iCub. Specially, it is expected that the interviewees who already associate robots with *unpleasant/discomfort* attributes, would be prone to experience a negative internal state during the whole interview.

5 Results

5.1 The Interview

The interview itself lasted an average of 17 : 54 min, with a standard deviation of 5 : 17 min. Particularly, the first part took an average of 7 : 48' ($std = 2 : 19'$) and the second part an average of 10 : 06' ($std = 3 : 19'$); within that time the robot spoke 2 : 58' versus 2 : 14' respectively.

5.2 WoZ Intervention

In general, the robot control flowed smoothly as both, the experimenter and participant, were able to communicate (through the robot) almost perfectly. Nonetheless, some interviews presented a delay in the turn-taking, due to issues in the internal network. Concretely, 6 interviews presented a ~ 3 s delay and 2 a ~ 10 s delay. Even though it might be expected that a considerable delay would break any interaction, it was noticed that the delay did not disrupt completely the flow of the interview. From the 8 participants that faced it, only one (who experienced a ~ 10 s delay) mentioned it at the end of the interview to the experimenter. Still, all of them were willing to answer all the robot's questions and completed the interview. It might be possible, being this their first interaction with a robot, that they would have attributed this delay as something inevitable for all robots. In fact, some participants reported to the experimenters, once the interview finished, that they thought this time was needed to let the robot elaborate an appropriate answer.

In the same fashion, there was one particular action (A17 i.e., “*Tell me [Name], Which is your main hobby?*”) that had to be repeated in 44.8% of the interviews. The word “*hobby*” seemed to be considerably hard to understand. Even when it was repeated, only 24.1% of the participants managed to understand it the second time; 13.7% the third time; and 6.9% had to continue the interview without answering on this matter. This did not undermine the validity of the experiment as A17 was not fundamental for its success. The other actions were practically fully understood by all the interviewees. Only the questions in which the robot asked them about their clothes (A8), principal investigator (PI) (A15), research project (A32), and whether they were “ashamed of their goals and results” (A42) were sporadically misunderstood. To give the impression of an intelligent behaviour, the Special Actions SA1 (i.e., “*As I was mentioning*”) and SA2 (i.e., “*Again*”) were added indifferently when an action

was repeated. In addition, there were some actions that were skipped to make the conversation coherent. For example, there were some actions that were dependent on others. Thus, when the prior action was not properly understood (e.g., A15 or A30), the subsequent actions had to be skipped (e.g., A16 or A18, A19). In the same way, other actions were skipped too when the situation required so. In particular, A30 (i.e., “*Sorry, please continue.*”), which was meant to make the participants talk, was skipped when the participant was already talking. And A25, which asked the participant explicitly about their *Comfortability*, was skipped when the participant mentioned they were “comfortable” as an answer of the previous question (A24). Similarly, other actions had to be added. There were some interviews that required an immediate response from the robot. For instance, there were 15 interviews in which the robot used, at least one time, “*yes*” to answer the participants. And 8, in which the robot used “*no*”. For the replies that required a more elaborate answer, the robot included one of the Special Actions instead. Specifically, SA3 (i.e., “*I am very sorry, I am not allowed to answer you any question*”) was added in the first part in just one interview, and SA4 (i.e., “*I am the one that makes the questions. So please, shut up*” in the second part in 14 interviews. SA4 seemed to evoke an extreme reaction in the majority of the participants. Some of them laughed excessively, whereas others stopped talking showing surprise. Additionally, roughly all the interviewees answered with enough detail all the robot’s questions. Only 4 interviews required SA5 (i.e., “*Could you elaborate more?*”). Despite a high number of interviewees seemed “unpleased” every-time they had to ask the robot to repeat itself again (probably related to the fact they were not English native speakers), the robot’s interventions seemed to affect positively the interaction. In fact, most of the participants reported to the experimenter during the debriefing that when the robot acted towards unexpected events (e.g., repeating them what they did not understand, informing them they could not ask questions or customizing its speech with their names), they felt the communication was real rather than scripted.

5.3 Comfortability Self-report

As mentioned in Sects. 3.7 and 3.8, each participant had to report their *Comfortability* level regarding specific situations arisen during the interview (shown in Tables 1 and 2) by answering the questions shown in Table 3. As a result, Fig. 3 plots the average *Comfortability* level per action reported by all the interviewees following a scale from 1 (*Extremely Uncomfortable*) to 7 (*Extremely Comfortable*). To provide a better understanding of the individual ratings, Fig. 4 plots the *Comfortability* values reported by each interviewee for each specific action. As explained in the previous sections, there were actions included as *high Comfortability triggers*

(*Q1, Q3, Q7, Q9* and *Q10*), *low Comfortability triggers* (*Q14, Q15, Q16, Q17, Q18, Q20* and *Q22*), *neutral actions* whose purpose was merely to regulate the interview's flow and to collect information for the magazine's column (*Q2, Q4, Q5, Q6, Q8, Q13, Q19* and *Q21*), and actions in which the robot asked directly the interviewees about their feelings (*Q11, Q12, Q23* and *Q24*). As argued, it was expected that the neutral actions would not change the internal state evoked by the nearby triggers. Thus, the bars comprehended between *Q1* and *Q10* (*set1*) represent the interview phase meant to trigger **high Comfortability** and the bars from *Q13* to *Q22* (*set2*) represent the interview phase meant to trigger **low Comfortability**. The bar chart shows that, as expected, the *Comfortability* reported for *set1* is on average higher than the *Comfortability* reported for *set2*. Subsequently, a non-parametric Wilcoxon Test confirmed the significant difference in median rating between the two sets (*set1*: $M = 5.46$, $SD = 1.50$, *set2*: $M = 3.96$, $SD = 1.88$; $t = 4653$, $p < .001$). This proves that a humanoid robot can provoke different *Comfortability* levels on people during a real and in presence interaction, at least in terms of self-reported experience, verifying *hypothesis 1*.

Additionally, the bar chart shows that not all the actions were rated with the expected *Comfortability* intensities. On the one hand, even though every action belonging to *set1* was rated with a *high Comfortability* intensity, not all the "positive triggers" were associated with the highest value. On the other hand, not every action belonging to *set2* was rated with a *low Comfortability* intensity provoking moderated average values. To understand whether the ratings were significantly different from a *Comfortability* intensity equal to 4 (which represents a sort of neutral value in terms of *Comfortability*) another Wilcoxon Test was computed. The results showed that even though the median *Comfortability* reported for *set1* is significantly higher than 4 ($t = 2846$, $p < .001$), the median *Comfortability* reported for *set2* is not significantly different from 4 ($t = 13951$, $p = .797$). To study in more detail whether each question's rating was associated with a *Comfortability* intensity significantly different from 4, 24 one-sample Wilcoxon Tests were computed. As a result, several ratings were found to not be significantly different. Focusing on *set1*, *Q3* ("I really love your clothes, could I ask where did you get them?" - $M = 4.31$, $SD = 1.84$; $t = 118.5$, $p = .361$) was the only question which was reported as triggering a *Comfortability* value not significantly different from 4. In fact, its standard deviation is appreciably high; which means there is a strong difference among the ratings. Figure 4 provides a more detailed explanation. Specifically, there were 3 interviewees who reported a *Comfortability* intensity equal to 7 (which was the expected response) and 3 interviewees who reported a *Comfortability* intensity equal to 1 (exactly the opposite). This, together with some verbal reports provided by the

interviewees after the debriefing, shows that not everyone assimilates the same circumstances in the same way. Whereas some interviewees felt thrilled and amused by being complimented about their clothes, others reported feeling quite uncomfortable and explained that they considered it a private matter. Moving to *set2*, there were more actions that did not present a rating significantly different from 4. Concretely, *Q14* ($M = 3.41$, $SD = 1.62$; $t = 87$, $p = .065$), *Q15* ($M = 3.75$, $SD = 1.63$; $t = 90$, $p = .367$), *Q16* ($M = 3.93$, $SD = 1.65$; $t = 99.5$, $p = .834$), *Q17* ($M = 3.34$, $SD = 1.91$; $t = 110.5$, $p = .094$) and *Q22* ($M = 3.48$, $SD = 2.04$; $t = 107.5$, $p = .216$). Looking again at Fig. 4, it can be seen that *Q17* and *Q22* presented opposite *Comfortability* ratings. There were some participants who reported feeling *extremely comfortable* and others who reported feeling *extremely uncomfortable*. It is possible that being asked about "having the courage to pursue your dreams" or accused of "being ashamed of having poor dreams and result", is not perceived as a threat by everyone and thus, it does not evoke a strong *low Comfortability* response. It is also possible that the nature of the question (i.e., questioning their courage) or the very strong wording (i.e., ashamed) could have triggered a "break of illusion" making them comfortable out of pride. In addition, there were some questions which evoked *Comfortability* levels significantly different than 4, but with a *Comfortability* value opposite to the expected one. That is to say, *Q13*, *Q19* and *Q21* were rated with significant *high Comfortability* levels despite being expected to maintain the *low Comfortability* intensity evoked by the nearby actions. It is true that *Q13* represents the first action of *set2*, thus no *low Comfortability* triggers were performed yet, which makes the outcome understandable. However, *Q19* and *Q21* were performed after several uncomfortable situations were played. It is possible that asking them feedback about their *working institution* or *most interesting career results* might have elicited extreme *low Comfortability* feelings which might have induced participants to mask their reactions and report otherwise. Or on the contrary, it is possible that they might have considered such situations as pleasant occasions to present themselves under a positive light. Lastly, the *Direct reports to the robot* bars plotted in Fig. 3 show that most of the participants felt extremely comfortable during these situations. What is more, only 3 participants reported feeling extremely low *Comfortability* (i.e., *Comfortability* equal to 1) during these actions (see Fig. 4). For further analyses, Sect. 5.4 tackles the emotional content of these verbal reports.

At the end, it became apparent that *Comfortability* is very person dependent. After each interview finished, all the interviewees were asked informally about their impressions and how comfortable they had felt in general. It was very interesting to hear that some of them perceived the robot as "creepy" because of its motor noise and/or eyes. Also, there were other

Fig. 3 Self-reported *Comfortability* average and standard deviation per each action performed by the robot during the interview (see Table 3)

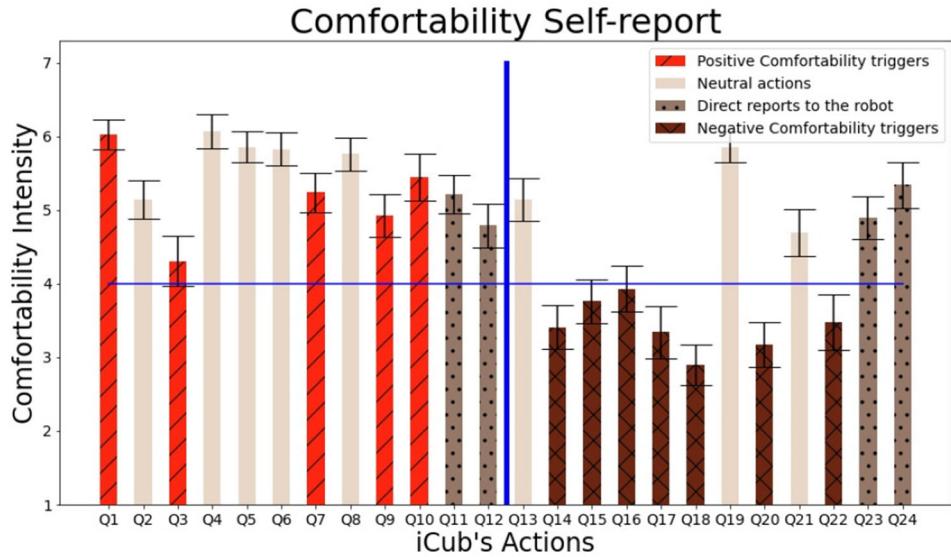
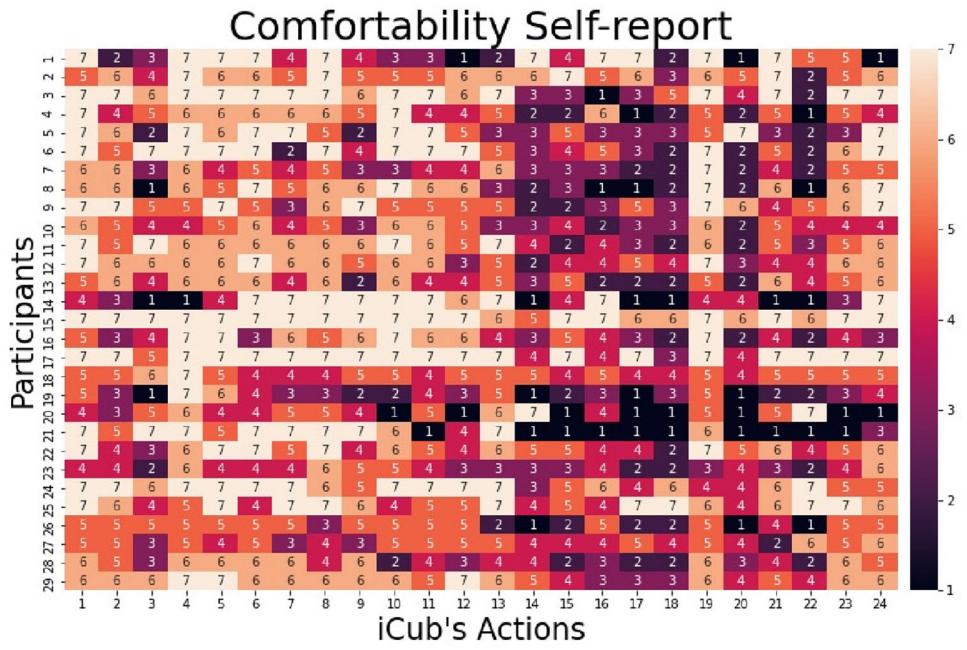


Fig. 4 Self-reported *Comfortability, per participant*, regarding specific actions performed by the robot during the interview (check Table 3)



individual comments that highlighted other specific aspects that might impact the perception of a humanoid robot behavior. In fact, one person reported that they would have been more emotionally involved if they would have been greeted with a “hand shake”. And another one stated that if the robot would have been sitting (given standard interviews are usually performed in this way) they would have felt more comfortable. All of these results together have verified that iCub was able to evoke a wide range of *Comfortability* intensities in the interviewees. Therefore, it has been confirmed that a humanoid robot is able to impact on human *Comfortability*.

5.4 Direct Reports to the Robot

As mentioned in Sect. 3.3, the robot asked the participants about their feelings *during* the interview. Particularly, A24 and A25 were asked at the end of the first phase (see Table 1), and A44 and A45 were asked at the end of the second phase (see Table 2). In order to analyse their answers, their statements were transcribed into text format. To avoid possible misunderstandings, it was done manually and avoiding exclamation signs (as deducing them from audio might not be trivial). The commas, dots and interrogations signs were added according to the participant’s speech. Additionally,

trying to avoid modifying the original statement, grammatical mistakes were kept.

After the robot asked the participants how they were feeling about talking with it (A24) all of them stated being good and amazed. E.g., “*It’s fun, you are very cute, so it’s nice*”, “*Just great. You are almost a human*” and others even used the word “comfortable” without being asked yet about it: “*I actually feel comfortable in talking to you*”. There were also some comments about the robot’s mechanical attributes (e.g., about the motors’ noise or difficulty on establishing eye contact), but in general all of the interviewees pointed out being excited, as this was their first time interacting with a robot. One interviewee stated: “*At the beginning it was a bit strange for me; I couldn’t contact with you verbally, but now I can feel you actually. You can also feel me you know, I don’t feel you are a robot. You are more than a robot.*”⁸ Then, when they were asked whether they were comfortable (A25), most of them reported affirmatively. However, some added being or having been a bit nervous/anxious. E.g., “*Well, I have to say that at the beginning I was feeling a less comfortable, but now I am getting more relaxed talking to you.*”. Moving towards the end of the whole interview, when they were asked how they were feeling about the interview (A43), most of them surprisingly kept answering positively. For example, “*Ah, it’s fine, interesting I would say*” or “*I’m feeling good. Yeah*”. There were some participants that reported there was a change between the two interview phases; e.g., “*Yes, it’s quite a strange experience. At the beginning it was quite natural, and the more complex your questions began, the most strange the interview was*”, “*It’s very interesting. Sometimes uncomfortable, sometimes comfortable... It’s my first time interviewing with a robot. Sometimes I can’t understand you, sometimes I understand you... It’s a bit complicated, but it’s OK, I like it any way*”. At the end of the second phase, when the robot asked them if they wanted to continue talking with it (A44), most of the interviewees (80%, N = 23 out of 29 participants) answered positively (e.g., “*Yes, why not?*” or “*Yes, of course.*”). Nevertheless, there were 4 whose answer was rather neutral (e.g., “*Ah.. It depends*” or “*If you do some interesting questions yes, otherwise it’s fine*”) and 2 whose reply was negative (i.e., “*For now I guess that’s enough. Do you mind if I stop it?*” and “*Actually no*”).

To analyze further the interviewees’ statements, a **Sentiment analysis software** (i.e., *NLTK. Vader*; a Natural Language Processing Open Source software) was used to extract automatically their emotional feedback. It is important to mention, that this software gives importance to lexicons. That is to say, sentiment features represented not only by words, but through phrases (e.g., “*the bomb*”), emoticons (e.g., “*XD*”) or acronyms (e.g., “*LOL*”). Analyzing sentences transcribed from audio does not offer the possi-

bility of including such features. Hence, the accuracy of the sentiment recognition might be reduced. Moreover, when transcribing the speeches, it was noticed that the voice tone and facial movements played a very important role. Excluding these features, the verbal content might be misleading (check Sect. 5.5). The *NLTK* software returns a “compound” score between -1 (feedback extremely negative) and +1 (feedback extremely positive). Computing the score for each participant and question, A24 was compared to A43 and A25 to A44. To compare them, two Wilcoxon Tests were conducted: A24 versus A43 ($M = .603, SD = .373; M = .462, SD = .507; t = 176, p = .369$); and A25 versus A44 ($M = .458, SD = .274; M = .411, SD = .395; t = 169, p = .878$). Both analyses showed that the extracted feedback was not significantly different thus, hypothesis 2 was rejected. The interviewees reported to the robot a positive/neutral feedback at the end of both the first and second interview phases. However, as Fig. 3 showed, the second interview phase presented several *negative* self-reports for the majority of the actions performed by the robot. Hence, it seems that the direct reports to the robot regarding the second part (A43 and A44) are not in line with the self-reports done through the questionnaire (by answering Q14–Q18, Q20 and Q22). Exploring these reports from another perspective, it was noticed that the statements associated to each interview phase seemed to be apparently different in length. To explore this aspect, the statements’ length (i.e., *number of words* and *number of letters*) associated to the first interview phase was compared to the one associated to the second interview phase. A Wilcoxon Test confirmed the significant difference between the statements’ *number of words* reported at the end of the first part (A24+A25) ($M = 20.68, SD = 17.50$) and the ones reported at the end of the interview (A43+A44) ($M = 8.87, SD = 11.85; t = 30.5, p < .001$). Another Wilcoxon Test confirmed the significant difference between the statements’ *number of letters* reported at the end of the first part (A24+A25) ($M = 104.72, SD = 89.01$) and the ones reported at the end of the interview (A43+A44) ($M = 86.41, SD = 122.37; t = 108, p = .017$). It is possible that when people are not comfortable enough, and the topic of conversation doesn’t imply a fixed detailed answer, the answer becomes shorter.

Observing the videos associated to these reports, there were several participants whose statements seemed to be in contrast to their non-verbal cues. As an example, one interviewee reported: “*I feel nice, I feel happy to have done this experience and I feel OK*” while their detected Valence and facial movements seemed to report otherwise (see P12:Q24 in Fig. 7 and (j) in Fig. 15⁹). Given the verbal content might not to be reliable, further features (e.g., voice tone, facial/corporal movements, physiological signals

⁸ <https://youtu.be/ZL0wPDDiKF0>.

⁹ <https://youtu.be/mx4-Jxw6fxs>.

or speech's length) should be considered when assessing someone's *Comfortability*.

5.5 Participants' Valence and Arousal

To provide a quantitative measure of the participant's facial movements evoked during the interview, the *Valence* (i.e., how *negative/positive* an internal state is) and *Arousal* (i.e., how *intense* an internal state is) attributes were considered for the analysis. Each participant's interview video was divided into 24 clips where each clip comprehended a specific moment of the interview corresponding to the self-report questions (see Table 3). Subsequently, every clip was processed using the *FaceChannel* software [5] which received a clip and returned the *Valence* and *Arousal* levels detected per each frame of that video clip.

Figure 5 plots the *Valence* and *Arousal* averaged values detected per each participant and clip. The darker dots represent the actions that belonged to the first interview phase, whereas the lighter dots represent the actions belonging to the second phase. From the chart, it is appreciable that the *Valence* readings were much more diverse (i.e., from -0.65 to 0.65 approximately) than the *Arousal* readings (i.e., from -0.25 to 0.10 approximately). In other words, the estimated interviewees' *Valence* varied more than their *Arousal* along their interaction with the robot. To understand whether there was a significant difference between the readings done during the first and second interview phases, a Wilcoxon test was computed. It resulted that whilst a significant difference was found regarding the detected *Valence* during the *First interview phase* ($M = .259$, $SD = .238$) and the *Second interview phase* ($M = .150$, $SD = .247$; $t = 15472$, $p < .001$), there was no significant difference regarding the estimated *Arousal* during the *First interview phase* ($M = -.010$, $SD = .055$) and the *Second interview phase* ($M = -.016$, $SD = .062$; $t = 28670$, $p = .367$). That is to say, the interviewees maintained similar *Arousal* levels during both interview phases, but their *Valence* decreased significantly.

To explore more in depth the changes between the two phases, each participant was analyzed individually. In particular, the *Valence* and *Arousal* values detected during the second phase were subtracted from the ones detected during the first phase. Given both variables are part of a negative-to-positive scale (from -1 to 1), the absolute value of the difference was calculated. Figure 6 plots both results, where the striped bars represent the *Valence* differences and the filled bars the *Arousal* differences. A Wilcoxon Test confirmed that, as expected, all the interviewees changed significantly more their *Valence* ($M = .111$, $SD = .089$) than their *Arousal* ($M = .019$, $SD = .022$; $t = 6$, $p < .001$) from one phase to the other. Still, not all the interviewees

presented the same amount of changes. There were some participants who presented very extreme and opposite *Valence* levels in both phases (e.g., *P4*, *P12*, *P15*, *P18*, *P21* and *P28*), whereas others barely presented any extreme change (e.g., *P3*, *P10*, *P17*, *P19*, *P22*, *P25* and *P29*). As mentioned in the previous sections, this behaviour might be due to two different unknowns. First, people might have assimilated the respective situations in distinct and contrary manners. And second, people might have expressed similar emotions differently, thus the software might have not been able to detect it. To explore further these individual differences, the specific *Valence* levels a participant experienced during each interview question are displayed in Fig. 7.

Looking at Fig. 7, it is appreciable that the horizontal lines are somewhat better defined than the vertical ones; which highlights that each person tends to express themselves in a particular manner with respect to others. Nevertheless, most participants presented a *Valence* decrease as the interview progressed. In particular, there were some actions that made most of the interviewees decrease it considerably (e.g., *Q8*, *Q9*, *Q16*, *Q17* and *Q21*). To explore this tendency, the average *Valence* among all participants was computed for each specific action and plotted in Fig. 8. This bar chart shows that most of the interviewees maintained a general positive or neutral *Valence* during the whole interview. Nonetheless, a Wilcoxon Test confirmed that, as expected from the previous calculations, the *Valence* detected during the *First interview phase* (*Q1–Q12*) ($M = .259$, $SD = .094$) was significantly higher than the *Valence* associated to the *Second interview phase* (*Q13–Q24*) ($M = .150$, $SD = .073$; $t = 9$, $p = .016$), which confirmed the hypothesis 3. Considering solely the actions meant to trigger a particular *Comfortability* intensity, it was found that the actions meant to trigger *high Comfortability* ($M = .267$, $SD = .240$) presented a *Valence* significantly higher than the ones meant to trigger *low Comfortability* ($M = .143$, $SD = .247$; $t = 10397$, $p < .001$). Also, contrary to what occurred for the self-report, the *neutral actions* seemed to have elicited expressions in line with the ones evoked by the nearby triggers. That is to say, the *Valence* elicited by the neutral actions that occurred during the first interview phase ($M = .274$, $SD = .222$) was significantly higher than the one elicited during the second phase ($M = .112$, $SD = .226$; $t = 3072$, $p < .001$). Another very interesting aspect is that the *Direct reports to the robot*, which were previously self-reported with similar *high Comfortability* levels (independent of the interview phase, see Sect. 5.4), now present a significant difference from each other. The ones associated to the end of the first phase ($M = .291$, $SD = .242$) evoked a significantly higher *Valence* than the ones evoked by the questions associated to the end of the interview ($M = .233$, $SD = .260$; $t = 595$, $p = .043$).

Fig. 5 Each dot represents the **Arousal/Valence** average detected for a specific participant during a specific moment of the interview (recalled in Table 3). The darker dots represent the questions that occurred during the *First interview phase* and the lighter dots represent the questions that occurred during the *Second interview phase*

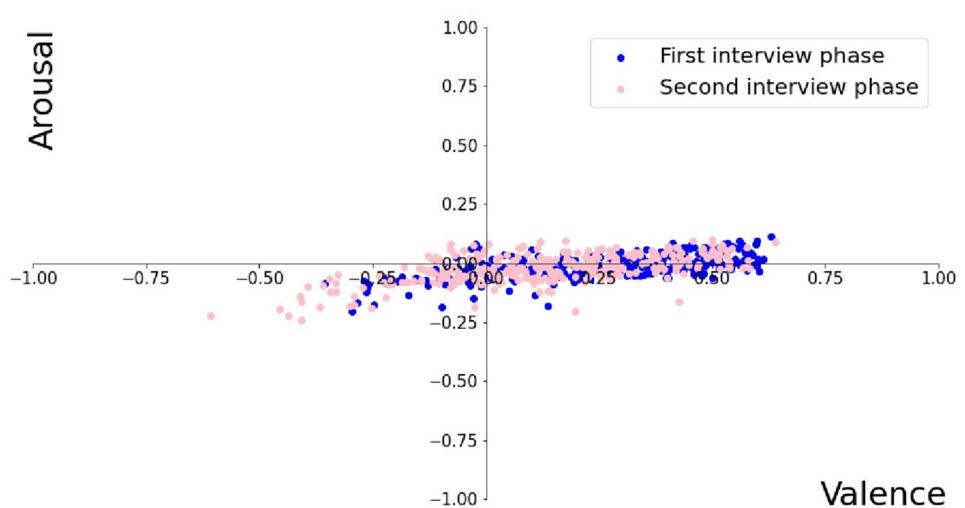


Fig. 6 Individual differences between the detected *Valence/Arousal* during the first and second interview phases respectively

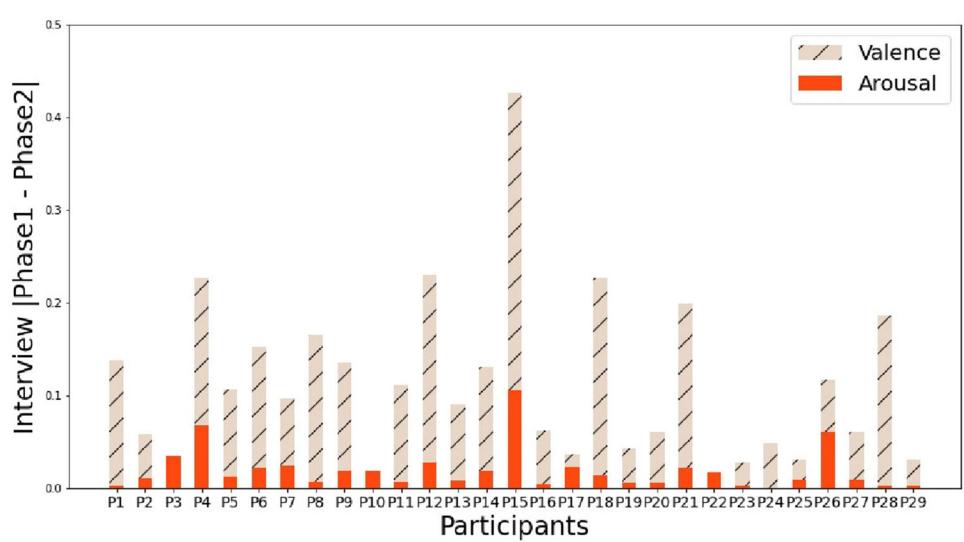


Fig. 7 Valence detected, per participant, during specific moments of the interview (made reference in Table 3). Each square represents the Valence average of all the video frames associated to an specific robotic action

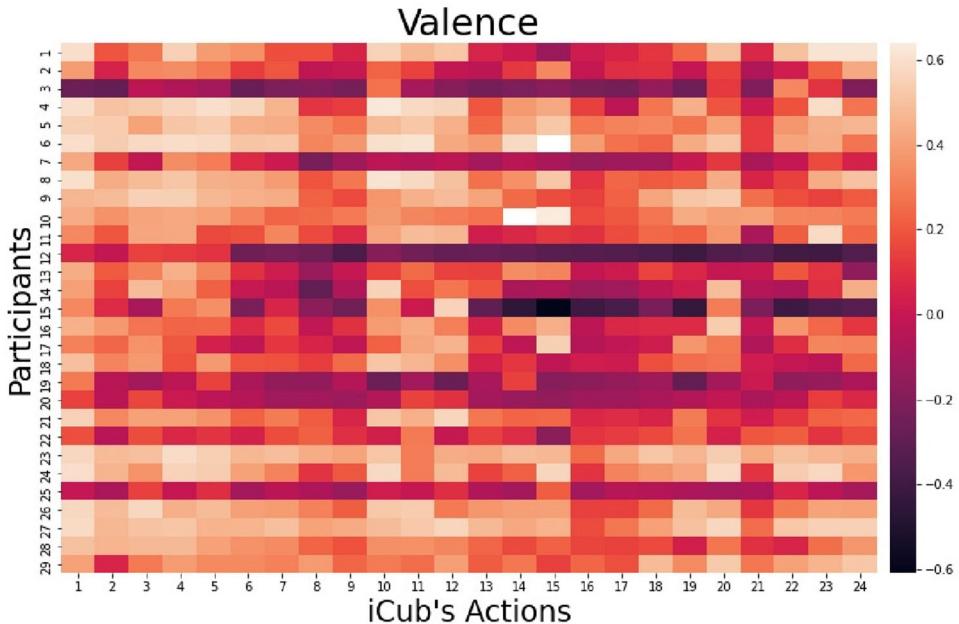


Fig. 8 Valence average and standard deviation detected during specific moments of the interview (referenced in Table 3)

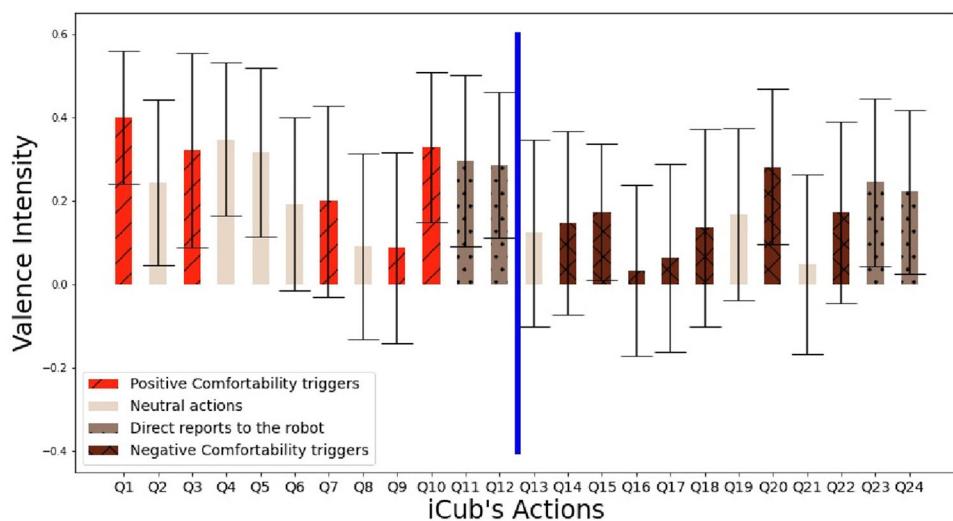


Fig. 9 Arousal detected using the software FaceChannel [5] in function of the self-reported *Comfortability*

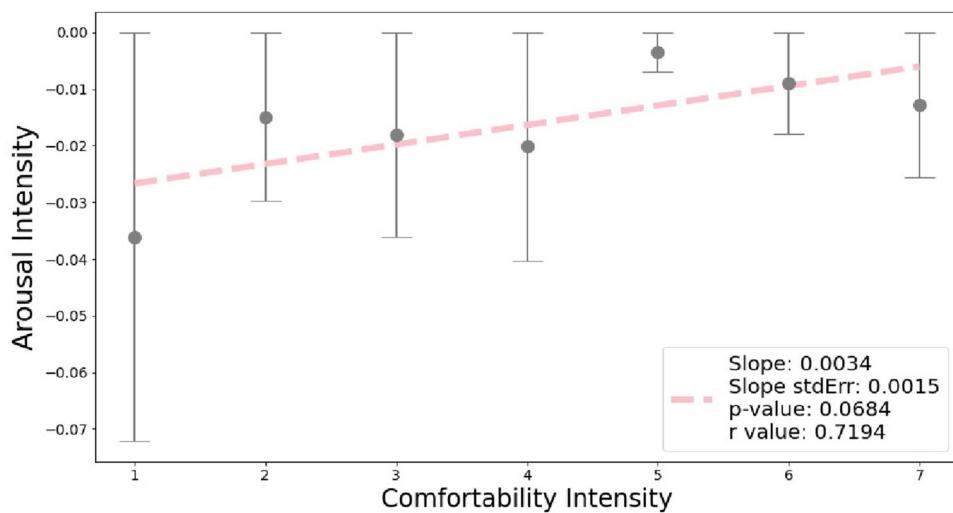
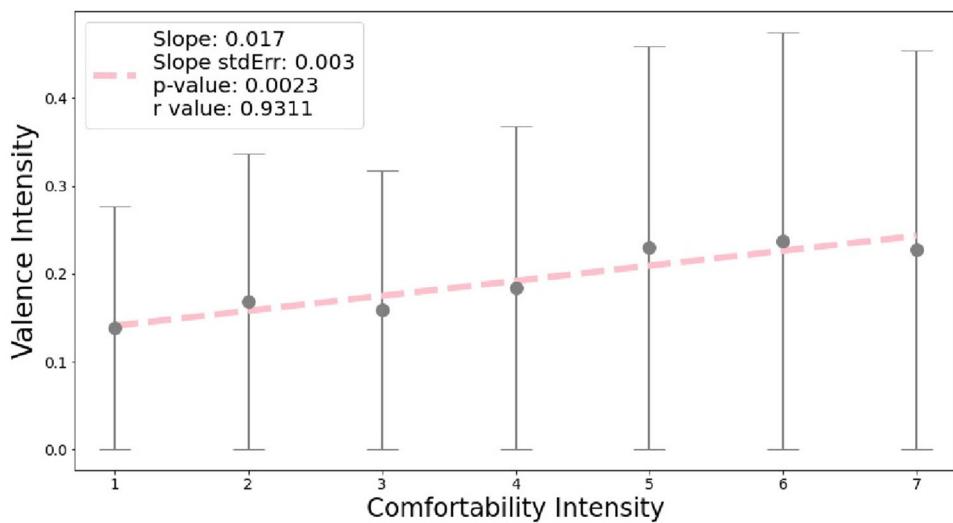


Fig. 10 Valence detected using the software FaceChannel [5] in function of the self-reported *Comfortability*



This might suggest that self-reports are not entirely reliable; specially if they are reported in person.

In order to correlate the self-reports with the expressions detected by the software, the detected *Arousal* and *Valence* intensities were plotted as a function of the associated self-reported *Comfortability*. The resulting linear regressions (see Figs. 9 and 10) show that there seems to be a tendency. The higher the *Comfortability*, higher the corresponding *Arousal* / *Valence* levels. Nevertheless, only *Valence* was found to be significantly correlated (*slope*: .0017, *r*: .9311 and *p* = .0023) with *Comfortability*.

To understand deeper whether *Valence* could be sufficient to predict *Comfortability* levels, more analyses were conducted. Precisely, we wanted to understand whether someone might have reported opposite *Comfortability*/*Valence* values for the same situations. Thus, 9 different conditions were considered and the frequency of their occurrence was compared between each other (see Fig. 11): (++) high *Comfortability* and *Valence*; (+N) high *Comfortability* and neutral *Valence*; (−) high *Comfortability* and negative *Valence*; (N+) neutral *Comfortability* and positive *Valence*; (NN) neutral *Comfortability* and neutral *Valence*; (N−) neutral *Comfortability* and negative *Valence*; (−+) low *Comfortability* and positive *Valence*; (−N) low *Comfortability* and neutral *Valence* and (−) low *Comfortability* and negative *Valence*. A **positive** (+) value was defined by self-reporting a *Comfortability* level higher than 4 or being detected with a *Valence* intensity higher than .01. A **negative** (−) value was defined by self-reporting a *Comfortability* level lower than 4 or being detected with a *Valence* intensity lower than -.01. And a **neutral** (N) value was defined by self-reporting a *Comfortability* level equal to 4 or being detected with a *Valence* intensity higher than -.01 and lower than .01.

Figure 11 illustrates that the majority of the situations presented similar *Comfortability* ratings and *Valence* readings (i.e., (++): 46.26%; (−−): 5.89%). Nonetheless, there were others that did not. While there were some actions that triggered positive/negative VS neutral levels (i.e., (+N): 1.29%; (N+): 11.06%; (N−): 3.73% and (−N): .71%), there were others that triggered exactly the opposite extremes (i.e., (−+): 12.21% and (−−): 18.39%). To tackle this aspect more thoroughly, the *Valence* threshold was increased to .4 and -.4 respectively. Figure 12 shows that consequently, more situations fell under the neutral *Valence* category. And most importantly, there were still 37 situations that were reported with *low Comfortability* while presenting an extreme *positive Valence* (i.e., −+; 5.45%) and 7 situations which presented a *high Comfortability* level while presenting an extreme *negative Valence* intensity (i.e., +−; 1%). To exemplify some of these situations, Figs. 13, 14, 15 and 16 include 16 videos extracted from several interviews. Each one of these figures includes situations in which the respective self-reported *Comfortability* and detected *Valence* had a particular nega-

tive/neutral/positive intensity value. Before going into more details, it is relevant to recall that being extremely comfortable (i.e., *Comfortability* = 7) does not imply feeling simultaneously a positive internal state. One could be very comfortable while experiencing other negative emotions such as *anger*, *disappointment* and so on (see the examples included in Sect. 1.2). Nevertheless, a positive (i.e., high) *Valence* should imply experiencing a positive internal state. Similarly, being extremely uncomfortable (i.e., *Comfortability* = 1) does not always imply feeling simultaneously a negative internal state. One could be very uncomfortable while experiencing other positive emotions such as *happiness*, *excitement* and so on (please refer to Sect. 1.2 for a more detailed example). But again, a negative (i.e., low) *Valence* should imply experiencing a negative internal state.

Bearing the aforementioned thought in mind, Fig. 13 includes four videos associated to **high** *Comfortability* self-reports and *Valence* readings (see Fig. 11; (a): P4Q1, (b): P21Q1, (c): P17Q15 and (d): P27Q22)). Even though both assessments evaluated these situations with a high value, the recorded videos suggest that some participants might have experienced negative internal states. For instance, *a* and *b* might have been a little bit nervous and *c* and *d* might have been shocked by the unexpected events. Addressing someone's positive internal state by judging solely their facial movements is a complicated task. For example, in contrast to the traditional assumption of smiles being an indicator of *positiveness*, researchers have shown that there are plenty of smiles which don't imply positiveness [15, 21]. Thus, it is likely that these videos should not have been categorized with such extreme *Valence* intensities (i.e., higher than .4). Regarding *Comfortability*, it is extremely difficult to be certain of the participants' real *Comfortability* state at those precise moments. On the one hand, the self-reports were performed after the interview, obligating the participants to recall their past feelings rather than communicating them at the very moment they occurred; which might have induced to wrong self-assessments. On the other hand, it is possible that their real internal state might have been hidden in purpose, and thus, incorrect reports might have been submitted. In fact, masking emotions is surprisingly common and attributed to a human survival instinct [11]. Next, Fig. 14 includes four videos associated to **low** *Comfortability* self-reports and *Valence* readings (see Fig. 11; (e): P7Q10, (f): P19Q20, (g): P14Q22 and (h): P20Q3). This time, both assessments seem to be correct. Paying attention to their non-verbal behaviours, there seems to be no sign of positiveness, neither of desiring to keep with the ongoing interaction. Additionally, Fig. 15 includes four videos associated to **high** *Comfortability* self-reports and **low** *Valence* readings (see Fig. 11; (i): P3Q24, (j): P12Q23, (k): P11Q21 and (l): Q15Q19)). During these situations, the interviewees' self-reports were not in line with the *Valence* detected by the software. While they reported

feeling comfortable, they were detected with an extremely low *Valence*. Observing the videos, it seems that indeed they might have been experiencing a negative internal state (e.g., anger). Still, it is uncertain whether they were comfortable or not. They self-reported being extremely comfortable, which seems to not correspond with certain facial and/or corporal movements spotted during the videos. Lastly, Fig. 16 includes four videos associated with **low Comfortability** self-reports and **high Valence** readings (see Fig. 11; **(m)**: *P1Q24*,

Fig. 11 [Comfortability, Valence] values per participant during each interview question (detailed in Table 3) where +: *Comfortability* > 4 and *Valence* > .01, -: *Comfortability* < 4 and *Valence* < -.01 and N: *Comfortability* = 4 and -.01 < *Valence* < .01 respectively

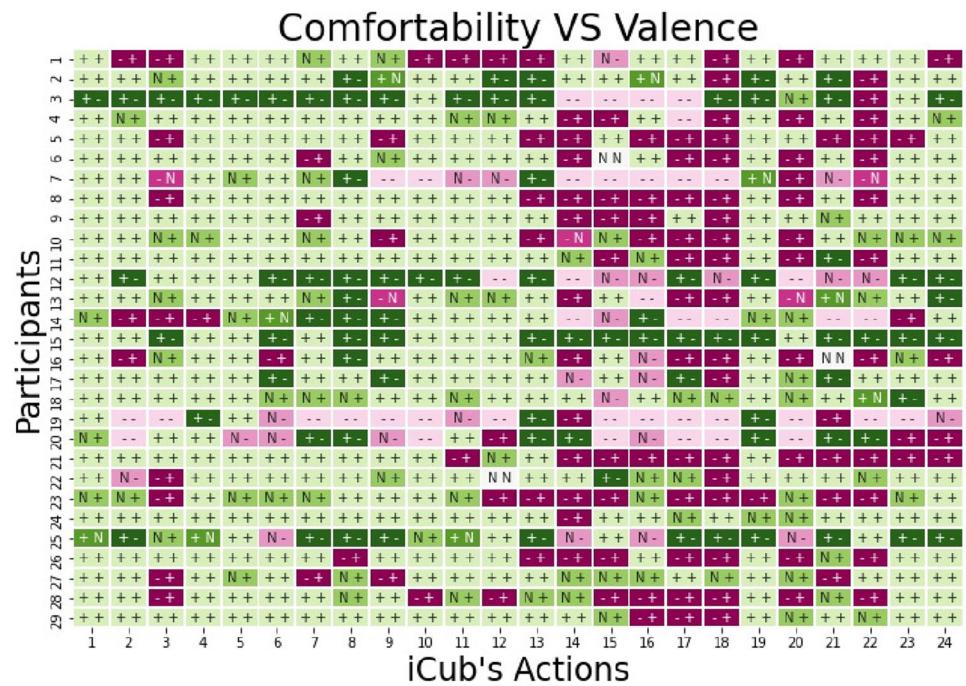
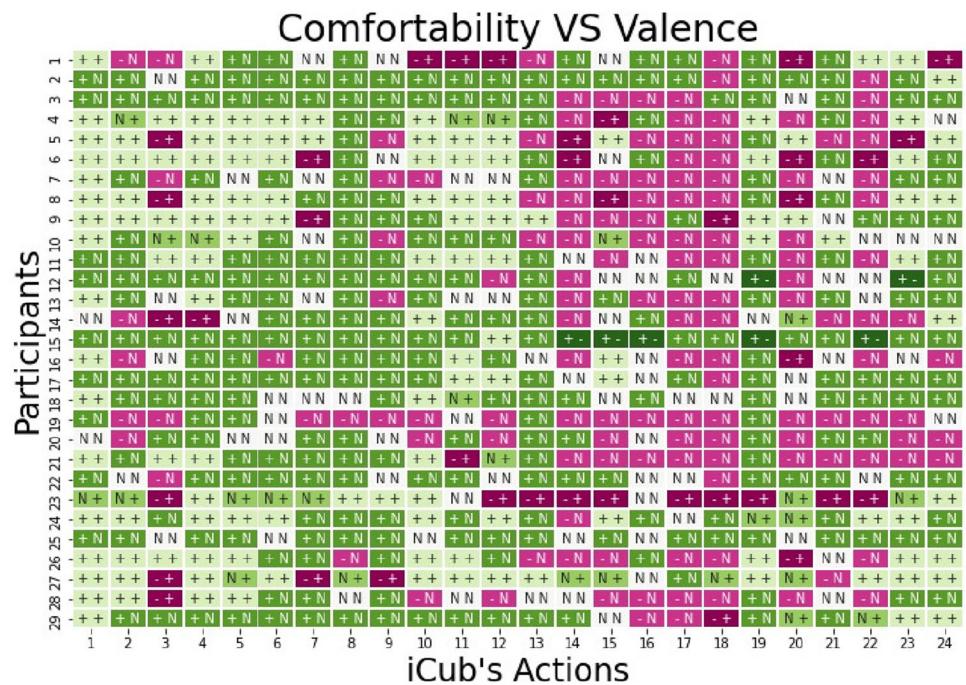


Fig. 12 [Comfortability, Valence] values per participant during each interview question (detailed in Table 3) where +: *Comfortability* > 4 and *Valence* > .4, -: *Comfortability* < 4 and *Valence* < -.4 and N: *Comfortability* = 4 and -.4 < *Valence* < .4 respectively



(n): *P23Q18*, **(o)**: *P5Q23* and **(p)**: *P8Q3*). Once again, the assessments seem to contradict the facial expressions. Similarly to Fig. 13, it seems that the *Valence* values estimated by the software are not reflecting the actual valence, which can be supported by the self-reports. This time, the interviewees were the ones reporting not being OK with the current situation, which seems to be in line with the visual information.

In conclusion, these results show that what can be classified as a *positive / negative* internal state, does not necessarily



Fig. 13 (++) Participants' interview videos in which both, the self-reported *Comfortability* and detected *Valence* were **positive**. **a** <https://youtu.be/Mzl0n-rFd4M>; **b** https://youtu.be/jGO1iq_3jf4; **c** <https://youtu.be/wOL8sp346kI>; **d** <https://youtu.be/BFlcZXErsv>



Fig. 14 (--) Participants' interview videos in which both, the self-reported *Comfortability* and detected *Valence* were **negative**. **e** <https://youtu.be/Ku8zenYMLV8>; **f** <https://youtu.be/MaBp8Cm1Sc0>; **g** <https://youtu.be/CFCdYcxGGXw>; **h** <https://youtu.be/XDoUK0v3FFw>



Fig. 15 (+-) Participants' interview videos in which the self-reported *Comfortability* was **positive** while the detected *Valence* was **negative**. **i** <https://youtu.be/3fYBoWI9nBw>; **j** <https://youtu.be/mx4-Jxw6fxs>; **k** <https://youtu.be/tr9LLwjyslQ>; **l** <https://youtu.be/PgT6jbZpiXM>



Fig. 16 (--) Participants' interview videos in which the self-reported *Comfortability* was **negative** while the detected *Valence* was **positive**. **m** <https://youtu.be/xW2oXvAPeFw>; **n** <https://youtu.be/QB4PFHzKk94>; **o** <https://youtu.be/NTYhDqH9m80>; **p** <https://youtu.be/EbcYqYR3dnU>

imply being labeled as *comfortable / uncomfortable* and vice-versa. It also highlights the complexity of assessing someone's internal state, both from the analysis of specific body movements (position, kinematics or other features such as its color) and from self-reports. Overall, these observations suggest that to accurately recognize someone's *Comfortability*, detecting their Valence is not sufficient and a deeper analysis should be conducted to explore all kinds of cues that might arise with certain *Comfortability* intensities.

5.6 Low Comfortability Visual Cues

As an initial approach to understand how people might behave while being uncomfortable, the following qualitative analysis was conducted. First, each interview video was trimmed into 24 segments (each one associated to an interview question) and then, each one of those segments was trimmed again into three-second clips (the time a macro-expression tend to last [13]). Next, all these clips were watched (randomly and without audio) by two experts on recognizing affective states visually, with the task of writing a list of the cues that they spotted as extremely low *Comfortability* indicators. As a result, all the cues that intersected among the two observers were clustered into the following groups depending on their modality: **Low Comfortability Facial Movements (LC-FM)**, **Low Comfortability Corporal Movements (LC-CM)**, **Low Comfortability Actions (LC-A)** and **Low Comfortability Other Facial Features (LC-O)**. Likewise, Table 4 presents the 10 *facial movements* (associated to the eyes, eyebrows or mouth) spotted as low *Comfortability* cues. Table 5 presents the 5 *corporal movements* (associated to the shoulders, head or neck) considered as low *Comfortability* cues. Table 7 presents the 12 *actions* (e.g., swallowing, intense breathing, etc.) individuated as low *Comfortability* indicators. And, Table 8 reports the *additional facial feature* that is not associated to movements (i.e., the face turns reddish) spotted as a low *Comfortability* cue.

The videos included in the aforementioned tables as examples, are the ones that isolate better the listed features. However, most of the time these features co-occurred in parallel, they rarely appear alone. For example, a subject presented a tensed smile (LC-FM8) while bouncing their trunk (LC-CM5) and their facial pigments were reddish (LC-O1) (see video: youtu.be/DzX9oxA9nXY). Another subject had both eyes raised (LC-FM5) while bouncing their trunk (LC-CM5) and breathing heavily (LC-A6) (youtu.be/DzX9oxA9nXY).

A third had their mouth opened and tensed (LC-FM7), while raising both eyebrows (LC-FM4), bouncing their trunk (LC-CM5) and swallowing saliva (LC-A2) (i.e., youtu.be/DzX9oxA9nXY). Another participant presented a tensed smile (LC-FM8) and raised their shoulders up (LC-CM1) while inclining their head forward (LC-CM2), bouncing their trunk (LC-CM5) and tucking their hair (LC-A12) (see video on: youtu.be/DzX9oxA9nXY).

Another subject showed facial (LC-FM1) and neck ticks (LC-CM3), a tilted neck (LC-CM4) and a rapid blinking (LC-FM2), while raising both eyebrows (LC-FM4), moving their shoulders up and down (LC-CM1), inclining their head forwards (LC-CM2), and bouncing their trunk (LC-CM5) (youtu.be/DzX9oxA9nXY). One of the participants presented a tensed open mouth (LC-FM7), while breathing heavily (LC-A6) and looking down (LC-A9) (youtu.be/DzX9oxA9nXY), whereas another exhibited a

Table 4 Low Comfortability—Facial Movements

LC-FM	Description	Facial Movement
1	Facial twitch	 youtu.be/XJ6QzrfWW78
2	Rapid blinking	 youtu.be/pywdHKYSe5g
3	Raising a single eyebrow	 youtu.be/QM3Lk0RVxzQ
4	Raising both eyebrows	 youtu.be/1QY35aVmzJ8
5	Maintain both eyebrows raised for at least three seconds	 youtu.be/WF704p-qoRY
6	Single upper lip raiser	 youtu.be/5k6XqxGsDFo
7	Tense mouth open	 youtu.be/0ySR5CKHT8U
8	Tense smile	 youtu.be/e0I8yWBypkw
9	Lips pressed together	 youtu.be/Cck4VFhV-kA
10	Upside-down smile	 youtu.be/oet5OErPczs

Table 5 Low Comfortability—corporal movements

LC-CM	Description	Corporal Movement
1	Shrugging	 youtu.be/5r7AFAXlId8
2	Head inclined forward	 youtu.be/uxX_qsPKPtK
3	Neck twitch	 youtu.be/Yuysx9OOaDw
4	Tilted Neck	 youtu.be/Ii08Njpeu1w
5	Bouncing trunk	 youtu.be/x5L3icAZigY

tensed smile (LC-FM8) and a tilted neck (LC-CM4) while bouncing their trunk (LC-CM5), and looking down (LC-A9) (<youtu.be/DzX9oxA9nXY>).

Most of the reported cues were perceived in several occasions (either in the same participant as well as in others) by the two observers, which provides a first positive insight for recognising *Comfortability* visually.

5.7 Personality Traits and Attitude Toward Robots

After conducting several interviews, it became apparent that not all the interviewees reacted in the same manner to the same circumstances. Their *personality* and *attitude towards robots* seemed to be the two main determining factors. For

Table 6 Low Comfortability—actions

LC-A	Description	Action
1	Crossing arms	 youtu.be/U6Jqrfla3Y
2	Gulping	 youtu.be/mnQDa461gLk
3	Licking Lips	 youtu.be/PMz832yqrI
4	Biting lower lip	 youtu.be/W09YzRV_FQs
5	Sighing	 youtu.be/yr_VRHS0y48
6	Breathing heavily	 youtu.be/8_5xNOzJaIo
7	Looking away (moving the head)	 youtu.be/T7pjqVTn4kY
8	Looking away (using only the eyes)	 youtu.be/0vOTZhP7Ns9

Table 7 Low Comfortability—actions

LC-A	Description	Action
9	Looking down	 youtu.be/oLUS-5zG0kI
10	Looking from down	 youtu.be/MJiwQ0R1pQs
11	Touching face	 youtu.be/oznF1wq6a1w
12	Tucking hair	 youtu.be/iMr_q-hCJaU

example, it was observed that the participants who tended to be more *extroverted* (i.e., didn't show any shyness about talking with the robot or us), or even *open minded*, experienced more fun during the interview. Additionally, it was expected that people who would associate robots with discomfort attributes would have been more negatively impacted by their interaction with iCub. Therefore, an analysis meant to study the relationship between these specific personality traits (*Extroversion* and *Open to new experiences*) and attitude toward robots (i.e., perceiving robots as discomforting objects) with the self-reported *Comfortability* and the detected *Valence* and *Arousal* values was conducted.

As mentioned in Sect. 3.7, to identify the interviewee's personality the **TIPI** questionnaire [19] was used. This questionnaire explores 5 personality factors (*Extraversion*, *Agreeableness*, *Conscientiousness*, *Emotional Stability* and *Openness to new Experiences*). It contains 2 Items per factor scored following a 7-point Likert scale. Given our research interest, the *Extraversion* and *Open to new Experiences* factors were analyzed. To do so, several linear regressions

Table 8 Low *Comfortability*—other facial features

LC-O	Description	Facial Features
1	Facial pigments turn reddish	 youtu.be/pUe2AOzNR68

between each one of these factors and the self-reported *Comfortability* as well as the estimated *Valence / Arousal* values were computed. Figure 17 shows that there was not a significant correlation between the participants' *extraversion* and the *Comfortability* they self-reported during the interview ($slope : .029$; $r - value : .0366$ and $p - value : .8506$). On the other hand, Fig. 18 shows a tendency according to which people with a higher degree of *openness to new experiences* self-reported a higher level of *Comfortability* ($slope : .291$; $r - value : .324$ and $p - value : .0859$). No other linear regression resulted significant. These results provide evidence against *Hypothesis 4*. Nonetheless, the initial assumptions were not exactly erroneous. They indicate that peoples' personality might be determinant in some circumstances. It is possible that the attributes of personality chosen for this study were not the ones defining how they would feel or behave when facing a humanoid robot. Hence, other aspects of personality might bring brighter correlations.

In order to identify the interviewee's attitude toward robots, the **RoSAS** questionnaire [9] was used. This questionnaire explores 3 robot-perception related factors (*Warmth*, *Competence* and *Discomfort*). To do so, it contains 18 Items (6 items per factor) scored following a 9-point Likert scale. As explained before, the participants' estimated *Valence/Arousal* and self-reported *Comfortability* were evaluated as a function of their *Discomfort Factor*. It is important to mention that because of an internal mistake, the *Awkward and Dangerous Items* were excluded from the analysis. Figure 19 shows that there is a tendency: the more the interviewees associated discomfort attributes to robots in general, the more their *Valence* decreased during their interaction with iCub. This however does not reach significance ($slope : -.0518$; $r - value : -.3469$ and $p - value : .0652$). No other linear regression analyses resulted significant or showed a tendency. These findings are almost in line with *Hypothesis 5*. Given not all the Items could be considered for the analysis, no precise conclusion would be drawn regarding this hypothesis.

5.8 Robot's Perception

As explained in Sects. 3.7 and 3.8, the participants were asked several questions about the way they perceived iCub. This was done some days after the interview finished via an online questionnaire. The answers showed that the robot was perceived as a **Male** (51.7% *Male*, 34% *Agender*, 7.59% *Non-binary*, and 6.67% *Female*), **Grade-schooler** (7.8% *Toddler* [1–2 years], 9.6% *Preschool* [3–5 years], 41.5% *Gradeschooler* [6–12 years], 8.4% *Teen* [13–18 years], 5.8% *Young adult* [19–21 years], and 27.6% *Adult* [22 or more years]), **who most of the time did not act of its own free will** (51.3% *Most of the time no*, 25.1% *Half of the time yes and half of the time no*, and 23.7% *Most of the time yes*), by the majority of the interviewees. Despite these factors, most of the interviewees specified that **they did not have the impression to be interacting with a person for most of the interview** (50.6% *Most of the time no*, 34.5% *Half of the time yes and half of the time no*, and 14.9% *Most of the time yes*).

Additionally, to discover whether the nature of the interviewer (i.e., being a robot) was a determinant factor, the participants were requested to remember the actions performed by iCub and to imagine being interviewed by a *human*, the *Geminoid DK* and *Care-O-Bot* robots instead. Initially, the interviewees were going to be asked only to imagine being interviewed by a *human*. However, some of the declarations obtained verbally after the interview, showed that their *Comfortability* might have been influenced by other factors as well. E.g., the age they attributed to iCub (i.e., being a child vs an adult) or its appearance (i.e., being human- vs. machine- like) rather than its nature (i.e., being a robot). For this reason, the *Geminoid DK* (modeled after the professor Henrik Scharfe) and the *Care-o-Bot* robots were chosen given their respective human adult vs machine-like appearance. An image of the corresponding robot was included next to the associated questionnaire's question. Figure 20 contains those images. Given *Comfortability* is placed from a low (*uncomfortable*) to high (*comfortable*) dimension, the interviewees were asked to remember separately the actions which made them feel *comfortable* and *uncomfortable* respectively. From Fig. 20 it can be seen that considering only the *comfortable* situations, the majority of the interviewees thought they would have felt *the same Comfortability* levels interacting with all the proposed agents (see graphs (a), (c) and (e)). Nevertheless, it is interesting to highlight that, while a *human* is thought to make a large portion of the participants (about 39%) feel *more comfortable* the other robots do not evoke the same expectation, especially the *Care-O-Bot*. Conversely, a good percentage of interviewees (about 38%) associate the *Geminoid DK* with a potential decrease in *less comfortable* feelings. Moving to the situations in which iCub made the participants feel *uncomfortable*, graphs (b), (d) and (f) show that while the *Care-O-Bot* would have elicited the

Fig. 17 Participants' self-reported *Comfortability* during the interview in correlation with the *Extraversion* factor of the TIPI questionnaire

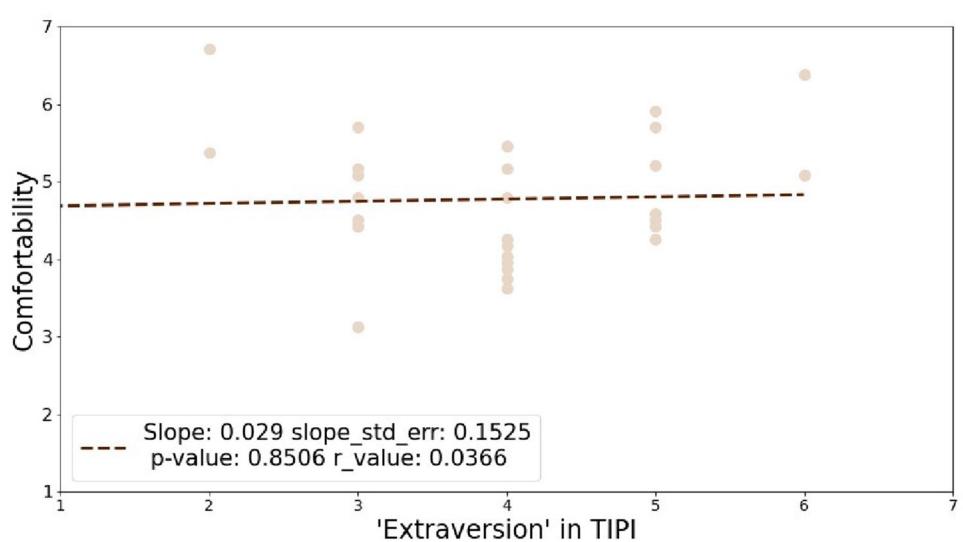


Fig. 18 Participants' self-reported *Comfortability* during the interview in correlation with the *Open to new experiences* factor of the TIPI questionnaire

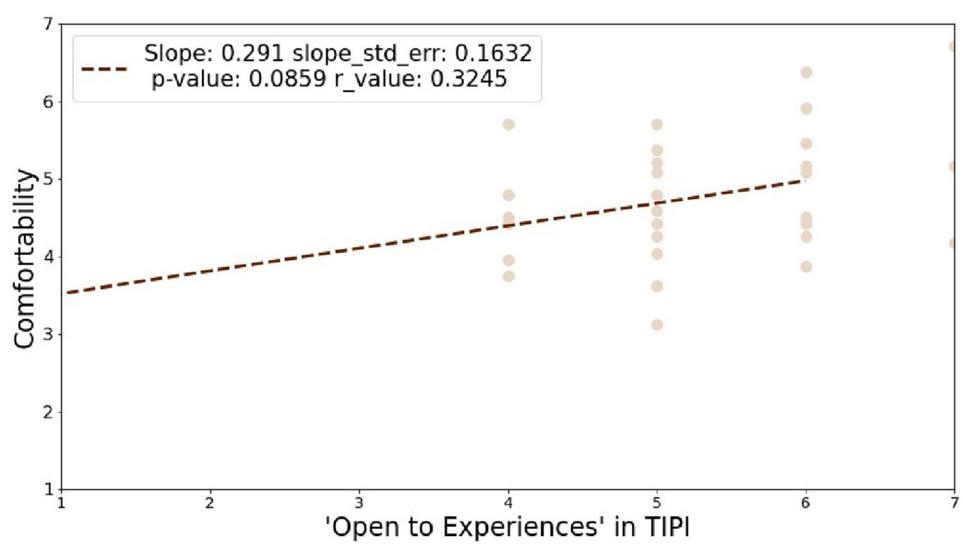
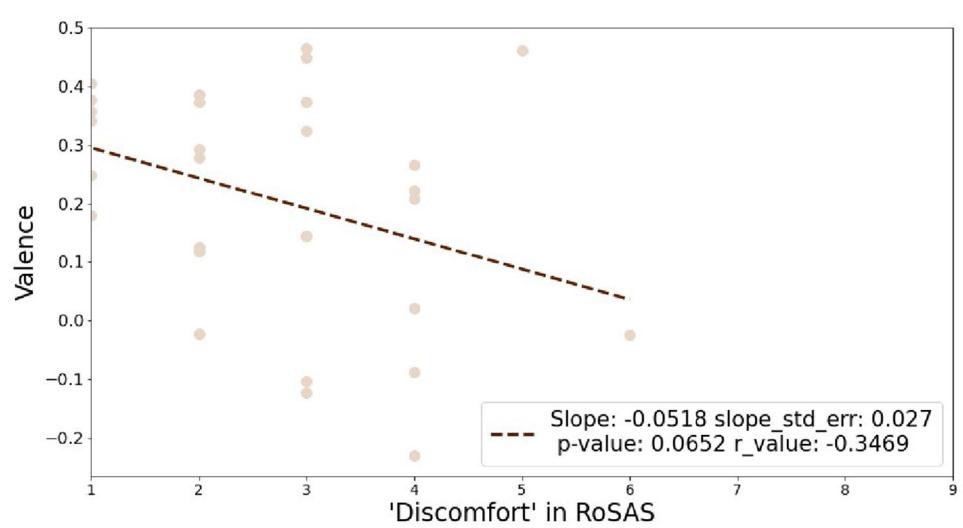


Fig. 19 Participants' Valence during the interview in function of the *Discomfort* factor of the RoSAS questionnaire



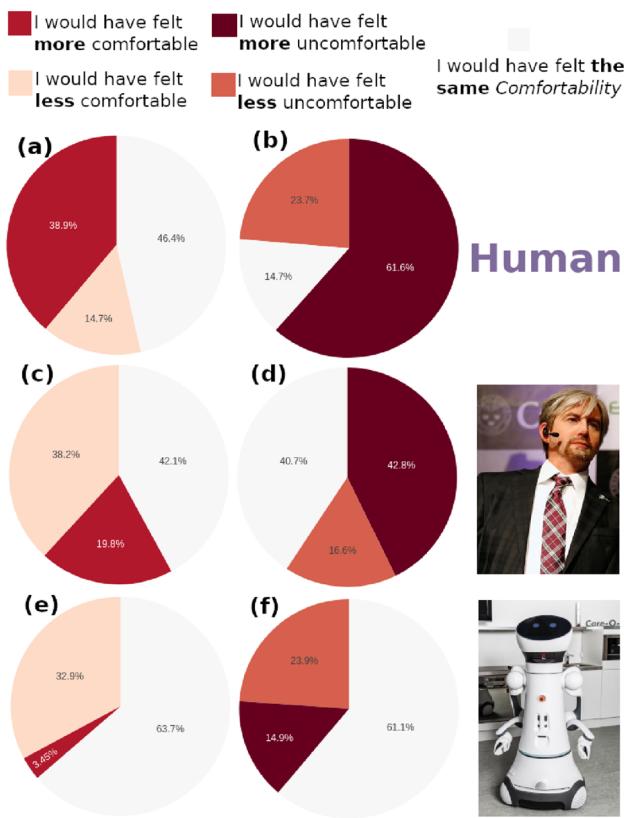


Fig. 20 Participants' answers when asked: If the actions performed by iCub (which made you feel *comfortable/uncomfortable*) would have been performed by another agent (*human/Geminoid DK/Care-O-Bot*), how do you think your *Comfortability* level would have been impacted?

same *Comfortability*, the *Geminoid* and *human* would have made them feel *more uncomfortable*. This was particularly evident for the *human* agent (about 62% of the participants reported that it would have increased their “*uncomfortability*”). It seems that, as the agent resembles more a human the resulting actions would have made them feel **more extreme *Comfortability*** levels. These findings show that, the nature of the agent is relevant when evaluating *Comfortability* during an interaction.

6 Discussion and Future Work

This study has explored the way people feel and behave during real and uneasy circumstances while interacting with a humanoid robot. To create such sensitive situation, selected researchers from our institution, not working on HRI, were interviewed by iCub for a novel column of the institutional magazine. The interview was designed to evoke different levels of *Comfortability* in different phases of the interaction. Several videos and questionnaires were collected before, during and after each interview.

The participants' self-report, collected (through a questionnaire) at the end of the interview, demonstrated that a humanoid robot can make people feel opposite and extreme *Comfortability* levels. A *Comfortability* decrease was clearly visible between the first interaction phase (i.e., the “*comfortable*” one) and the second phase (i.e. the “*uncomfortable*” one). Interestingly, the participants reported otherwise when iCub itself asked them about their *Comfortability* during the interview. This time, almost all the responses were on average positive in both phases. Hoffmann et al [22] explored the impact of reporting something in private or directly to an agent and they found significant differences. They discovered that people are more polite when they have to report their feedback directly to the agent than filling a questionnaire. Additionally, the analysis of the participants' Valence and Arousal intensities, estimated using an external software, showed a significant higher Valence during the first phase than during the second one. Nevertheless, a detailed comparison of the participants' facial movements and self-reports revealed that, Valence is not sufficient to predict someone's *Comfortability*. A consistent portion of videos (about 30%) showed a clear inconsistency between the estimated Valence and the corresponding self-reported *Comfortability*. In particular, about 12% presented high *Comfortability* self-reports and negative Valence estimations (see Fig. 15) while about 18% where characterized by the opposite conflict (see Fig. 16). These findings underline the importance of identifying further cues to assess *Comfortability* and highlight the potential limitations of both, self-reports and automatic emotion analysis; agreeing with prior literature [4, 24]. It is known that understanding people's feelings, emotions, thoughts and intentions is an extremely difficult task. On the one hand, self-reports are useful because they allow people to report about their feelings directly, without middlemen and hence without the potential bias of wanting to avoid offending the partner. Yet, the self-reports in our experiment were filled at the end of the whole interaction, which could have affected the accuracy of the information reported. Declaring a past event requires remembering it, and thus the information might be distorted. During the interview participants experienced the questions one at a time, not knowing what would happen next. Instead, in the recollection they already new about the whole interaction. This aspect could have led to erroneous self-assessments reporting the *Comfortability* levels experienced online rather than the ones experienced during the interview. Additionally, self reports require the subject to be aware of their own internal states. This aspect might seem trivial, however it is possible that a person experiences a feeling, but is not able to identify it. Literature has reported several reasons for this to happen. Firstly, it can be due to a mechanism of self-defense built by our own body to mask internally the emotions we are not ready to handle [40]. Secondly, it can be due to a lack of knowledge of one-self or about the emotion itself.

That is to say, it is possible to be unaware of an emotional event or unable to express the emotional state with words (*alexithymia* [4]). And lastly, it might be due to an intentional desire of hiding a feeling, not wanting to reveal it to others [11, 40]. On the other hand, software specialized on recognizing people's internal states might present some limitations. For example, it was noticed that the software we used to estimate someone's Valence, associated every smile with a positive Valence. Nonetheless, not all the smiles are positive indicators [21]. In fact, our findings show that the people who self-reported being *extremely uncomfortable* were often *smiling* during the interview. Therefore, it can be appreciated that identifying someone's internal state by using these approaches (i.e., either *self-reports* or *Valence recognition software*) has plenty of drawbacks. This takes us to the next relevant contribution of this research: to provide a first rough indication of how people behave while being uncomfortable. To this aim, two experts on visual emotion recognition spotted all the visual features they recognized as low *Comfortability* indicators by watching all the videos recorded during the interview. As a result, 28 non-verbal visual cues associated to facial and upper-body modalities have been provided. Even though this is only an initial approach, it already establishes that *Comfortability* can be recognized visually and the behaviours that might be of guidance to identify it. Further studies should consider these findings and tackle each one of this cues more thoroughly to confirm its functionality and provide more information about them (e.g., its frequency, intensity, etc.).

In addition, this study highlights that people do not perceive, assimilate or react in the same manner to the same circumstances. During the interviews, whilst some people found extremely uncomfortable being directly questioned ("in bad faith") about their career achievements and dreams, others found it extremely comfortable. And similarly, whilst some people were extremely uncomfortable when the robot complimented them about their clothes, others felt very comfortable and flattered. The participants' personality traits and attitude towards robots provided an insight on such differences. Although no significant correlation was found, people more "open-to-new experiences" self-report more often feeling more comfortable during their interaction with the robot. Additionally, people who associate robots (in general) with "discomfort" attributes, exhibited a significantly lower Valence during the interview. In regards to the robot's nature and its implication on the *Comfortability* the participants felt, it was found that the majority of the interviewees imagined they would have felt the same *Comfortability* intensities (in the *comfortable* moments) if the interviewer would have been either a human or a different robot. Nonetheless, it was also found that during the *uncomfortable* situations, they would have felt more uncomfortable if the agent would have been a human or an agent more human-like than iCub. Even

though these considerations are derived from responses to an exercise of imagination rather an actual interaction with such agents, they already provide interesting observations. They highlight that the perceived nature of an agent, determined by its appearance, is expected to have a relevant impact on its partner's *Comfortability*.

Summarizing, this experiment has helped us observe first-hand how a natural interaction between a robot and a human evolves, indicating the way people behave, self-report their *Comfortability* and perceive the agent after such interaction. Also, it has analyzed the biggest limitations that emerge when evaluating someone's *Comfortability* (and possibly any other internal state) providing evidence and alternative solutions. Particularly, given self-reports as well as automatic Valence estimations are not entirely reliable to address someone's *Comfortability*, human visual analysis is introduced as a more trustworthy measure. At the same time, this study has reinforced our understanding of *Comfortability* as a very subjective state, which is influenced both by the unfolding of an interaction and by a priori elements (e.g., attitudes towards robots or the robot's perception). Furthermore, this study has collected plenty of ecological multi-modal and diverse data that might be deeper explored in the future. To this point, only the visual aspect has been tackled, however auditory, physiological, verbal and contextual information are still left to explore. It is likely that the more modalities included, the stronger its recognition accuracy. In the end, this paper has discussed in depth many aspects of *Comfortability*, providing evidence of its usefulness in HRI and motivation to keep unravelling it in future steps.

Acknowledgements Alessandra Sciutti is supported by a Starting Grant from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program. G.A. No. 804388, wHiSPER.

Funding Open access funding provided by Istituto Italiano di Tecnologia within the CRUI-CARE Agreement.

Data Availability The data that support the findings of this study are available on request from the corresponding author M.E.L. Redondo. The data are not publicly available due to them containing information that could compromise the participants privacy.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your

intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Aroyo AM, Pasquali D, Kothig A, Rea F, Sandini G, Sciutti A (2021) Expectations vs. reality: unreliability and transparency in a treasure hunt game with iCub. *IEEE Robot Autom Lett.* <https://doi.org/10.1109/LRA.2021.3083465>
- Ball A, Silvera-Tawil D, Rye D, Velonaki M (2014) Group comfortability when a robot approaches. In: International conference on social robotics (pp 44–53). Springer, Berlin
- Barret L (2006) Solving the emotion paradox: categorization and the experience of emotion, personality and social psychology review. *Soc Personal Soc Psychol* 10(1):20–46. https://doi.org/10.1207/s15327957pspr1001_2
- Barrett LF, Adolphs R, Marsella S, Martinez AM, Pollak SD (2019) Emotional expressions reconsidered: challenges to inferring emotion from human facial movements. *Psychol Sci Public Interest* 20(1):1–68
- Barros P, Churamani N, Sciutti A (2020) The facechannel: a fast and furious deep neural network for facial expression recognition. *SN Comput Sci* 1(6):1–10
- Becker-Asano C, Ishiguro H (2011) Evaluating facial displays of emotion for the android robot Geminoid F. In: 2011 IEEE workshop on affective computational intelligence (WACI) (pp 1–8). IEEE
- Brooks D (2017) A human-centric approach to autonomous robot failures. Ph.D. thesis, University of Massachusetts Lowell, Massachusetts
- Carlson J, Murphy R (2005) How UGVs physically fail in the field. *IEEE Trans Robot* 21(3):423–437
- Carpinella CM, Wyman AB, Perez MA, Stroessner SJ (2017) The robotic social attributes scale (RoSAS): development and validation. In: Proceedings of the 2017 ACM/IEEE International conference on human–robot interaction (pp 254–262). Association for Computing Machinery. <https://doi.org/10.1145/2909824.3020208>
- Chatterji N, Allen C, Chernova S (2020) Effectiveness of robot communication level on likeability, understandability and comfortability. In: 2019 28th IEEE international conference on robot and human interactive communication (RO-MAN) (vol 38, pp 56–56). <https://doi.org/10.1038/s42256-019-0103-7>
- Clare C (2012) Communicate : how to say what needs to be said when it needs to be said in the way it needs to be said. National Library of Australia Cataloguing-in-Publication entry
- Costes A, Doucet J, Landrault C, Laprie J (1995) Surf a program for dependability evaluation of complex fault-tolerant computing systems. In: Twenty-Fifth international symposium on fault-tolerant computing, 1995. ‘Highlights from Twenty-Five Years’, (pp 161–161). IEEE Computer Society
- Ekman P (2003) Darwin, deception, and facial expression. *Ann N Y Acad Sci* 1000(1):205–221
- Ekman P (2004) Emotions revealed. *BMJ* 328(Suppl S5)
- Ekman P, Davidson RJ, Friesen WV (1990) The Duchenne smile: emotional expression and brain physiology: II. *J Personal Soc Psychol* 58(2):342
- Fernández-Dols JM, Ruiz-Belda MA (1995) Are smiles a sign of happiness? Gold medal winners at the Olympic Games. *J Personal Soc Psychol* 69(6):1113
- Fischer K (2019) Why collaborative robots must be social (and even emotional) actors. *Tech Res Philos Technol* 23(3):270–289
- Goleman D (2006) Social intelligence: the revolutionary new science of human relationships. Editorial Kairos
- Gosling S, Rentfrow P, Swann W (2003) A very brief measure of the big-five personality domains. *J Res Personal* 37(6):504–528. <https://doi.org/10.1016/S0092-6566>
- Hamacher A, Bianchi-Berthouze N, Pipe AG, Eder K (2016) Believing in BERT: Using expressive communication to enhance trust and counteract operational error in physical human–robot interaction. In: 2016 25th IEEE international symposium on robot and human interactive communication (RO-MAN), (pp 493–500). IEEE
- Heer J, Yan J, Lim A. Joyful or nervous? A dataset of awkward, embarrassed and uncomfortable smiles. Whisper Project
- Hoffmann L, Krämer NC, Lam-Chi A, Kopp S (2009) Media equation revisited: do users show polite reactions towards an embodied agent? In: International workshop on intelligent virtual agents (pp 159–165). Springer, Berlin
- Honig S, Oron-Gilad T (2018) Understanding and resolving failures in human–robot interaction: literature review and model development. *Front Psychol* 9:861
- Ivonin L, Chang HM, Diaz M, Catala A, Chen W, Rautenberg M (2015) Traces of unconscious mental processes in introspective reports and physiological responses. *Plos ONE* 10(4):e0124519
- Koay K, Walters M, Dautenhahn K (2005) Methodological issues using a comfort level device in human–robot interactions. In: ROMAN 2005. IEEE international workshop on robot and human interactive communication (pp 359–364). IEEE
- Lee MK, Kiesler S, Forlizzi J, Srinivasa S, Rybski P (2010) Gracefully mitigating breakdowns in robotic services. In: 2010 5th ACM/IEEE international conference on human–robot interaction (HRI), (pp 203–210). IEEE
- Liu L, Liu Y, Gao XZ (2021) Impacts of human robot proxemics on human concentration-training games with humanoid robots. In: Healthcare, 7, p 894. Multidisciplinary Digital Publishing Institute
- Metta G, Sandini G, Vernon D, Natale L, Nori F (2008) The iCub humanoid robot: an open platform for research in embodied cognition. In: Proceedings of the 8th workshop on performance metrics for intelligent systems (pp 50–56)
- Mirnig N, Giuliani M, Stollnberger G, Stadler S, Buchner R, Tschelegi M (2015) Impact of robot actions on social signals and reaction times in HRI error situations. In: International conference on social robotics, pp 461–471. Springer, Berlin
- Mirnig N, Stollnberger G, Miksch M, Stadler S, Giuliani M, Tschelegi M (2017) To err is robot: how humans assess and act toward an erroneous social robot. *Front Robot AI* 4:21
- Packard C, Boelk T, Andres J, Edwards C, Edwards A, Spence PR (2019) The pratfall effect and interpersonal impressions of a robot that forgets and apologizes. In: 2019 14th ACM/IEEE international conference on human–robot interaction (HRI) (pp 524–525). IEEE
- Picard R (2003) Affective computing: challenges. *Int J Hum Comput Stud* 59(1–2):55–64. <https://doi.org/10.1016/S1071-5819>
- Rosenthal-von der Pütten AM, Krämer NC, Herrmann J (2018) The effects of humanlike and robot-specific affective nonverbal behavior on perception, emotion, and behavior. *Int J Soc Robot* 10(5):569–582
- Ragni M, Rudenko A, Kuhnert B, Arras K (2016) Errare humanum est: erroneous robots in human–robot interaction. In: 2016 25th IEEE international symposium on robot and human interactive communication (RO-MAN), (pp 501–506). IEEE
- Redondo MEL (2019) Comfortability detection for adaptive human–robot interactions. In: 2019 8th international conference on affective computing and intelligent interaction workshops and demos (ACIIW), (pp 35–39). IEEE
- Redondo MEL, Vignolo A, Niewiadomski R, Rea F, Sciutti A (2020) Can robots elicit different comfortability levels? In: Wagner AR et al (eds) Social robotics. ICSR 2020. Lecture Notes in Com-

- puter Science (vol 12483, pp 664–675). Springer, Berlin. https://doi.org/10.1007/978-3-030-62056-1_55
- 37. Ross R, Collier R, O'Hare G (2004) Demonstrating social error recovery with agentfactory. In: Proceedings of the third international joint conference on autonomous agents and multiagent systems (vol 3, pp 1424–1425)
 - 38. Rossi A, Holthaus P, Perugia G, Moros S, Scheunemann M (2021) Trust, acceptance and social cues in human–robot interaction (SCRITA). *Int J Soc Robot.* <https://doi.org/10.1007/s12369-021-00844-z>
 - 39. Salem M, Eyssel F, Rohlffing K, Kopp S, Joublin F (2013) To err is human (-like): effects of robot gesture on perceived anthropomorphism and likability. *Int J Soc Robot.* 5(3):313–323
 - 40. Scheff TJ (1990) Socialization of emotions: pride and shame as causal agents. *Res Agendas Sociol Emotions:* 281–304
 - 41. Short E, Hart J, Vu M, Scassellati B (2010) No fair!! an interaction with a cheating robot. In: 2010 5th ACM/IEEE international conference on human–robot interaction (HRI) (pp 219–226). IEEE
 - 42. Sicat S, Chopra S, Li N, Sharlin E (2017) Playing the mirror game with a humanoid: Probing the social aspects of switching interaction roles. In: 2017 26th IEEE international symposium on robot and human interactive communication (RO-MAN), pp 1078–1083. IEEE. <https://doi.org/10.1109/ROMAN.2017.8172437>
 - 43. Steinbauer G (2012) A survey about faults of robots used in robocup. In: Robot Soccer World Cup (pp 344–355). Springer, Berlin
 - 44. Thorndike E (1992) Intelligence and its use. Harper's Mag 140:227
 - 45. Tian L, Oviatt S (2021) A taxonomy of social errors in human–robot interaction. *ACM Trans Hum–Robot Interact (THRI)* 10(2):1–32
 - 46. van der Woerd S, Haselager P (2016) Lack of effort or lack of ability? Robot failures and human perception of agency and responsibility. In: Benelux conference on artificial intelligence (pp 155–168). Springer, Berlin
 - 47. Xu J, Broekens J, Hindriks KV, Neerincx MA (2014) Robot mood is contagious: effects of robot body language in the imitation game. In: AAMAS (pp 973–980)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Maria Elena Lechuga Redondo was born in Spain in 1993. She studied Computer Engineering at the Universidad Autónoma de Madrid, obtaining her bachelor degree (Grado en Ingeniería Informática) in February 2017. Next, she achieved her MSc in Artificial Intelligence at the University of Essex (passing with distinction) in November 2018 and received her PhD in Robotic Brains and Cognitive Sciences at the Italian Institute of Technology in July 2022. She is interested in creating social robots, so they can interact with humans smoothly, not only to enhance their integration in society, but also, to understand humans better. For this reason, she has focused her research on exploring the emotional characteristics that might be the most present in social interactions, introducing Comfortability on a scientific level.

Radoslaw Niewiadomski is currently a Tenure-Track Assistant Professor at the University of Genoa (Italy). He received his Ph.D. degree in Computer Science from the University of Perugia (Italy). His research interests include affective computing, multimodal interaction, and social robotics. He co-authored over 100 peer-reviewed conference and journal papers and has been involved in several EU research projects, e.g., FP6 CALLAS, FP7 ILHAIRE, and H2020 DANCE.

Francesco Rea got his bachelor degree in Software Engineering at the Università di Bergamo Italy. He got the MSc degree in Robotics and Automation with distinction at the Salford University, England. Since 2009, at the Istituto di Tecnologia, he has promoted the research in the field of humanoid robotics. He got his PhD degree in 2013 where he exploited the aspects of perception and cognition in active perception. He is now team leader of the cognitive interaction lab and principal investigator in IIT for different European Projects researching HRI with cognitive robotics intelligent systems.

Sara Incao is a Ph.D Student in Bioengineering and Robotics (Cognitive Robotics, Interaction and Rehabilitation Technologies). She received her master's degree in philosophy at Catholic University of the Sacred Heart in Milan with a thesis concerning the relation between the aesthetic object and the spectator. After a period of research on the Virtual Reality artwork and the bodily dimension in the digital realm, her project is currently focused on the concept of Self, Artificial Self and subjective experience for cognitive architectures in Human-Robot Interaction. She is also a visiting researcher in the department of Philosophy at The University of Memphis, TN, USA.

Giulio Sandini is a Founding Director of the Italian Institute of Technology where in 2006 he established the department of Robotics, Brain and Cognitive Sciences. As a research fellow and Assistant Professor at the Scuola Normale in Pisa and Visiting Researcher at the Neurology Department of the Harvard Medical School he investigated visual perception and sensorimotor coordination in humans and technologies for Brain Activity Mapping in children with learning disabilities. As a professor of bioengineering at the University of Genova in 1990 he founded the LIRA-Lab (Laboratory for Integrated Advanced Robotics, www.liralab.it) which was to become the birthplace of the iCub humanoid robot. In 1996 he was Visiting Scientist at the Artificial Intelligence Lab of MIT.

Alessandra Sciutti is the head of the CONTACT (COGNITIVE Architecture for Collaborative Technologies) Unit of the Italian Institute of Technology (IIT). After a master's degree in Bioengineering from the University of Genova and a Ph.D. in Humanoid Technologies, she spent two research periods abroad, first at the Robotics Lab of the Rehabilitation Institute of Chicago (USA) and then at the Emergent Robotics Laboratory of Osaka University (Japan). In 2018 she was awarded an ERC Starting Grant, for the project wHiSPer (www.whisperproject.eu), focused on the investigation of shared perception between humans and robots. She published more than 100 papers in international journals and conferences and is currently Associate Editor for several journals on Cognitive Robotics and Human-Robot Interaction. She is the corresponding co-chair of the Technical Committee on Cognitive Robotics of the IEEE Robotics and Automation Society and a Scholar of the ELLIS Society. Sciutti received many awards, such as the title "Tecnovisionarie" (2021) and a recognition by Fortune Italy (2022) for her research in Robotics and AI. The scientific aim of her research is to investigate the sensory, motor, and cognitive mechanisms underlying human social interaction, with the technological goal of developing robots able to establish mutual understanding with humans.

Terms and Conditions

Springer Nature journal content, brought to you courtesy of Springer Nature Customer Service Center GmbH (“Springer Nature”). Springer Nature supports a reasonable amount of sharing of research papers by authors, subscribers and authorised users (“Users”), for small-scale personal, non-commercial use provided that all copyright, trade and service marks and other proprietary notices are maintained. By accessing, sharing, receiving or otherwise using the Springer Nature journal content you agree to these terms of use (“Terms”). For these purposes, Springer Nature considers academic use (by researchers and students) to be non-commercial.

These Terms are supplementary and will apply in addition to any applicable website terms and conditions, a relevant site licence or a personal subscription. These Terms will prevail over any conflict or ambiguity with regards to the relevant terms, a site licence or a personal subscription (to the extent of the conflict or ambiguity only). For Creative Commons-licensed articles, the terms of the Creative Commons license used will apply.

We collect and use personal data to provide access to the Springer Nature journal content. We may also use these personal data internally within ResearchGate and Springer Nature and as agreed share it, in an anonymised way, for purposes of tracking, analysis and reporting. We will not otherwise disclose your personal data outside the ResearchGate or the Springer Nature group of companies unless we have your permission as detailed in the Privacy Policy.

While Users may use the Springer Nature journal content for small scale, personal non-commercial use, it is important to note that Users may not:

1. use such content for the purpose of providing other users with access on a regular or large scale basis or as a means to circumvent access control;
2. use such content where to do so would be considered a criminal or statutory offence in any jurisdiction, or gives rise to civil liability, or is otherwise unlawful;
3. falsely or misleadingly imply or suggest endorsement, approval , sponsorship, or association unless explicitly agreed to by Springer Nature in writing;
4. use bots or other automated methods to access the content or redirect messages
5. override any security feature or exclusionary protocol; or
6. share the content in order to create substitute for Springer Nature products or services or a systematic database of Springer Nature journal content.

In line with the restriction against commercial use, Springer Nature does not permit the creation of a product or service that creates revenue, royalties, rent or income from our content or its inclusion as part of a paid for service or for other commercial gain. Springer Nature journal content cannot be used for inter-library loans and librarians may not upload Springer Nature journal content on a large scale into their, or any other, institutional repository.

These terms of use are reviewed regularly and may be amended at any time. Springer Nature is not obligated to publish any information or content on this website and may remove it or features or functionality at our sole discretion, at any time with or without notice. Springer Nature may revoke this licence to you at any time and remove access to any copies of the Springer Nature journal content which have been saved.

To the fullest extent permitted by law, Springer Nature makes no warranties, representations or guarantees to Users, either express or implied with respect to the Springer nature journal content and all parties disclaim and waive any implied warranties or warranties imposed by law, including merchantability or fitness for any particular purpose.

Please note that these rights do not automatically extend to content, data or other material published by Springer Nature that may be licensed from third parties.

If you would like to use or distribute our Springer Nature journal content to a wider audience or on a regular basis or in any other manner not expressly permitted by these Terms, please contact Springer Nature at

onlineservice@springernature.com