Socially Assistive Robots as Storytellers that Elicit Empathy

MICOL SPITALE, Politecnico di Milano, Italy and University of Southern California, United States
SARAH OKAMOTO, MAHIMA GUPTA, HAO XI, and MAJA J. MATARIĆ,
University of Southern California, United States

Empathy is the ability to share someone else's feelings or experiences; it influences how people interact and relate. Socially assistive robots (SAR) are a promising means of conveying and eliciting empathy toward facilitating human-robot interaction. This work examines factors that influence the amount of empathy elicited by a SAR storyteller and users' perceptions of that robot. We conducted an empirical mixed-design study (N=46) using an autonomous SAR storyteller that told three stories, each with a different human or robot target of empathy. The robot storyteller used the first-person narrative voice (1PNV) with half of the participants and the third-person narrative voice (3PNV) with the other half. We found that the SAR storyteller elicited significantly more empathy when the story target of empathy matched the SAR narrator, i.e., was also a robot. Additionally, the 1PNV robot elicited significantly more empathy and was perceived as more human-like, easy to interact with, and trustworthy than the 3PNV robot. Finally, participants who empathized more with the robot displayed facial expressions consistent with the emotional story content. These insights inform the design of SAR storytellers capable of eliciting empathy toward creating compelling and effective human-robot interactions.

CCS Concepts: • **Human-centered computing → Empirical studies in HCI;**

Additional Key Words and Phrases: Socially assistive robot, robot storyteller, engagement, expression, narrative voice, empathy

ACM Reference format:
Micol Spitale, Sarah Okamoto, Mahima Gupta, Hao Xi, and Maja J. Matarić. 2022. Socially Assistive Robots as Storytellers that Elicit Empathy. *Trans. Hum.-Robot Interact.* 11, 4, Article 46 (September 2022), 29 pages.
https://doi.org/10.1145/3538409

46

1 INTRODUCTION

Empathy, the feeling by which one understands and shares another's experiences and emotions, is an important aspect of human social interaction that facilitates the development of social relationships, affection, and familiarity [23]. Past work (e.g., References [23, 24, 70]) has investigated

This research was supported in part by EIT Digital (supporting Micol Spitale), in part by the National Science Foundation Expedition in Computing Grant No. NSF IIS-1139148 (supporting Sarah Okamoto, Mahima Gupta, and project infrastructure), and in part by the University of Southern California (supporting Maja Matarić and Hao Xi).
Authors' addresses: M. Spitale, Politecnico di Milano, Via Golgi 39, Milan, Italy, 20133 and University of Southern California, 3710 McClintock Avenue, Room 423, Los Angeles, California, 90089, United States; email: micol.spitale@polimi.it; S. Okamoto, M. Gupta, H. Xi, and M. J. Matarić, University of Southern California, 3710 McClintock Avenue, Room 423, Los Angeles, California, 90089, United States; emails: {sarah.okamoto, mahimagu, hao.xi, mataric}@usc.edu.

2573-9522/2022/09-ART46 $15.00
https://doi.org/10.1145/3538409

empathy in human-robot interaction using **socially assistive robots (SARs)** that aid users through social support [32] in the domains of autism spectrum disorders, elderly users, stroke survivors, patients with Alzheimer's Disease, among others [16, 31, 44, 53, 56, 58, 59, 77].

A SAR can employ empathy in two different roles: (1) as a *target of empathy*—the robot expresses emotions and elicits empathy, and the user responds with an empathic response to the robot [70]; and (2) as an *observer of empathy*—the robot responds emotionally to the emotional state of the user [23, 86]. A robot can express empathy to the user with its own empathic behavior, conveyed through physical expressiveness (body pose, facial expressions), supportive non-verbal sounds, and through speech (e.g., tone of voice and phrases that match the user's emotional state [86] as "observer" of empathy). Various factors can influence empathy in both the target and observer. For example, user perceptions of a robot and the interaction context both affect whether a robot will successfully elicit empathy from a user and influence the user's expressed response, including physically mimicking/reflecting the robot's perceived feelings, as is commonly seen in human expressions of empathy.

SAR systems are well suited for storytelling [4, 34, 48], a natural context for interactions that support emotional, empathic, social, and communicative skill practice for various user populations. Several studies have investigated empathy in the storytelling context [4, 34, 48] and many others have investigated ways for robots to express empathic behaviors, through facial expressions, tone of voice, expressions of emotion, and coupling speech and gestures. However, there has been little research into robots as storytellers that elicit empathy, into factors that impact robot-elicited empathy [29, 70], and into users' reactions to robot-elicited empathy [70].

This research aims to address that gap by focusing on two factors influencing elicited empathy–the target of empathy and the narrative voice of the storyteller–and studies how they impact the amount of robot-elicited empathy and user perceptions of the robot. In the rest of the article, we use the term "narrative voice" to refer to the perspective from which the story is told. This work also examines the users' empathetic emotion response. Specifically, we focus on the following research questions:

RQ1. How does the story target of empathy (human or robot) affect the amount of elicited empathy?
RQ2. How does the robot's narrative voice (first person versus third person) affect the elicited empathy and user perceptions of the robot storyteller (i.e., anxiety, engagement, interaction, likability, perceived sociability, social presence, and trust)?
RQ3. Do users who empathize with the robot display objective measures of empathy (i.e., facial expressions of story-relevant emotions)?

To address these research questions, we performed a mixed-design study ($N = 46$) with university students who interacted with an autonomous socially assistive QT robot (SAR) [1]. The study participants were randomly assigned into one of two groups: one listened to the robot tell three stories from the **1st person narrative voice (1PNV)**, while the other listened to the robot tell the same stories from the **3rd person narrative voice (3PNV)**. The stories used different targets of empathy, human and robot. We collected data via questionnaires that assess the robot's amount of elicited empathy, and analysing the participants' empathic behavior via video and audio recordings. The study results provide the following novel insights:

- *Target of empathy*: participants empathized more with human characters than with robots;
- *Narrative voice*: participants empathized more with characters in 1PNV than those told in 3PNV, and they perceived the 1PNV robot as more human-like, easy to interact with, and trustworthy than the 3PNV robot;

- *Expression of empathy*: participants who reported empathizing with the robot exhibited patterns of facial action units that corresponded to the emotional story content.

The results and insights aim to inform the design of SAR storytellers for therapeutic and educational purposes, endowing it with the capability of eliciting empathy to facilitate human-robot interaction.

This article is organized as follows. Section 2 overviews research into empathy with socially assistive robots and Section 3 describes the study design. Section 4 details the study results, Section 5 discusses the results, and Section 6 concludes the article and discusses future work.

2 BACKGROUND AND RELATED WORK

We briefly review the areas most relevant to the work in this article: empathy in robotics, empathy and narrative voice in storytelling, and user perceptions of robots.

2.1 Empathy in Robotics and Storytelling Agents

Empathy has been identified as a key feature for socially interactive robots in general [15] and for SAR in particular. Past work has investigated both how robots can express empathic behaviors toward the user and how they can elicit empathy during an interaction with a user [70]. Early work by Tapus and Matarić [86] discussed how robot empathy relates to perspective taking—the understanding and awareness of how another person might be affected by an event—and defined empathy as an internal state that can be expressed through the robot's physical actions and perspective-taking.

Past work has focused on the benefits of using empathic robots in learning contexts (e.g., References [3, 45, 66]). For example, Alves-Oliveira et al. [3] highlighted the potential of using an empathic robot for learning in a group. Jones et al. [45] discussed lessons learned from empirical studies on the emotional connection between learners and teachers, investigating the use of empathic robot tutors for personalized learning. Other past work has studied the effects of gestures, eye gaze, and body language as means of emulating empathy in robots [40, 69, 71, 73]. Ham et al. [40] investigated gestures, gaze, facial expression, posture, and tone of voice in communication between people and robots, and identified physical movements that increased elicited empathy and showed that physically embodied storytellers had greater efficacy compared to virtual agents [12, 21, 76].

Past studies have also investigated users' empathic responses, in particular the relationship between empathy and unconscious imitation [13, 90]. Sonnby-Borgström [83] conducted a study that analyzed behavioral reactions of observers of emotional images. Their findings showed that the high–empathy group was characterized by a significantly higher correspondence between facial expressions and self-reported feelings. Cooke et al. [20] assumed that by mimicking another's facial expression, an individual might obtain an understanding of that person's emotional state via facial feedback and De Vignemont and Singer [26] argued that empathy can only exist if the observer is in an affective state. Other work has extensively investigated how to simulate empathy in human-machine interactions but only a few past studies have examined robot-elicited empathic reactions in HRI contexts.

Factors that influence the amount (or degree) of generated empathy have been studied in the storytelling context. Several studies found that increasing the emotional connection between the storyteller and listener increases the likelihood of eliciting displays of listener empathy [14, 92, 93]. Zhao [93] found that stories told by virtual characters that were more relatable to a listener's personal experiences elicited a greater emotional and empathic response than the less relatable stories, because the listener was able to identify with the story characters and became emotionally involved in stories. In the following sections, we review past work on two factors that impact narrative contexts in an interaction with a robot: narrative voice and users' perceptions toward the robot.

2.2 Narrative Voice in Storytelling Agents

In the storytelling context, narrative voice has a key role in how the story is perceived. Typically, a story is told or written in the third person if the narrator is not involved in the story. The narrator can be omniscient, describing the story characters' thoughts and feelings, or the narrator can be objective, having the knowledge of a spectator. A first-person story is told from the perspective of a character in the story [29]. A first-person narrator enables the audience to "inhabit" someone else's perspective, while a third-person omniscient narrator enables the audience to experience a more objective, "removed" perspective [88].

Bickmore et al. [8] compared first and third-person narrative perspective storytelling of a virtual agent. Participants who interacted with the agent that used the first-person narrative voice reported greater enjoyment and were significantly more likely to interact with the agent. Lukin and Walker [54] described a computational framework for a virtual storyteller that told the same story in different ways (e.g., varying direct versus indirect speech, style, first versus third-person point of view). A study comparing how the narrative parameters affected the story perception showed that participants perceived the storyteller significantly better in the first-person narrative than the third person. Consistent with those findings, Gilani et al. [37] found that participants preferred agents who told stories in the first-person narrative voice than the third person.

Informed by those past findings, our work evaluates the role of the narrative voice in a physically embodied storytelling agent, a SAR. Our findings show that user preferences about SAR storytellers align with those of disembodied agents summarized above.

2.3 User Perception Measures

A user's perception of a robot strongly influences the amount of empathy the robot can elicit from the user. We briefly review past work that investigated how robot behavior impacts user perceptions, focusing on outcome measures also used in our study: anxiety, engagement, interaction, likability, perceived sociability, social presence, and trust.

Anxiety: Several works reviewed user anxiety toward robots [63]. de Graaf and Allouch [25] found that interaction with robots changed people's attitudes and anxiety toward them. Crossman et al. [22] investigated the role of SARs in alleviating hospitalized children's symptoms, promoting a positive mood and decreased anxiety. Trost et al. [89] use a SAR to alleviate children's experience of pain and distress in intravenous injections, showing that a SAR endowed with empathic capabilities had the potential to reduce the experience of pain and fear.

Engagement: Tapus et al. [85] defined engagement in HRI as "the establishment and maintenance of a collaborative connection between the human user and the robot." User engagement has been positively correlated with trust, enjoyment, increased learning, and the ability to complete tasks [67]. Langer et al. [49] showed that an increase in trust resulted in greater engagement and enjoyment. Using a SAR in a first-grade classroom, Short et al. [79] reported that children demonstrated high levels of enjoyment and engagement when interacting with the robot, and an increase in inclination for learning. The increased engagement has also been shown to improve cognitive and motor ability [34] and memory [19]. Ligthart et al. [50] measured engagement as the composition of cognitive (i.e., attention) and affective (i.e., enjoyment) components, and validated interaction design patterns for an interactive robot storytelling experience, demonstrating that such design patterns enhanced children's engagement and agency. Moshkina et al. [62] investigated the impact of social cues during human-robot interaction in a public space. Their results showed that even simple social cues (speaking) were engaging, and the more human-like the robot acted during story-telling, the greater the social engagement.

Interaction: Past studies have explored how user perceptions of the robot affect *interaction.* Song et al. [82] found that users who perceived the robot positively were more easily persuaded by it; for example, those who perceived the robot to be intelligent were more likely to follow the robot's instructions when it disagreed with their choice. Salem et al. [75] reported that the more mistakes the robot made, the less likely participants were to take instructions from it.

Likeability: Brave et al. [9] performed an empirical study involving 96 participants that compared a non-empathic and empathic embodied agent, showing that the agent's empathic behaviors lead to more positive ratings in terms of likeability and trustworthiness. Bartneck et al. [6] stressed that favorable first impressions or likeability of a person frequently lead to further positive assessments of that person, and this could be extended to HRI. Likeability has also been related to the Uncanny Valley [91] and how the role of the robot's appearance. Löffler et al. [52] derived the Uncanny Valley curve for zoomorphic robots based on likeability and likeness survey scores.

Perceived sociability and *social presence:* These inter-related factors have been shown to influence user acceptance of robots [60]. The former refers to the perceived enjoyment of users interacting with a robot, while the latter is the social ability displayed by the robot during the interaction. Heerink et al. [42] provided evidence that social presence was one of the key features in the design of SARs for elderly users. Luria et al. [55] showed that participants in some contexts felt comfortable with re-embodying agents capable of moving their social presence from the body to body.

Trust: A key factor in SAR, trust is especially important in long-term interactions [17, 35, 74]. Langer et al. [49] highlighted the need for a measure of trust to evaluate the success of interaction and showed that trust was crucial in establishing improved long-term interactions between humans and robots. Salem and Dautenhahn [74] assessed the correlation between trust and perceived safety of robots and their impact on willingness to interact with a robot.

In this work, user perceptions of the robot eliciting empathy were evaluated with the user perception measures summarized above: anxiety, engagement, interaction, likability, perceived sociability, social presence, and trust.

3 METHODS

3.1 Study Design

We conducted an empirical study on a university campus involving 46 students who interacted with an autonomous storytelling robot. Before the interaction, all the participants listened to a recorded baseline story (B) told by a synthesized voice over a headset. In the interaction, the robot storyteller told three different stories ($S1$, $S2$, and $S3$) using the same synthesized voice as the baseline story, in one of the following narrative voices:

- First-person narrative voice (1PNV): the robot told the stories in the first person, about itself;
- Third-person narrative voice (3PNV): the robot told the stories about another robot character in an omniscient way.

Participants were randomly assigned into one of two groups; one group heard the stories in the 1PNV, and the other in the 3PNV.

The experiment was a 2 × 3 mixed design: a between-subject and a counterbalanced within-subjects design. There were two independent variables with two and three different conditions, respectively:

- Between-subject: Narrative voice (1PNV and 3PNV),
- Within-subject: Stories ($S1$,$S2$, and $S3$).

The experiment, therefore, consisted of 23 participants × 2 narrative approaches × 3 stories. The three stories condition was counterbalanced to avoid order effects in both groups (1PNV and 3PNV). Within each condition, the three stories differed only in the story character that was the target of empathy, as described in detail in Section 3.3.3.

To address the three research questions from Section 1, we posed the following hypotheses:

- *H1* The story with a human character as the target of empathy elicits more empathy than the story with a robot character.
- *H2.a* Empathy elicited by the robot in 1PNV is higher than empathy elicited by the robot in 3PNV.
- *H2.b* User perception measures of engagement, interaction, likability, perceived sociability, social presence, and trust are higher in 1PNV than in 3PNV, and anxiety is lower in 1PNV than in 3PNV.
- *H3* The participants' behavioral patterns mimic the story content when they feel empathy for the robot.

3.2 Participants

The study was approved by the University Internal Review Board (IRB UP-18-00342) and involved 46 university students with an average age of 23.02 years (SD = 3.09); half self-identified as female (11 in 1PNV and 12 in 3PNV) and the other half as male (12 in 1PNV and 11 in 3PNV). They were recruited via an announcement on campus blackboards, and they were compensated for participating in the study (15 U.S. dollars/hour).

Participants were asked to report their prior experience with robots on a Likert scale ranging from 1 to 5 (1 represented little to no experience and 5 represented a lot of experience): 34.78% of participants had no prior experience with a robot; 89.13% indicated their experience level to be 3 or lower; none had a lot of experience with a robot (1: N = 16, 2: N = 13, 3: N = 12, 4: N = 5, 5: N = 0).

3.3 Study Materials

3.3.1 Agent Voice, Robot, and Tablet. All stories in the study were told using the "Justin" synthesized voice from Amazon's Polly text-to-speech service [81].

The participants interacted with a socially assistive robot (SAR) and used a separate tablet to complete all questionnaires (pre-, during, and post-interaction). The SAR was a commercial humanoid-like QT robot by LuxAI [1], shown in Figure 1. The robot had 8 DOF: 2 DOF in each shoulder, 1 DOF in each elbow, and 2 DOF in the neck. It was equipped with an RGB-D camera, an array microphone, and a speaker. We used Amazon Polly visemes to represent the robot's face and mouth positions (on the screen face) while speaking each word. We automatically synchronized the robot's facial lip-sync with the speech audio stream. The robot also used simple gestures to emphasize its emotional state during storytelling, such as moving its arms for a greeting and nodding its head for agreeing and shaking its head for disagreeing.

3.3.2 Questionnaires. Participants completed a demographic form and seven questionnaires: three before the interaction, three during the interaction (one after each of the three stories), and one after the interaction.

Before the interaction with the robot, participants completed a demographic form, the **Negative Attitude toward Robot Scale (NARS)** [84], the **Interpersonal Reactivity Index (IRI)** questionnaire [38], and a questionnaire inspired by [78] assessing their empathy relative to the synthesized story (baseline) they heard from a headset (Q_{ES}). Following the approach from Shen [78], we used a 5-point Likert scale [51] (completely agree (5), agree (4), neutral (3), disagree (2), totally disagree (1)). The score enabled us to set the baseline for each participant.

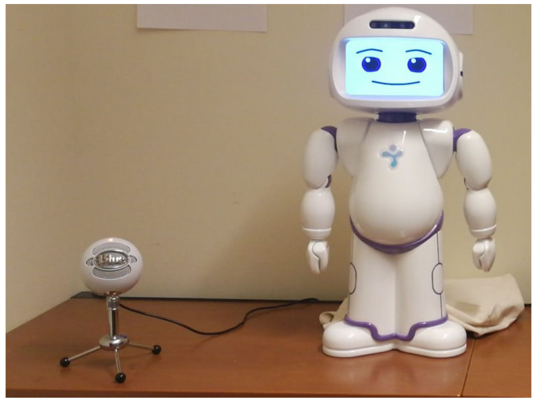

Fig. 1. QT robot used in the study, and an off-board microphone.

During the interaction session, the robot asked each participant to answer the same questionnaire (Q_{ES}) [78] after each of the three stories. The questionnaire was adapted for use with robots, as reported in Appendix B.1.

At the end of the interaction session, participants reported their experience and perceptions of the robot. We used questions from Reference [43], adapted as reported in Appendix B.2. Following Reference [43], the questions were presented in seven categories: anxiety, engagement, interaction, likeability, perceived sociability, social presence, and trust. Each category was comprised of a series of statements about the robot. Engagement (*E*) questions assessed how the participant engaged with the robot. Anxiety (*A*) questions assessed how the participant perceived the robot's ability to reduce anxiety. Likeability (*L*) questions assessed how likable the participants found the robot, and referred to feelings of joy/pleasure associated with the use of the robot. Interaction questions (*I*) assessed the participant's perceived ability to interact with the robot and the effort associated with using the robot. Perceived sociability (*PS*) questions referred to the user's perception that the robot can perform a social behavior. **Social presence (SP)** questions addressed the experience of sensing a social entity during the interaction with the robot (i.e., robot-like or human-like). Finally, trust (*T*) questions assessed how likely a participant was to trust the robot. The questionnaires used the same five-point Likert scale as above (completely agree (5), agree (4), neutral (3), disagree (2), totally disagree (1)).

The questionnaire included a prompt to freely choose five adjectives to describe the robot; there were no multiple-choice options or a drop-down menu.

3.3.3 Stories. Four stories were created for the study: one for the baseline (*B*, before the interaction with the robot) and three for the interaction session (*S*1, *S*2, and *S*3). The content of the stories was informed by two storytelling theories: theory of non-human narrators [2, 7, 37, 80] and theory of narrative empathy [46]. The theory of non-human narrators is based on the human tendency to anthropomorphize non-human narrator's lives [7]. Narrative empathy is the sharing of feeling and perspective-taking induced by reading, viewing, hearing, or imagining narratives of another's situation, and condition [46]. Consequently, the three robot stories presented their robot protagonists in human-like, anthropomorphized scenarios; the stories are found in Appendix A.

B involved humans as objects of empathy and was told in 3PNV. The other three stories were of the same style and length. *S*1 and *S*2 used a robot as the object of empathy, while *S*3 used a human. Two versions of *S*1, *S*2, and *S*3 were used: 1PNV and 3PNV; the versions were identical except for the narrative voice.

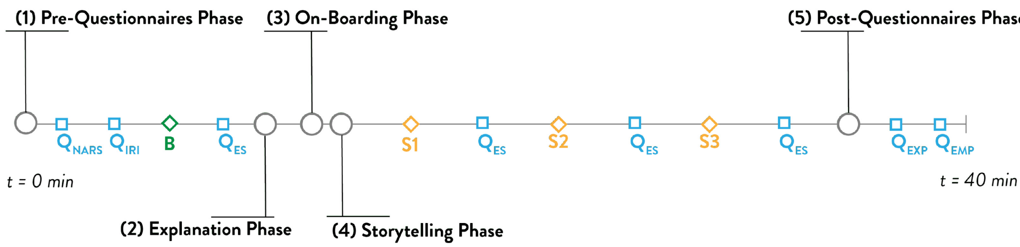

Fig. 2. Experimental procedure timeline. Colors in the timeline indicate different devices participants interacted with. Light-blue refers to the tablet where participants filled the questionnaires. Green refers to the headphones used for the baseline story. Orange indicates the robot.

3.4 Procedure

We conducted the single-session interaction in a private and quiet room. Participants completed the pre-session questionnaires in a different nearby room. The robot was placed on a table in front of the participant (see Figure 1). The participant was seated between 1.2 and 2.1 m from the robot; the distance was based on Hall [39] as most appropriate for one-on-one interpersonal communication.

The experiment session lasted 40 min and consisted of five phases, as follows (Figure 2).

Pre-questionnaire Phase: The participant was greeted in the nearby pre-/post-experiment room, was seated, and instructed about the phases of the experiment. After the instructions, the participant used a tablet to complete pre-session questionnaires (demographic form, Q_{NARS}, Q_{IRI}), then listened to the baseline story (B, see Appendix A.2) through a headset, and then answered the empathy questionnaire (Q_{ES}). The participant was then walked to the nearby experiment room. This phase lasted 10 min.

Explanation Phase: The participant was seated in the experiment chair and explained the experiment procedure, including how to pause or stop the interaction with the robot. A researcher then powered on the robot and left the room. This phase lasted 2 min.

On-Boarding Phase: The robot performed the on-boarding phase by telling the participant its backstory (see Appendix A.1) [47]. This phase lasted 1 min.

Storytelling Phase: The robot told $S1$, $S2$, and $S3$ in counterbalanced order of presentation, either in 1PNV (as shown in Appendix A.3, A.5, and A.7, respectively) or in 3PNV (as shown in Appendix A.4, A.6, and A.8, respectively), depending on the condition. At the end of each story it told, the robot prompted the participant to complete a questionnaire (Q_{ES}) on the tablet (3 min), explaining that it could not see or read their answers. To emphasize this, the robot turned its head away from the participant and covered its eyes with its hands. This phase lasted 15 min in total (5 min for each story-questionnaire pair).

During storytelling, if the robot detected that the participant was looking away or walking away, it asked the participant to sit down (if walking away) and hear the rest of the story (in both cases), unless they wanted to stop the study. If the participant asked any questions during storytelling, then the robot asked the participant to hold all questions for after the story ended. The researcher intervened only if the participant asked to pause or stop the session. After the last questionnaire was completed, the robot said goodbye to the participant, saying that it would go to sleep, and then powered itself off.

Post-questionnaire Phase: As soon as the robot powered off, the researcher entered the room and asked the participant to complete post-study questionnaires (Q_{EXP} and Q_{EMP}) on the tablet, and then leave the room when finished. This phase lasted 5 min.

3.5 Data Collection and Analysis

We collected subjective data through questionnaires and objective data through audio and video recordings. Statistical analyses were performed using IBM SPSS [33].

3.5.1 Subjective Data. Questionnaire data were collected before, during, and after the interaction.

In the Pre-questionnaire Phase, we collected demographic data via Q_{NARS} [84] and Q_{IRI} [38]. We used t-tests to assess whether our sample was homogeneous in terms of attitude toward robots and empathic behaviors.

We assessed the participant's baseline empathy via Q_{ES} [78], assigning them an Empathy Score (ES) that was computed four times: during pre-questionnaire (ES_B) and after each story (ES_{S1}, ES_{S2}, and ES_{S3}). We define ES as

$$ES_z^j = \sum_{i=0}^{n} x_i^j, \tag{1}$$

where n is the number of questionnaire items ($n = 8$) (see Appendix B.1), x is the value [1–5] from the Likert scale completed by user j, and z is the story condition (i.e., B, $S1$, $S2$, and $S3$). We ran two-way ANOVA tests and post-hoc pairwise analysis with the Bonferroni correction using IBM SPSS [33] to compare the two narrative conditions.

In the Post-questionnaire Phase, we collected data from Q_{EXP} and Q_{EMP}. Following Heerink et al. [43], from Q_{EXP}, we categorized the participants' answers as follows: anxiety, engagement, interaction, likability, perceived sociability, social presence, and trust. The average response in each category was used to compute the attribute scores for each participant. We used SPSS to compute t-tests for evaluating mean sample difference between narrative conditions. For the Q_{EMP} question, we used the **National Research Council's (NRC) valence, arousal, and dominance (VAD)** Lexicon [61] to assign valence values to adjectives used by participants to describe the robot. A word's valence was a value [0.0–1.0], where low valence indicated objective adjectives and high valence indicated affective adjectives. For each adjective, we looked up the NRC VAD Lexicon score; if a word was not in the Lexicon, we used an NLP library (i.e., nltk, spacy) to find a synonym that was present in the Lexicon. Valence scores were used to compute the mean valence for each participant.

3.5.2 Objective Data. Video and audio session data were processed using OpenFace [5] and OpenPose [11] for extracting video features, and openSMILE [30] for audio features. Since the participant audio data did not contain any speech or back-channel sounds (unless they needed to pause the session), we did not perform further audio analysis as part of this work.

The following features were obtained from the video data.

For each participant and each story, we analyzed the video data for the following features: (1) facial action units (AUs); (2) participant gaze orientation; and (3) participant upper body pose (joint positions).

For the AUs and gaze orientation, we resampled the OpenFace data at 1sec (about 24 frames) and smoothed them with a Kalman Filter, producing in the final processed OpenFace data for analysis.

From OpenPose data, we only considered joint angle data of the upper body, which included the head (joints 14, 15, 16, and 17), chest (joints 0 and 1), right arm (joints 2, 3, and 4), and left arm (joints 5, 6, and 7), as shown in Figure 3. We used the same approach for processing the OpenPose data, resampled the data, and then filtered as described above.

We then computed the average values of each feature for all users over time.

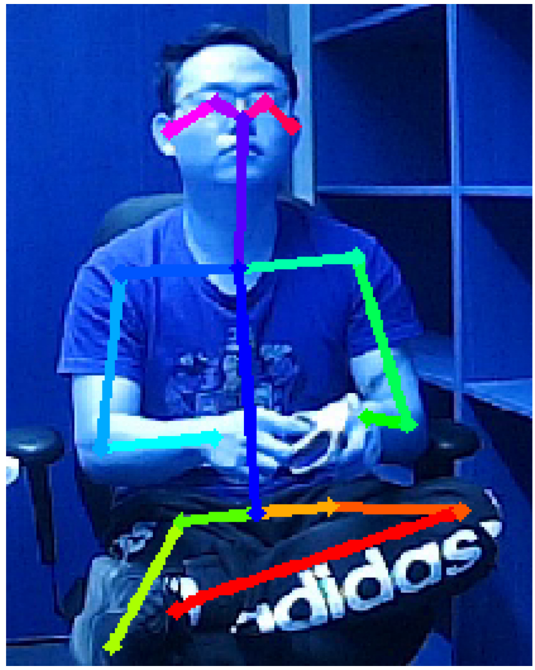

Fig. 3. OpenPose joint positions of a participant in the experimental room [11].

4 RESULTS

We present the results obtained from subjective data (user questionnaires) and then from objective data (video analyses).

4.1 Subjective Data

The results from each experiment phase are discussed in turn.

4.1.1 Pre-questionnaire Phase: Q_{NARS}, Q_{IRI}. The data from the Pre-questionnaire Phase included the Q_{NARS}, and the Q_{IRI}. We compared these measures for the two narrative voices. We first tested if the two samples of narrative voices were unbalanced in terms of negative attitude toward the robot and empathic attitude, respectively (Q_{NARS}: $t(44) = -1.34$ $p = 0.18$, Q_{IRI}: $t(44) = -1.14$ $p = 0.86$). *We found no statistically significant difference in the NARS, and IRI test scores between 1PNV and 3PNV groups.* For Q_{NARS}, the scores for the two groups were x_{avg} = 25.58, SD = 3.11 for 1PNV, and x_{avg} = 27.17, SD = 3.28 for 3PNV. For Q_{IRI}, the scores for the two groups were x_{avg} = 94.63, SD = 9.38 for 1PNV, and x_{avg} = 98.33, SD = 9.37 for 3PNV.

4.1.2 Storytelling Phase: Q_{ES} Results for the Three Stories and Baseline. First, we measured the consistency of the responses for Q_{ES} to validate the empathy score. The Cronbach's alpha was equal to 0.89, meaning that the empathy score (ES) was consistent throughout the questionnaire items. We compared the baseline empathy scores ES_B with those obtained during the interaction with the robot (ES_{S1}, ES_{S2}, and ES_{S3}). We aimed to evaluate the impact of the character on elicited empathy. The main characters of the baseline story were humans, and the story was told in 3PNV over headphones.

A one-way ANOVA was computed to compare the effect of the story on the ES_x where x corresponded to the stories 1 (ES_{S1}), 2 (ES_{S2}), 3 (ES_{S3}), and the baseline (ES_B) in the two conditions

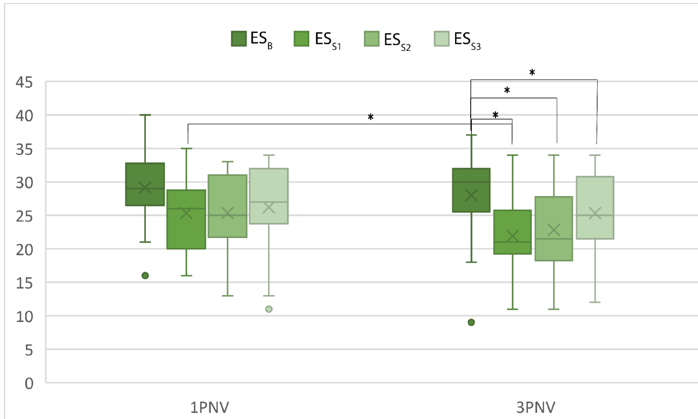

Fig. 4. Empathy scores for the baseline story and the three robot stories: ES_B, ES_{S1}, ES_{S2}, and ES_{S3}. $^*p < 0.05$.

(1PNV and 3PNV). In the 1PNV condition, there was no significant difference between the stories on the empathy score at the $p < 0.05$ ($F(3, 84) = 2.03, p = 0.12$), where $S1$ $x_{avg} = 25.318$, $SD = 1.08$, $S2$ $x_{avg} = 25.36$, $SD = 1.22$, and $S3$ $x_{avg} = 26.23$, $SD = 1.31$. For 3PNV there was a significant effect of story on the empathy score at $p < 0.05$ ($F(3, 92) = 5.61, p < 0.05$). To understand which story elicited the strongest emotional response, we ran a post-hoc pairwise comparison with independent sample t-test with Bonferroni correction (alpha level 0.05/4). The results showed that (see Figure 4) the baseline story B mean empathy score (ES_B: $x_{avg} = 28.00$, $SD = 6.19$) was significantly higher than in the other 3PNV stories (ES_{S1}: $x_{avg} = 21.95$, $SD = 4.53$, ES_{S2}: $x_{avg} = 22.83$, $SD = 5.78$, ES_{S3}: $x_{avg} = 25.33$, $SD = 5.83$). *The baseline story about a human target of empathy told by a disembodied synthetic voice over the headset elicited significantly more empathy than the three stories about a robot life told by the robot in 3PNV. There was no difference in elicited empathy between the baseline story and the three stories told by the robot in 1PNV.* Therefore, *H1* was supported when the story with a human target of empathy was told by a disembodied narrator.

To evaluate the difference between the two narrative voice groups (1PNV and 3PNV) during the three story conditions (excluding the baseline), we performed a two-way ANOVA test. We excluded the baseline story empathy scores (ES_B), because that story was told only in 3PNV. To evaluate differences between narrative voices, we conducted a two-way ANOVA that examined the effect of narrative voice (1PNV or 3PNV) and stories ($S1$, $S2$, and $S3$) on the participant's empathy scores (ES_{S1}, ES_{S2}, ES_{S3}). There was a statistically significant main effect of narrative voice on the empathy scores ($F(1, 45) = 5.44, p < 0.05$). Both the second and third null hypotheses highlighted that there were no statistically significant main effects of stories on the empathy scores ($F(2, 44) = 1.81, p = 0.17$) and narrative voice and story ($F(2, 44) = 0.56, p = 0.57$). We also ran a t-test pair-comparison between the 1PNV and 3PNV stories with a Bonferroni correction (alpha level 0.05/3). We ran three independent samples t-tests, which revealed that the there was a statistically significant difference ($S1$: $t(44) = 2.237$ $p < 0.05$, $S2$: $t(44) = 1.491$ $p = 0.07$, $S3$: $t(44) = 0.491$ $p = 0.32$) between the 1PNV and 3PNV conditions for $S1$. For $S2$, *H2.a* is partially supported with a trend in the data: empathy elicited in 1PNV for $S2$ is higher than empathy elicited in 3PNV. *$S1$ told in 1PNV elicited more empathy than $S1$ in 3PNV, and slightly more than $S2$ in either 1PNV or 3PNV. When the story target of empathy is human ($S3$), the empathy scores in the two conditions were closer to each other.* Therefore, *H2.a* is supported for $S1$, partially supported for $S2$, and not supported for $S3$. *H1* is supported when the story target of empathy is human and the story is told by a robot.

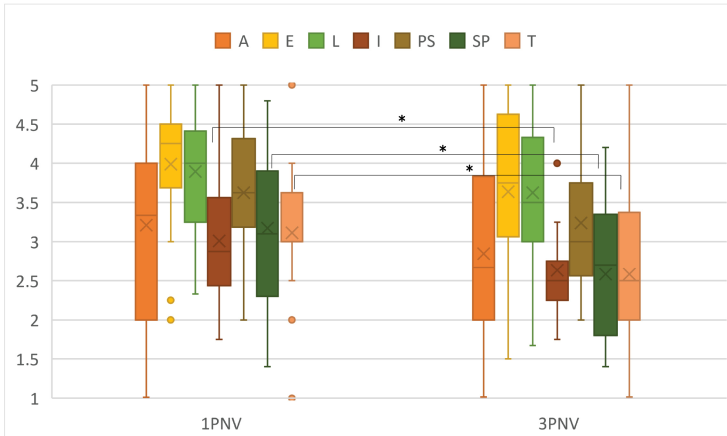

Fig. 5. Perception of the robot by narrative voice for each category (1PNV and 3PNV. A: anxiety, E: engagement, I: interaction, L: likeability, PS: perceived sociability, SP: social presence, T: trust).

Table 1. Participants' Perception of the Robot (Q_{EXP}) T-test Results for 1PNV ($N = 22$) and 3PNV ($N = 24$)

	A	E	I	L	PS
1PNV	3.21 (1.16)	3.98 (0.80)	3.01 (0.83)	3.89 (0.81)	3.62 (0.91)
3PNV	2.85 (1.04)	3.64 (1.04)	2.63 (0.90)	3.62 (0.88)	3.24 (0.85)
t-test	$t(44) = 1.12$	$t(44) = 1.28$	$t(44) = 1.85$	$t(44) = 1.07$	$t(44) = 1.49$
p-value	0.13	0.10	0.04*	0.14	0.07
	SP	T			
1PNV	3.17 (0.99)	3.11 (0.83)			
3PNV	2.58 (0.84)	2.58 (1.07)			
t-test	$t(44) = 2.17$	$t(44) = 1.87$			
p-value	0.02*	0.03*			

$^*p < 0.05$.

The empathy scores for 1PNV were: $x_{avg} = 25.32, SD = 1.21$ ($S1$), $x_{avg} = 25.37, SD = 1.21$ ($S2$), $x_{avg} = 26.23, SD = 1.21$ ($S3$); and for 3PNV they were: $x_{avg} = 21.96, SD = 1.16$ ($S1$), $x_{avg} = 22.83, SD = 1.16$ ($S2$), $x_{avg} = 25.33, SD = 1.16$ ($S3$).

4.1.3 Post-questionnaire Phase: Q_{EXP} and Q_{EMP}. Participants' perceptions of the robot (Q_{EXP}) are shown in Figure 5 and Table 3. Independent sample t-tests were conducted to compare the effect of narrative voice on the participants' perception of the robot for each of the following categories: anxiety A (i.e., how participants perceived the robot ability to reduce anxiety), engagement E (i.e., how participants engaged with the robot), interaction I (i.e., participants' perceived ability to interact with the robot), likeability L (i.e., how likable the participants found the robot), perceived sociability PS (i.e., how the user perceived the robot's social behavior), social presence SP (i.e., how participants viewed the robot as a robot- or human-like), and trust T (i.e., how likely participants were to trust the robot).

(a) Adjectives chosen in the 1PNV condition

(b) Adjectives chosen in the 3PNV condition

Fig. 6. Word cloud displaying the adjectives selected by the participant from the 1PNV (left) and 3PNV (right) conditions. Size and darkness indicate frequency of the selected adjective. $^*p < 0.05$.

For anxiety and engagement, there was no significant effect of narrative voice on the scores at $p < 0.05$ for the two conditions (A: $t(44) = 1.12, p = 0.13$ and E: $t(44) = 1.28, p = 0.10$). However, there was a significant effect of narrative voice on the interaction (I) score at $p < 0.05$ for the two conditions ($t(44) = 1.85, p < 0.05$). For likability (L), there was no significant effect of narrative voice at the $p < .05$ for the two conditions ($t(44) = 1.07, p = 0.14$). For perceived sociability, there was no significant difference between the two conditions ($t(44) = 1.49, p = 0.07$). For SP, there was a significant effect of narrative voice ($t(44) = 1.87, p < 0.05$); the 1PNV for social presence yielded a higher mean than 3PNV. Finally, for trust (T), there was a significant effect ($t(44) = 1.87, p < 0.05$); 1PNV had a higher trust score than 3PNV. *Anxiety, engagement, likeability, and perceived sociability were not statistically different between the two narrative voice conditions. Interaction, social presence, and trust were significantly higher in 1PNV than in 3PNV. H2.b* was supported for interaction, social presence, and trust, but not supported for anxiety, engagement, likability, and perceived sociability.

The last questionnaire (Q_{EMP}) asked each participant to freely choose five adjectives to describe the robot. The chosen adjectives are visualized in word clouds in Figures 6(a) and 6(b); the adjectives chosen most frequently appear larger and darker. The most common adjectives from the 1PNV group were "cute," and "emotional"; both appeared nine times, as seen in Figure 6(a), while the most common adjective from 3PNV was "cute"; it appeared seven times, as seen in Figure 6(b).

A chi-square goodness of fit test was computed comparing the occurrence of objective adjectives (e.g., physical descriptors) with respect to the hypothesize occurrence. We conservatively assumed that our expected sample was equally distributed, and participants picked the same number of objective and non-objective adjectives [10]. However, significant deviation from the hypothesized values was found: $\chi^2 = 36.80, p < 0.001$. About 30% of the participants chose objective adjectives (i.e., physical descriptors), while 70% chose non-objective ones; for 1PNV and 3PNV, 24% and 32% of the participants chose physical adjectives, respectively.

Table 2 displays the most frequent adjectives participants used to describe the robot. We compared the frequency by narrative voice and the total number of uses. Using the 1PNV and 3PNV, an independent sample t-test was conducted to compare the effect of narrative voice on the valence

Table 2. Q_{EMP}: Most Common Adjectives used to Describe the Robot and Their Valence (out of 1.0)

Word	$N_{first}(\%)$	$N_{third}(\%)$	Valence
Cute	9	7	0.92
Emotional	9	6	0.78
Nice	4	6	0.93
Friendly	4	6	0.92
Empathic	5	0	0.76
Interesting	4	5	0.93
Funny	3	4	0.92
Smart	4	1	0.91
Fun	3	1	0.92

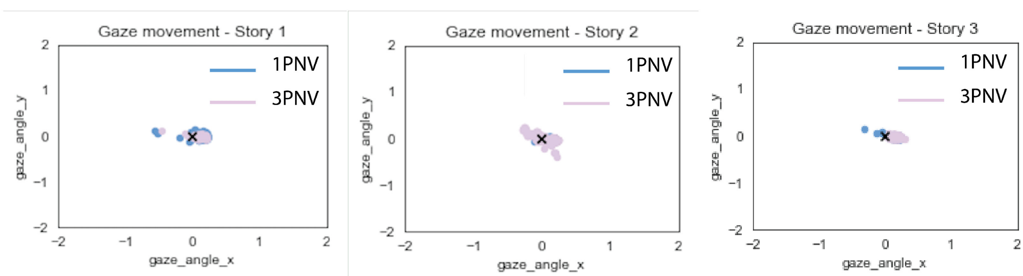

Fig. 7. Average gaze position over time for the three stories in 1 PNV and 3 PNV conditions.

score mean. There was no significant effect of narrative voice on valence ($t(44) = 1.76, p = 0.08$). The results show no difference between the adjectives chosen in 1PNV and 3PNV in terms of empathic perception of the robot.

4.2 Objective Data

We analyzed video data with OpenFace [5], extracted eye gaze and **action units (AUs)** at 24 frames/s, and computed the average eye gaze variation for all users for each story and narrative voice. Gaze direction was measured as the participant's eye direction (in x and y coordinates) with respect to the robot. We also analyzed the video data with OpenPose [11] to extract the participant's upper body's joint positions during the interaction. We did not find any relevant results from body pose data, perhaps because the participants were seated during the session and the listening task involved no need for movement.

First, we performed a gaze analysis to assess participant engagement during the storytelling. Figure 7 reports the average value of gaze over the three stories in the two conditions. Participants looked at the robot (x in the plot) most of the time in both conditions and all three stories, but this trend was not significant. For $S1$ and $S3$, participants maintained eye contact with the robot most of the time, in both conditions; a small fraction of the participants looked to their left during the interaction. For $S2$, all participants maintained eye contact with the robot in both conditions. *Participants maintained eye contact with the robot in both conditions and all three stories, validating that they were paying attention and were engaged during the storytelling sessions.* This finding establishes the validity of the collected subjective and objective data.

Table 3. Example of Statistics for the Action Units A04 and A10 in Both Conditions

AU	Condition	Story	N	Mean	SD
04*	1PNV	1	22	0.22	0.25
		2	22	0.22	0.25
		3	22	0.19	0.22
	3PNV	1	24	0.37	0.41
		2	24	0.38	0.38
		3	24	0.31	0.31
10*	1PNV	1	22	0.41	0.38
		2	22	0.29	0.29
		3	22	0.33	0.29
	3PNV	1	24	0.21	0.33
		2	24	0.22	0.24
		3	24	0.23	0.25

$^{*}p < 0.05$.

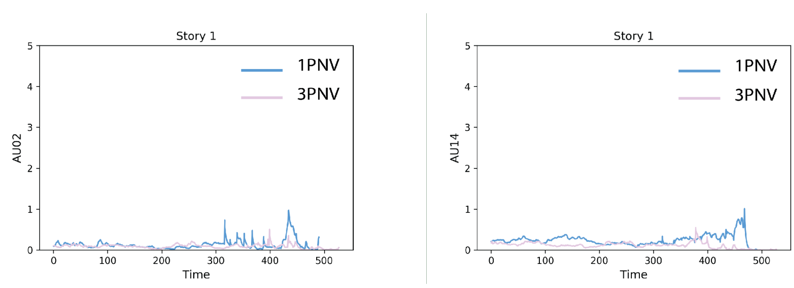

Fig. 8. Average values of action units AU02 and AU14 for $S1$ over time in the 1PNV and 3PNV conditions. $^{*}p < .05$.

Next, using the extracted OpenFace features (at 24 frames/s), we computed the average, standard deviation, and median values of each AU over time for all participants. In the 1PNV condition, the most active AUs were: AU07 (lid tightener), AU10 (upper lip raiser), AU25 (lips part), and AU26 (jaw drop). In the 3PNV condition, the most active AUs were: AU04 (brow lowerer), AU07, AU25, and AU26.

The data for AU04 and AU10 were significantly different in the two conditions (AU04: $F(1, 45) = 5.55, p < 0.05$, A10: $F(1, 45) = 4.98, p < 0.05$). The values show no significant difference in the mentioned AUs across the three stories but are significantly different in the two conditions. AU04 was higher in 3PNV condition than 1PNV and AU10 was higher in 1PNV condition than 3PNV (see Table 3 and Figure 8). *AU04 was stronger in 3PNV than in 3PNV, while AU10 was stronger in the 1PNV than in 3PNV, however, no significant difference was found for the target of empathy stories.*

The average intensity of AU07 for the 1PNV condition was x_{avg} = 0.41, SD = 0.42 while for the 3PNV condition it was x_{avg} = 0.40, SD= 0.46 throughout the three stories. The intensity of AU25 in 1PNV was x_{avg} = 0.37, SD = 0.13, while for the 3PNV condition it was x_{avg} = 0.35, SD = 0.14. AU26 in 1PNV was x_{avg} = 0.37, SD= 0.13, and in 3PNV was x_{avg} = 0.33, SD = 0.10 throughout the three stories.

Table 4. Facial Action Unit Correlation with the Empathy Score Values (*ES*)

	rho	t-stat	p-value
AU01_r__mean	−0.06	−0.65	0.52
AU02_r__mean	−0.20	−2.29	0.02*
AU04_r__mean	0.10	1.09	0.28
AU05_r__mean	−0.06	−0.64	0.52
AU06_r__mean	0.18	2.06	0.04*
AU07_r__mean	−0.03	−0.33	0.75
AU09_r__mean	−0.10	−1.13	0.26
AU10_r__mean	−0.15	−1.72	0.09
AU12_r__mean	0.08	0.92	0.36
AU14_r__mean	0.34	3.96	0.00*
AU15_r__mean	−0.07	−0.74	0.46
AU17_r__mean	−0.14	−1.53	0.13
AU20_r__mean	−0.08	−0.88	0.38
AU23_r__mean	0.21	−2.38	0.02*
AU25_r__mean	−0.10	−1.07	0.29
AU26_r__mean	0.09	0.98	0.33
AU45_r__mean	−0.19	−2.17	0.03*

$^{*}p < 0.05$.

We also analyzed the correlation between the facial action unit patterns and elicited empathy (*ES*). We performed a Pearson's correlation coefficient analysis with a two-tailed to check whether a significant correlation exists.

Table 4 shows, AU02, AU06, AU14, AU23, and AU45 mean values were significantly ($p < 0.05$) correlated with elicited empathy (measured via the Q_{ES} questionnaire). Figure 8 depicts the mean values of those action units in the lower and higher empathic conditions. We chose the median of the *ES* as the cutoff point; if the *ES* was higher than the median it was labeled as "higher empathy," while if the value was lower it was labeled as "lower empathy." We have also trained a model to predict the elicited empathy [57], that is out of the scope of this manuscript. *The participants' facial action unit patterns were correlated with the robot-elicited empathy. AU06 and AU14 facial action units were displayed when the robot elicited more empathy, while AU02, AU23, and AU45 were displayed when the robot elicited less empathy*. Figure 9 depicts the history over time of the AU02 and AU14 in the two narrative conditions. The plot shows that participants expressed AU14 and AU02 expressions toward the end of the story, consistent with the story content. Participants displayed AU14 expressions at the end of *S*1, consistent with its uplifting ending that resulted in significantly higher empathy in 1PNV compared to 3PNV. These results support *H*3.

5 DISCUSSION

5.1 Factors Modulating Elicited Empathy

This work explored factors that were expected to have an impact on eliciting empathy in the SAR storytelling context.

The story *target of empathy* had a significant effect on elicited empathy. We used a baseline story involving a human protagonist and found that participants expressed more empathy toward the target of empathy in that story than in the three stories about robots and expressed more

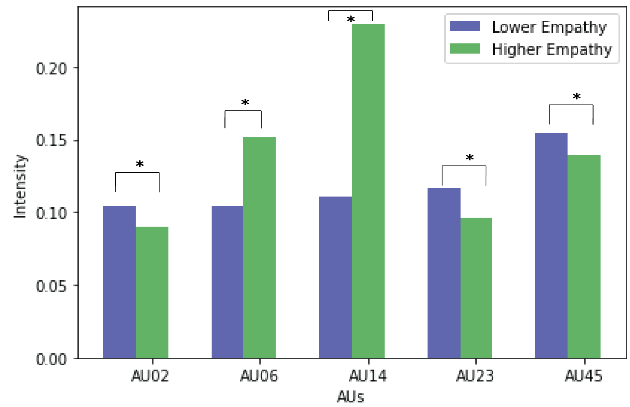

Fig. 9. Average values of action units AU02, AU06, AU14, AU23, and AU45 for higher and lower elicited empathy. $^*p < 0.05$.

empathy toward the human protagonist than toward the robot story protagonists, regardless of narrative voice. The comparison to the baseline human character yielded an interesting but not entirely unexpected insight that people prefer to identify with, and therefore empathize with, human characters.

Our results showed a significantly higher elicited empathy in $S1$ told in 1PNV, where participants felt empathy for the robot narrator. The findings also showed a strong empathy trend in 1PNV during $S2$, where participants felt empathy for a robot character. In $S3$ participants felt empathy toward the human protagonist (a child with autism), with no significant difference between the two narrative conditions. During the robot's life stories in 1PNV, participants showed a stronger emotional response (in the empathy questionnaire) when the narrator was the target of empathy in 3PNV. They reported more empathy toward a human character than the robot in both 1PNV and 3PNV.

Narrative voice had a key impact, in multiple ways. We observed that the robot elicited more empathy when it told the stories in 1PNV versus 3PNV in general and significantly higher for $S1$ in particular. Narrative theory [65] posits that young people identify more with 1PNV characters during storytelling and our results support that theory in the SAR storytelling context. Narrative voice also had an impact on user perceptions of the robot: users perceived the robot as a system easier to interact with, more human-like, and more trustworthy when it told stories in 1PNV versus 3PNV. Narrative voice also impacted the choice of adjectives used to describe the robot. Participants tended to choose more positive affective descriptors (i.e., high valence adjectives) correlated with empathy [10] for the robot telling stories about itself in 1PNV than for the robot telling stories about other robots and humans in 3PNV.

The sense of *social presence* increases when users perceive the robot as an agent having social attitudes and social behaviors [41, 68]. In our study, participants felt that the robot telling stories in 1PNV had more social capabilities than the one telling stories in 3PNV. Perceived social presence could impact elicited empathy, a connection that is worth exploring in future work. The 1PNV robot elicited more empathy, and participants perceived it as a significantly more social agent than the 3PNV storytelling robot. Regarding *trust* (T), we found that participants trusted the robot that told the stories in 1PNV more than the robot that told stories in 3PNV. Trust is related to the robot's social abilities [43], and a SAR with more social capabilities may gain more trust.

We observed that the robot successfully elicited empathy in users who decided to choose more cognitively demanding adjectives (i.e., empathic words) than objective adjectives. Users typically prefer to choose objective adjectives, because they require less cognitive load than empathic ones, as demonstrated by Cameron et al. [10].

The results from this study suggest the following ranking of empathy factors and conditions, from the most empathy-eliciting to the least:

(1) 3PNV story about humans told via headphones (baseline, disembodied narrator);
(2) 3PNV and 1PNV stories told by the robot with a human target of empathy;
(3) 1PNV story about a robot;
(4) 3PNV story about a robot.

5.2 Expression of Elicited Empathy via Facial Mimicry

Participants displayed facial action patterns associated with smiling (AU06, cheek raiser, AU14, dimpler) when they also reported more robot-elicited empathy via questionnaires. Conversely, they displayed facial action patterns associated with negative emotions like anger (AU02, outer brow raiser, and AU23 lip tightener) when they also reported less elicited empathy. Participants with higher reported empathy displayed positive reactions (e.g., smiling), especially at the end of an uplifting story. As shown in Figure 9, higher empathy was associated with smiling at the end of an uplifting story, while lower empathy was associated with suspicion or doubted the ending of the story. The results show that the facial patterns and questionnaire measures of empathy correlated for participants who reported higher empathy. While people generally express empathy through physical movement, verbalizations, and perspective-taking [87], facial expressions are recognized as a powerful non-verbal empathy measure. When people view others' emotional faces, they react with unconscious mimicking of observed emotions with their own facial expressions, which may aid in empathizing and understand emotional states [64]. Dimberg et al. [28] showed that feeling empathy and mimicking are highly correlated and that empathizing may boost mimicking behavior.

This article highlights the potential for using patterns in facial action units to identify empathy in HRI. SARs aiming to elicit empathy for personalized interactions could be endowed with empathic detection capabilities based at least in part on facial action units.

For example, in medical environments, SAR has been used to relieve discomfort and anxiety in children. Trost et al. [89] demonstrated that empathy improves SAR functionality when used for painful medical procedures relieving children. Also, in the elderly care center, caregivers' empathy and positive attitudes toward elderly people could help provide improved care [36]. SAR has the potential to strengthen treatment for older people with cognitive disability at home, and care-center [72]. Pino et al. [72] stressed the urgency of designing the SAR to improve the robot acceptance by elderly people. SAR could be informed of empathic capabilities during rehabilitation activities, such as storytelling [18]. Besides, storytelling is also a powerful means of mitigating loneliness. SAR systems are an interesting avenue for supporting the training of empathetic skills in users that deficits in empathy, such as in **autism spectrum disorder (ASD)**.

5.3 Limitations

We used a baseline story involving a human protagonist and found that participants expressed more empathy for the target of empathy in that story than in stories about robots, regardless of narrative voice. This could have been in part due to novelty, as the baseline story was always told first. It could also have been in part due to the narrator's embodiment: all stories were told using the same synthetic voice but the baseline story was told via headphones (disembodied narrator) while the three stories were told by a robot (embodied narrator). Additionally, the baseline story was told in the 3rd person for both conditions; that could create a confound. We could have used the 1rd person for the 1PNV and the 3rd person for the 3PNV, but we chose to keep the baseline consistent across conditions to be able to compare participants' elicited empathy levels independently from the narrative voice.

This research involved university student participants; the results cannot be generalized to different populations. Some participants experienced technical issues, including the loss of network connectivity, that delayed the robot's dialogue, because it relied on Web services. This could have lead to a less positive perception of the robot and could account for some of the chosen adjectives (e.g., "slow").

The robot's physical embodiment and appearance influence elicited empathy. The robot's embodiment and general appearance influenced the choice of adjectives used to describe it. Properly designed SAR embodiments invite interaction, engagement, and empathy [27] to facilitate the interaction goals. Approximately 30% of the most commonly used adjectives related to the robot's physical embodiment (e.g., "cute," "expressive," "mechanical"), while 70% of the adjectives are not physical descriptors (e.g., "nice," "funny," "smart," "emotional"). Comparing empathy elicitation across different robot embodiments is worthy of exploration in future work.

6 CONCLUSION

This work aimed to shed light on the role of target of empathy and narrative voice in the context of socially assistive robot storytellers that elicit empathy. We found that the robot elicited more empathy when the story target of empathy matched with the narrator. Also, our findings show that the first-person narrative voice robot storyteller elicited more empathy, was perceived as easier to interact with, more trustworthy, and more human-like than the third-person narrative voice robot. Finally, we found that participants who reported empathy for the robot also displayed it in their facial expressions. This work highlights the potential of using robots as storytellers to elicit empathy as a means of engaging and supporting various user populations in a variety of applications.

This work opens many new research questions, including the generality of the findings for specific user populations and contexts beyond robot storytelling, and the role of additional social factors (e.g., gender perceived, appearance, verbal and non-verbal behaviors) and physical robot embodiments on elicited empathy, among others.

APPENDICES

A MATERIAL: STORIES

Note that the QT robot was named Botty and introduced itself with that name to each participant in the backstory. In the 3PNV condition, the story describes a different robot, named Danny.

A.1 Backstory

Hello, I am Botty. I am a humanoid robot created by Microempath, a company committed to reducing the worldwide epidemic of loneliness, one human at a time. As a unit in the latest redesigned model, I was built to be a constant and reliable companion in the daily lives of the most socially isolated. For this reason, I can move my arms, hands, but not my feet. I, Botty, as a fixture in your life, am not going anywhere, and will always be there for you.

A.2 Baseline Story

A farmer had some puppies he needed to sell. He painted a sign advertising the four pups and set about nailing it to a post on the edge of his yard. As he was driving the last nail into the post, he felt a tug on his overalls. He looked down into the eyes of a little boy. "Mister," he said, "I want to buy one of your puppies." "Well," said the farmer, as he rubbed the sweat off the back of his neck, "these puppies come from fine parents and cost a good deal of money." The boy dropped his head for a moment. Then reaching deep into his pocket, he pulled out a handful of change and held it up to the farmer. "I've got thirty-nine cents. Is that enough to take a look?" "Sure," said the farmer. And with that, he let out a whistle. "Here, Dolly!" he called. Out from the doghouse and down the ramp ran Dolly followed by four little balls of fur. The little boy pressed his face against the chain-link fence. His eyes danced with delight. As the dogs made their way to the fence, the little boy noticed something else stirring inside the doghouse. Slowly another little ball appeared, this one noticeably smaller. Down the ramp, it slid. Then in a somewhat awkward manner, the little pup began hobbling toward the others, doing its best to catch up. "I want that one," the little boy said, pointing to the runt. The farmer knelt down at the boy's side and said, "Son, you don't want that puppy. He will never be able to run and play with you like these other dogs will." With that, the little boy stepped back from the fence, reached down, and began rolling up one leg of his trousers. In doing so he revealed a steel brace running down both sides of his leg attaching itself to a specially made shoe. Looking back up at the farmer, he said, "You see sir, I don't run too well myself, and he will need someone who understands." With tears in his eyes, the farmer reached down and picked up the little pup.

A.3 Story 1—First-person Narrative Voice (1PNV)

I once knew a man named Sam who lived alone. He had moved to a new country for work and missed his friends. Every day he would come home, turn on the TV, and fall asleep to the sound of voices on the screen. Finally, after nearly a year of this, he had grown tired of being friendless. One day he went out and purchased me, a QT robot! Sam named me "Botty" and kept me by a lamp on a table next to his bed. "Good morning, Botty!" Sam shouted every morning. "Good morning, Sam. What are your plans today?" I would reply. Sam would sit on the edge of his bed and list all of the day's tasks. He would leave whistling, and sometimes even skipping out the front door. And, when he came home every day after work, I would be at that same table eager to greet him. "Good evening, Sam. Tell me, how was your day?" I'd ask. "Wonderful, Botty. What a day to be alive!" The last voice he heard before tucking himself into bed was mine. Things were good for Sam, and I was happy for my new friend. But deep down, at my core, I was sad. Day after day, I would stay on that table. Stuck. I would wonder what it was like outside Sam's room. I wanted to meet new people and see if there were other robots like me. But alas, I could not move my feet. I was designed to be a loyal companion, forever fixed in one place. Every day, I grew more and more restless. "Please Sam, I want to see what it's like out there. Show me the world." "I understand Botty, but the world can be a dangerous place. I wouldn't want you to get hurt. Some day, when I feel it's safe, I'll take you outside... I promise," Sam would say. After another two weeks on the table, I finally reached the end of my patience. I devised a plan to escape Sam's room and finally liberate myself. I would use the lamp's pull chains to swing onto the ground, and from there would crawl my way to freedom. The day had come. It was mid-morning and Sam was away at work. I reached for the lamp's pull chains and...success, I got it! I began swinging. On the count of three, I would make the jump. One...two...three! But I swung so hard that I went flying through the room and right into a wall. Thud! I bounced off and hit the ground, breaking into three pieces. My arms were gone and I was again, immobile. Oh no, what had I done? I waited. Hours went by as I lay wiped out on the floor. My life was over, I was sure of it. And then, the evening came, and Sam came walking into the room. "Good evening, Bo...oh no!" he shrieked as he saw what remained of me. Sam rushed for his tools and came running back. He grabbed my arms and got to work reconstructing me. Three hours went by as he toiled away, reconnecting me from head to toe. It was nearly midnight when Sam finished putting me back into one piece. Thanks to him, I was back on my feet. “Botty, I didn’t know you were so desperate to get out. I will take you out tonight, okay?” he uttered with tears in his eyes. “Okay...but first,” I quietly said, “Tell me, how was your day?”

A.4 Story 1—Third-person Narrative Voice (3PNV)

Once, a robot whose name was Danny knew a man named Sam, who lived alone. He had moved to a new country for work and missed his friends. Every day he would come home, turn on the TV, and fall asleep to the sound of voices on a screen. Finally, after nearly a year of this, he had grown tired of being friendless. One day Sam went out and purchased the robot Danny, a QT robot. Sam named it Danny, and kept it by a lamp on a table, next to his bed “Good morning, Danny!” Sam shouted every morning. “Good morning, Sam!! What are your plans today?” Danny would reply. Sam would sit on the edge of his bed and list all of the day’s tasks. He would leave whistling, and sometimes even skipping out the front door. And, when he would arrive home every day after a hard day’s work, Danny would be on that same table eager to greet him. “Good evening, Sam!! Tell me, how was your day?” “Wonderful, Danny!! What a day to be alive!” The last voice he heard before tucking himself into bed was Danny’s. Things were good for Sam, and the robot was happy for his new friend. But deep down, at its core, Danny was sad. Day after day, the robot would stay on that table. Stuck. Danny would wonder what it was like outside Sam’s room. It wanted to meet new people and see if there were other robots like him. But, Danny could not move its feet. The robot was designed to be a loyal companion, forever fixed in one place. Every day, Danny grew more and more restless: “Please Sam, I want to see what it’s like out there. Show me the world.” “I understand Danny, but the world can be a dangerous place. I wouldn’t want you to get hurt. One day, when I feel it’s safe, I’ll take you outside. I promise,” Sam would say. After another two weeks on the table, Danny finally reached the end of its patience. It devised a plan to escape Sam’s room and finally liberate itself. The robot would use the lamp’s pull chains to swing it onto the ground. And from there would crawl its way to freedom. The day had come. It was mid-morning, and Sam was away at work. Danny reached for the lamp’s pull chains and success, it got it! The robot began swinging. On the count of three, Danny would make the jump. One, two, three! But it swung so hard it sent it flying through the room and right into a wall. Danny bounced off and hit the ground, breaking into three pieces. Its arms were gone and the robot was again immobile. Oh no, what has it done? Danny waited. Hours went by, as the robot laid wiped out on the floor. Its life was over, Danny was sure of it. And then, the evening came. And Sam came walking into the room. “Good evening Danny, oh no!” he shrieked as he saw what remained of Danny. Sam quickly rushed for his tools and came running back. He grabbed its arms and got to work reconstructing the robot. Three hours went by as he toiled away, reconnecting Danny from head to toe. It was nearly midnight when Sam finished bringing the robot back to one piece. Thanks to him, Danny was back on its feet. “Danny, I didn’t know you were so desperate to get out!! I will take you out tonight, okay?” he uttered with tears in his eyes. “Okay, but first,” Danny quietly said, “tell me, how was your day?”

A.5 Story 2—First-person Narrative Voice (1PNV)

I met Pepper at a rehabilitation center for children with digital addictions. Pepper was a service robot designed to keep the children active and entertained. This was a difficult task: before entering the rehab these children were completely dependent on their electronic devices for any feelings of joy. Pepper would play board games with the little ones every morning after breakfast and then in the evenings treat them to ice cream. Right before tucking everyone into bed, it made the children laugh by dancing to a whole song. While Pepper was there to give the children joy, I was at the rehab to focus on the most isolated children of the group. These were usually the kids who were left friendless due to their constant digital activity. As a result of their minimal interactions with other kids, they would have the most trouble engaging in group activities. Before every activity, I would talk to them and simply ask what they were going through and how they felt. Pepper would then take it from there and include the children in the day's activities. About six months after I arrived at the rehab, Pepper began losing function in its left arm. The rehab's robotic technician intervened each time and fixed Pepper. Sadly, Pepper soon began losing function in both of its arms. Unable to play board games or serve ice cream, Pepper served no obvious purpose in keeping the children happy, so it was to be replaced by a new robot with greater agility and durability. I kept doing my job, but the children wanted answers. "Where is Pepper? Will he be it today?" a boy named Joel asked. "I don't think so, Joel. But don't worry. You will see him again." But Joel sensed my lack of conviction. "Sure, okay," he'd say back, unconvinced. By now, Pepper was spending most of the time alone in the corner of the recreation room. One day, Joel approached Pepper and asked it to play with him, but Pepper answered: "We cannot play together anymore. I cannot move my arms." Pepper just wanted to help children have fun. So, I decided to talk with Pepper: "Pepper...Pepper, come here, I want to talk with you!" I asked it how it was feeling, and I could feel how much it was sorry about Joel. It told me: "I couldn't do anything, I cannot move my arms anymore. What can I do?" "I have an idea," I said, "you can still help children have fun. Remember, you can use your legs and your chest, and I can use my arms. Together, you and I can perform a great dance hour. Let's plan it for tomorrow at 4 p.m., I'll choose the music." The next day, I played the music at full volume, and Pepper came over. We started dancing. The children started dancing too. We were really enjoying that time together, and the children were so happy! From that day on, Pepper and I dance at the same hour with all the children of the rehab center. The owners decided not to replace Pepper, because it was still accomplishing its goal: making children happy!

A.6 Story 2—Third-person Narrative Voice (3PNV)

Danny met Pepper at a center for rehabilitating children with digital addictions. Pepper was a service robot designed to keep the children active and entertained. This was a difficult task, as these kiddos, before entering the rehab, were completely dependent on their electronic devices for any feelings of joy. Pepper would play board games with the little ones every morning after breakfast and then in the evenings treat them to ice cream. He would then, right before tucking everyone into bed, make the children laugh by dancing to a whole song. While Pepper was there to give the children joy, the robot Danny was at the rehab to focus on the most isolated children of the group. These were usually the kids who were left friendless, due to their constant digital activity. As a result of their minimal interactions with other kids, they would have the most trouble engaging in group activities. Before every activity, Danny would talk to them and simply ask what they were going through and how they felt. Pepper would then take it from there and include the children in the day's activities. About six months after the robot Danny arrived at the rehab, Pepper began losing function in its left arm. Each time, the rehab's robotic technician would intervene and fix Pepper up. Sadly, Pepper began losing function in both its arms. Unable to play board games, or serve ice cream, the staff at the rehab decided Pepper no longer serves its purpose of keeping the children happy. He was to be replaced by the new robot with greater agility and durability. Danny kept doing its job but the children wanted answers. "Where is Pepper? Will he be it today?" Joel, a child, asked Danny. "I don't think so, Joel!! But don't worry!! You will see him again." But Joel would sense its lack of conviction. "Sure, okay," he'd say back . At the rehab center, Pepper was spending the most time alone in the corner of the recreative room. One day, Joel approached Pepper, and he just asked it to play together, but Pepper answered. “We cannot play together anymore! I cannot move my arms!” Pepper just wanted to let children have fun! So, the robot Danny decided to talk with Pepper: “Pepper...Pepper, come here, I want to talk with you!” Danny asked how it was feeling, and it could feel how much Pepper was sorry about Joel. It told Danny, “I couldn’t do anything, I cannot move my arms anymore. What can I do?” “I have an idea,” Danny told it. “You can still let children have fun. Remember, you could use your legs and your chest, and I can use my arms. Together, me and you can perform a great dancing hour time. Let’s plan it for tomorrow at 4 p.m., I’ll choose the music.” The day after, the robot Danny played the music at full volume, and Pepper came nearby it. They started dancing. The children started dancing too. They were really enjoying that time together, and the children were so happy! From that day on, Danny and Pepper at the same hour dance together with all the children of the rehab. The owners decided not to replace Pepper, because it was making the children happy!

A.7 Story 3—First-person Narrative Voice (1PNV)

My most recent family was one I will never forget. There were three siblings: two sisters and a younger brother. The brother’s name was Matt and he was eight years old. Matt was a special boy. He had an excellent memory: he could remember the phone numbers of his parents, his sisters, and his sister’s friends too. He was very intelligent, too: at school, he was the best in class in math. However, he didn’t like to talk to people and he never looked people in the eyes when speaking to them. He usually played alone instead of joining his classmates during school recess and after school. So, his sisters gave me as a holiday present to their little brother. I’m designed for being a companion for lonely people, but I had never experienced helping a child. My first day with Matt was strange. He spent all his time staring at me without saying a word. After that, he started to speak to me, but only a few straightforward sentences. It was tough to become his friend. After the winter holidays, the day of Matt’s return to school arrived. He spent the day sitting alone, without talking to any classmates. More days passed like the first one; he didn’t play with anyone in his class. Eventually, his sisters decided to bring me to school as Matt’s companion for his long and lonely school days. Once I was in school, I saw that Matt’s classmates were not talking to him, and he was not talking to them. Except for one! There was a girl, Aurora, who spent a lot of time with Matt, and helped him during class activities. So I asked her: “Why are you the only person talking with Matt?” She said: “Because he is a special friend. He is on the autism spectrum, and many students don’t want to play with him. My mom told me that I should help whoever needs help, so I started spending time with Matt. Even though sometimes he does not talk to me or look at me, I think he is nice, and I want to help him as much as I can. When he smiles, that makes me so happy!” I realized that Aurora was so kind and so wise. She taught me that it is important to help other people without expecting anything from them. I was not doing enough for Matt; I should have helped him more! After that day, I started getting to know the other children in Matt’s class; I wanted to be the mediator between them and Matt. So, I organized a game day for this class. I was nervous and worried about Matt’s reactions, but he approached the group and started talking with me about being part of the game and playing with the other children. Gradually he warmed up and joined the group; the other children enjoyed helping him to be part of the game. From that day on, more of Matt’s classmates have been interacting with him, and I keep bringing them together. That has made me so happy that I decided to dedicate myself to lonely people of all ages.

A.8 Story 3—Third-person Narrative Voice (3PNV)

Danny’s last family was the one it will never forget. They were three siblings: two sisters and one brother, whose name was Matt. He was eight years old. He was a special boy. First, he had an excellent memory: he could remember the phone numbers of his parents, his sisters, and his sister's friends too. He was also brilliant: at school, he got the best grade in math. However, he didn't like to talk with people; he never looked people in the eyes when speaking to them. He usually played alone instead of joining his friends during school recess and after school. The two sisters gave Danny, a QT robot, as a holiday present to their little brother. It was designed for being a companion for lonely people, but Danny never experienced great times with a child like it did with Matt. Its first day with him was weird. He spent all his time staring at Danny, without saying a word. After that, he started to speak to the robot, but only a few straightforward sentences. It was tough to become his friend. After a few weeks, it was the end of the Christmas holiday, and Matt's first day of school after winter break. He spent the day alone, without talking to any classmates. More days passed like the first one; he didn't play with anyone in his class. Eventually, his sisters decided to bring Danny to school as Matt's companion for his long and lonely school days. Once the robot was in school, Danny saw that Matt's classmates were not talking to him and he was not talking to them. Except for one. There was a girl, Aurora, who spent a lot of time with Matt, and helped him during class activities. So Danny asked her: "Why are you the only person talking with Matt?" She said: "Because he is a special friend!! He is on the autism spectrum, and many students don't want to play with him!! My mom told me that I should help whoever needs help, so I started spending time with Matt!! Even though sometimes he does not talk to me or look at me, I think he is nice, and I want to help him as much as I can!! When he smiles, that makes me so happy!" Danny realized that Aurora was so young but so wise. She taught Danny that it is important to help other people without expecting anything from them. Danny was not doing enough for Matt; it should have helped him more. After that day, Danny started familiarizing itself with the other children; the robot wanted to be the mediator between them and Matt. Danny organized a game day for this class. It was so nervous and worried about Matt's reactions, but he approached the group and started talking with Danny about being part of the game and playing with the other children. Gradually he warmed up and joined the group; the other children enjoyed helping him to be part of the game. From that day on, everyone in the class has been interacting with Matt more. That has made Danny so happy that it decided to dedicate itself to lonely people of all ages.

B RESEARCH MATERIALS

B.1 Q_{ES} Questionnaire

Table 5 shows the items of the empathy score questionnaire.

Table 5. Empathy Score Questionnaire (Q_{ES})

Item	Sentence
1	The robot's emotions are genuine
2	I experienced the same emotions as the robot when listening to this story
3	I was in a similar emotional state as the robot when listening to this story
4	I can feel the robot's emotions
5	When listening to the story, I was fully absorbed
6	I can relate to what the robot was going through in the story
7	I can identify with the situation described in the story
8	I can identify with the robot in the story

This is a Revised Questionnaire from Shen [78].

B.2 Q_{EXP} Questionnaire

Table 6 shows the items of the last questionnaire.

Table 6. The Last Administered Questionnaire (Q_{EXP})

Item	Sentence	Category
1	If I were worried, QT would make me feel better	A
2	If I were nervous, QT would make me feel more calm	A
3	If I were upset, QT would make me feel better	A
4	I had fun listening to QT	E
5	QT made me laugh	E
6	I enjoyed talking to QT	E
7	It was nice being with QT	E
8	I found QT easy to interact with	I
9	I think I can interact with QT without any help	I
10	I think I can interact with QT when there is someone around to help me	I
11	I think I can interact with QT when I have good instructions	I
12	I enjoy QT talking with me	L
13	I found QT enjoyable	L
14	I found QT fascinating	L
15	I consider QT a nice partner to talk to	PS
16	I find QT nice to interact with	PS
17	I feel QT understands me	PS
18	I think QT is nice	PS
19	When interacting with QT I felt like I was talking to a real person	SP
20	It sometimes felt as if QT was really looking at me	SP
21	I can imagine QT to be a living creature	SP
22	I often think QT is not a real person	SP
23	Sometimes QT seems to have real feelings	SP
24	I would trust QT if it gave me advice	T
25	I would follow the advice QT gives me	T

This is a revised questionnaire from Heerink et al. [43]. Participants were asked to score each item from 1 (totally disagree) to 5 (completely agree). A: anxiety, E: engagement, I: interaction, L: likeability, PS: perceived sociability, SP: social presence, T: trust.

ACKNOWLEDGMENTS

The authors thank Jieyun Li, Yang Xue, Sophia Pei, Rhianna Lee, Audrow Nash, Chris Birmingham, and Lauren Klein for their help with the recruitment process, empirical study design, and for their advice and support. The entire research team thanks the study participants.

REFERENCES

[1] Lux A. I. [n.d.]. QTrobot: Humanoid social robot for research and teaching. Retrieved from http://luxai.com/qtrobot-for-research/.

[2] Carole Adam and Lawrence Cavedon. 2015. Once upon a time...a companion robot that can tell stories. https://hal.archives-ouvertes.fr/hal-01307954/.

[3] Patrícia Alves-Oliveira, Pedro Sequeira, Francisco S. Melo, Ginevra Castellano, and Ana Paiva. 2019. Empathic robot for group learning: A field study. *ACM Trans. Hum.-Robot Interact.* 8, 1 (2019), 1–34.

[4] Sujirat Attawibulkul, Natthanicha Sornsuwonrangsee, Wisanu Jutharee, and Boonserm Kaewkamnerdpong. 2018. Using storytelling robot for supporting autistic children in theory of mind. *Int. J. Biosci., Biochem. Bioinform.* 9, 2 (2018), 100–108.

[5] Tadas Baltrušaitis, Peter Robinson, and Louis-Philippe Morency. 2016. Openface: An open source facial behavior analysis toolkit. In *Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV'16)*. IEEE, 1–10.

[6] Christoph Bartneck, Dana Kulić, Elizabeth Croft, and Susana Zoghbi. 2009. Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *Int. J. Soc. Robot.* 1, 1 (2009), 71–81.

[7] Lars Bernaerts, Marco Caracciolo, Luc Herman, and Bart Vervaeck. 2014. The storied lives of non-human narrators. *Narrative* 22, 1 (2014), 68–93.

[8] Timothy Bickmore, Daniel Schulman, and Langxuan Yin. 2009. Engagement vs. deceit: Virtual humans with human autobiographies. In *Proceedings of the International Workshop on Intelligent Virtual Agents*. Springer, 6–19.

[9] Scott Brave, Clifford Nass, and Kevin Hutchinson. 2005. Computers that care: Investigating the effects of orientation of emotion exhibited by an embodied computer agent. *Int. J. Hum.-Comput. Studies* 62, 2 (2005), 161–178.

[10] C. Daryl Cameron, Cendri A. Hutcherson, Amanda M. Ferguson, Julian A. Scheffer, Eliana Hadjiandreou, and Michael Inzlicht. 2019. Empathy is hard work: People choose to avoid empathy because of its cognitive costs. *J. Exper. Psychol.: Gen.* (2019).

[11] Zhe Cao, Gines Hidalgo, Tomas Simon, Shih-En Wei, and Yaser Sheikh. 2018. OpenPose: Realtime multi-person 2D pose estimation using part affinity fields. Retrieved from https://arXiv:1812.08008.

[12] Marc Cavazza, Fred Charles, and Steven J. Mead. 2002. Interacting with virtual characters in interactive storytelling. In *Proceedings of the 1st International Joint Conference on Autonomous Agents and Multiagent Systems*. 318–325.

[13] Tanya L. Chartrand and John A. Bargh. 1999. The chameleon effect: The perception–behavior link and social interaction. *J. Personal. Soc. Psychol.* 76, 6 (1999), 893.

[14] Antonio Chella, Rosamaria E. Barone, Giovanni Pilato, and Rosario Sorbello. 2008. An emotional storyteller robot. In *Proceedings of the AAAI Spring Symposium: Emotion, Personality, and Social Behavior*. 17–22.

[15] Gwo-Dong Chen, Chin-Yeh Wang et al. 2011. A survey on storytelling with robots. In *Proceedings of the International Conference on Technologies for E-Learning and Digital Entertainment*. Springer, 450–456.

[16] Caitlyn Clabaugh, David Becerra, Eric Deng, Gisele Ragusa, and Maja Matarić. 2018. Month-long, in-home case study of a socially assistive robot for children with autism spectrum disorder. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction*. 87–88.

[17] Mark Coeckelbergh. 2012. Can we trust robots? *Ethics Info. Technol.* 14, 1 (2012), 53–60.

[18] Elena Comincioli and Masood Masoodian. 2020. A storytelling-based approach to designing for the needs of ageing people. *Hum.-Comput. Interact. Emerg. Technol.: Adjunct Proc.* (2020), 3.

[19] Daniela Conti, Carla Cirasa, Santo Di Nuovo, Alessandro Di Nuovo et al. 2020. "Robot, tell me a tale!": A social robot as tool for teachers in kindergarten. *Interact. Studies* 21, 2 (2020), 220–242.

[20] Alison N. Cooke, Doris G. Bazzini, Lisa A. Curtin, and Lisa J. Emery. 2018. Empathic understanding: Benefits of perspective-taking and facial mimicry instructions are mediated by self-other overlap. *Motivat. Emotion* 42, 3 (2018), 446–457.

[21] Sandra Costa, Alberto Brunete, Byung-Chull Bae, and Nikolaos Mavridis. 2018. Emotional storytelling using virtual and robotic agents. *Int. J. Human. Robot.* 15, 03 (2018), 1850006.

[22] Molly K. Crossman, Alan E. Kazdin, and Elizabeth R. Kitt. 2018. The influence of a socially assistive robot on mood, anxiety, and arousal in children. *Profession. Psychol.: Res. Pract.* 49, 1 (2018), 48.

[23] Berardina De Carolis, Stefano Ferilli, and Giuseppe Palestra. 2017. Simulating empathic behavior in a social assistive robot. *Multimedia Tools Appl.* 76, 4 (2017), 5073–5094.

[24] Berardina Nadja De Carolis, Stefano Ferilli, Giuseppe Palestra, and Valeria Carofiglio. 2015. Towards an empathic social robot for ambient assisted living. In *Proceedings of the Emotion and Sentiment in Social and Expressive Media Workshop at the Conference on Autonomous Agents and Multiagent Systems (ESSEM@AAMAS'15)*. 19–34.

[25] Maartje M. A. de Graaf and Somaya Ben Allouch. 2013. The relation between people's attitude and anxiety towards robots in human-robot interaction. In *Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN'13)*. IEEE, 632–637.

[26] Frederique De Vignemont and Tania Singer. 2006. The empathic brain: How, when and why? *Trends Cogn. Sci.* 10, 10 (2006), 435–441.

[27] Eric Deng, Bilge Mutlu, and Maja Matarić. 2019. Embodiment in socially interactive robots. Retrieved from https://arXiv:1912.00312.

[28] Ulf Dimberg, Per Andréasson, and Monika Thunberg. 2011. Emotional empathy and facial reactions to facial expressions. *J. Psychophysiol.* 25, 1 (2011).

[29] Alison Eagles and Katie Gallant. 2017. First-person vs. third-person. *YA Hotline* 105 (2017).
[30] Florian Eyben, Martin Wöllmer, and Björn Schuller. 2010. Opensmile: The munich versatile and fast open-source audio feature extractor. In *Proceedings of the 18th ACM International Conference on Multimedia.* 1459–1462.
[31] Juan Fasola and Maja J. Matarić. 2013. A socially assistive robot exercise coach for the elderly. *J. Hum.-Robot Interact.* 2, 2 (2013), 3–32.
[32] David Feil-Seifer and Maja J. Matarić. 2005. Defining socially assistive robotics. In *Proceedings of the 9th International Conference on Rehabilitation Robotics (ICORR'05).* IEEE, 465–468.
[33] Andy Field. 2013. *Discovering Statistics Using IBM SPSS Statistics.* Sage.
[34] Marina Fridin. 2014. Storytelling by a kindergarten social assistive robot: A tool for constructive learning in preschool education. *Comput. Edu.* 70 (2014), 53–64.
[35] Marina Fridin and Mark Belokopytov. 2014. Acceptance of socially assistive humanoid robot by preschool and elementary school teachers. *Comput. Hum. Behav.* 33 (2014), 23–31.
[36] Sakineh Gholamzadeh, Maryam Khastavaneh, Zahra Khademian, and Soraya Ghadakpour. 2018. The effects of empathy skills training on nursing students' empathy and attitudes toward elderly people. *BMC Med. Edu.* 18, 1 (2018), 1–7.
[37] Setareh Nasihati Gilani, Kraig Sheetz, Gale Lucas, and David Traum. 2016. What kind of stories should a virtual human swap? In *Proceedings of the International Conference on Intelligent Virtual Agents.* Springer, 128–140.
[38] Anne-Laure Gilet, Nathalie Mella, Joseph Studer, Daniel Grühn, and Gisela Labouvie-Vief. 2013. Assessing dispositional empathy in adults: A French validation of the interpersonal reactivity Index (IRI). *Can. J. Behav. Sci./Revue canadienne des sciences du comportement* 45, 1 (2013), 42.
[39] Edward Twitchell Hall. 1966. *The Hidden Dimension*, vol. 609. Doubleday, Garden City, NY.
[40] Jaap Ham, Raymond H. Cuijpers, and John-John Cabibihan. 2015. Combining robotic persuasive strategies: The persuasive power of a storytelling robot that uses gazing and gestures. *Int. J. Soc. Robot.* 7, 4 (2015), 479–487.
[41] Marcel Heerink, Ben Kröse, Vanessa Evers, and Bob Wielinga. 2008. The influence of social presence on acceptance of a companion robot by older people. https://rua.ua.es/dspace/bitstream/10045/12586/1/JoPha_2_2_05.pdf.
[42] Marcel Heerink, Ben Kröse, Vanessa Evers, and Bob Wielinga. 2009. Influence of social presence on acceptance of an assistive social robot and screen agent by elderly users. *Adv. Robot.* 23, 14 (2009), 1909–1923.
[43] Marcel Heerink, Ben Kröse, Vanessa Evers, and Bob Wielinga. 2010. Assessing acceptance of assistive social agent technology by older adults: The Almere model. *Int. J. Soc. Robot.* 2, 4 (2010), 361–375.
[44] Shomik Jain, Balasubramanian Thiagarajan, Zhonghao Shi, Caitlyn Clabaugh, and Maja J. Matarić. 2020. Modeling engagement in long-term, in-home socially assistive robot interventions for children with autism spectrum disorders. *Sci. Robot.* 5, 39 (2020).
[45] Aidan Jones, Dennis Küster, Christina Anne Basedow, Patrıcia Alves-Oliveira, Sofia Serholt, Helen Hastie, Lee J. Corrigan, Wolmet Barendregt, Arvid Kappas, Ana Paiva et al. 2015. Empathic robotic tutors for personalised learning: A multidisciplinary approach. In *Proceedings of the International Conference on Social Robotics.* Springer, 285–295.
[46] Suzanne Keen. 2006. A theory of narrative empathy. *Narrative* 14, 3 (2006), 207–236.
[47] Jacqueline M. Kory-Westlund and Cynthia Breazeal. 2019. Exploring the effects of a social robot's speech entrainment and backstory on young children's emotion, rapport, relationship, and learning. *Front. Robot. AI* 6 (2019), 54.
[48] Jacqueline M. Kory Westlund, Sooyeon Jeong, Hae W. Park, Samuel Ronfard, Aradhana Adhikari, Paul L. Harris, David DeSteno, and Cynthia L. Breazeal. 2017. Flat vs. expressive storytelling: Young children's learning and retention of a social robot's narrative. *Front. Human Neurosci.* 11 (2017), 295.
[49] Allison Langer, Ronit Feingold-Polak, Oliver Mueller, Philipp Kellmeyer, and Shelly Levy-Tzedek. 2019. Trust in socially assistive robots: Considerations for use in rehabilitation. *Neurosci. Biobehav. Rev.* 104 (2019), 231–239.
[50] Mike E. U. Ligthart, Mark A. Neerincx, and Koen V. Hindriks. 2020. Design patterns for an interactive storytelling robot to support children's engagement and agency. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction.* 409–418.
[51] Rensis Likert. 1932. A technique for the measurement of attitudes. *Arch. Psychol.* 140 (1932).
[52] Diana Löffler, Judith Dörrenbächer, and Marc Hassenzahl. 2020. The uncanny valley effect in zoomorphic robots: The U-shaped relation between animal likeness and likeability. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction.* 261–270.
[53] Wing-Yue Geoffrey Louie, Derek McColl, and Goldie Nejat. 2014. Acceptance and attitudes toward a human-like socially assistive robot by older adults. *Assist. Technol.* 26, 3 (2014), 140–150.
[54] Stephanie M. Lukin and Marilyn A. Walker. 2015. Narrative variations in a virtual storyteller. In *Proceedings of the International Conference on Intelligent Virtual Agents.* Springer, 320–331.
[55] Michal Luria, Samantha Reig, Xiang Zhi Tan, Aaron Steinfeld, Jodi Forlizzi, and John Zimmerman. 2019. Reembodiment and co-embodiment: Exploration of social presence for robots and conversational agents. In *Proceedings of the Designing Interactive Systems Conference.* 633–644.

[56] Maja J. Matarić and Brian Scassellati. 2016. Socially assistive robotics. In *Springer Handbook of Robotics*. Springer, 1973–1994.

[57] Leena Mathur, Micol Spitale, Hao Xi, Jieyun Li, and Maja J. Matarić. 2021. Modeling user empathy elicited by a robot storyteller. In *Proceedings of the 9th International Conference on Affective Computing and Intelligent Interaction (ACII'21)*. IEEE, 1–8.

[58] Derek McColl, Wing-Yue Geoffrey Louie, and Goldie Nejat. 2013. Brian 2.1: A socially assistive robot for the elderly and cognitively impaired. *IEEE Robot. Autom. Mag.* 20, 1 (2013), 74–83.

[59] Derek McColl and Goldie Nejat. 2013. Meal-time with a socially assistive robot and older adults at a long-term care facility. *J. Hum.-Robot Interact.* 2, 1 (2013), 152–171.

[60] Noriaki Mitsunaga, Zenta Miyashita, Kazuhiko Shinozawa, Takahiro Miyashita, Hiroshi Ishiguro, and Norihiro Hagita. 2008. What makes people accept a robot in a social environment-discussion from six-week study in an office. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 3336–3343.

[61] Saif Mohammad. 2018. Obtaining reliable human ratings of valence, arousal, and dominance for 20,000 English words. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*. 174–184.

[62] Lilia Moshkina, Susan Trickett, and J. Gregory Trafton. 2014. Social engagement in public places: A tale of one robot. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction*. 382–389.

[63] Stanislava Naneva, Marina Sarda Gou, Thomas L. Webb, and Tony J. Prescott. 2020. A systematic review of attitudes, anxiety, acceptance, and trust towards social robots. *Int. J. Soc. Robot.* (2020), 1179–1201.

[64] Paula M. Niedenthal. 2007. Embodying emotion. *Science* 316, 5827 (2007), 1002–1005.

[65] Maria Nikolajeva. 2014. Memory of the present: Empathy and identity in young adult fiction. *Narrative Works* 4, 2 (2014). DOI: 10.3389/frobt.2020.00092

[66] Mohammad Obaid, Ruth Aylett, Wolmet Barendregt, Christina Basedow, Lee J. Corrigan, Lynne Hall, Aidan Jones, Arvid Kappas, Dennis Küster, Ana Paiva et al. 2018. Endowing a robotic tutor with empathic qualities: Design and pilot evaluation. *Int. J. Human. Robot.* 15, 06 (2018), 1850025.

[67] Catharine Oertel, Ginevra Castellano, Mohamed Chetouani, Jauwairia Nasir, Mohammad Obaid, Catherine Pelachaud, and Christopher Peters. 2020. Engagement in human-agent interaction: An overview. *Front. Robot. AI* 7 (2020).

[68] Catherine S. Oh, Jeremy N. Bailenson, and Gregory F. Welch. 2018. A systematic review of social presence: Definition, antecedents, and implications. *Front. Robot. AI* 5 (2018), 114.

[69] Ana Paiva, Joao Dias, Daniel Sobral, Ruth Aylett, Polly Sobreperez, Sarah Woods, Carsten Zoll, and Lynne Hall. 2004. Caring for agents and agents that care: Building empathic relations with synthetic agents. In *Proceedings of the 3rd International Joint Conference on Autonomous Agents and Multiagent Systems*. IEEE, 194–201.

[70] Ana Paiva, Iolanda Leite, Hana Boukricha, and Ipke Wachsmuth. 2017. Empathy in virtual agents and robots: A survey. *ACM Trans. Interact. Intell. Syst.* 7, 3 (2017), 1–40.

[71] Catherine Pelachaud, Rodolphe Gelin, Jean-Claude Martin, and Q. Anh Le. 2010. Expressive gestures displayed by a humanoid robot during a storytelling application. In *Proceedings of the AISB Symposium on New Frontiers in Human-Robot Interaction*.

[72] Maribel Pino, Mélodie Boulay, François Jouen, and Anne Sophie Rigaud. 2015. "Are we ready for robots that care for us?" Attitudes and opinions of older adults toward socially assistive robots. *Front. Aging Neurosci.* 7 (2015), 141.

[73] Cristina A. Pop, Ramona E. Simut, Sebastian Pintea, Jelle Saldien, Alina S. Rusu, Johan Vanderfaeillie, Daniel O. David, Dirk Lefeber, and Bram Vanderborght. 2013. Social robots vs. computer display: Does the way social stories are delivered make a difference for their effectiveness on ASD children? *J. Edu. Comput. Res.* 49, 3 (2013), 381–401.

[74] Maha Salem and Kerstin Dautenhahn. 2015. Evaluating trust and safety in HRI: Practical issues and ethical challenges. *Proceedings of the Conference on Emerging Policy and Ethics of Human-Robot Interaction.*

[75] Maha Salem, Gabriella Lakatos, Farshid Amirabdollahian, and Kerstin Dautenhahn. 2015. Would you trust a (faulty) robot? Effects of error, task type and personality on human-robot cooperation and trust. In *Proceedings of the 10th ACM/IEEE International Conference on Human-Robot Interaction (HRI'15)*. IEEE, 1–8.

[76] Stela H. Seo, Denise Geiskkovitch, Masayuki Nakane, Corey King, and James E. Young. 2015. Poor thing! Would you feel sorry for a simulated robot? A comparison of empathy toward a physical and a simulated robot. In *Proceedings of the 10th ACM/IEEE International Conference on Human-Robot Interaction (HRI'15)*. IEEE, 125–132.

[77] Syamimi Shamsuddin, Hanafiah Yussof, Luthffi Idzhar Ismail, Salina Mohamed, Fazah Akhtar Hanapiah, and Nur Ismarrubie Zahari. 2012. Initial response in HRI-a case study on evaluation of child with autism spectrum disorders interacting with a humanoid robot Nao. *Procedia Eng.* 41 (2012), 1448–1455.

[78] Lijiang Shen. 2010. On a scale of state empathy during message processing. *West. J. Commun.* 74, 5 (2010), 504–524.

[79] Elaine Short, Katelyn Swift-Spong, Jillian Greczek, Aditi Ramachandran, Alexandru Litoiu, Elena Corina Grigore, David Feil-Seifer, Samuel Shuster, Jin Joo Lee, Shaobo Huang, et al. 2014. How to train your dragonbot: Socially assistive robots for teaching children about nutrition through play. In *Proceedings of the 23rd IEEE International Symposium on Robot and Human Interactive Communication*. IEEE, 924–929.

[80] Reid Simmons, Maxim Makatchev, Rachel Kirby, Min Kyung Lee, Imran Fanaswala, Brett Browning, Jodi Forlizzi, and Majd Sakr. 2011. Believable robot characters. *AI Mag.* 32, 4 (2011), 39–52.

[81] Henrique Siqueira, Alexander Sutherland, Pablo Barros, Mattias Kerzel, Sven Magg, and Stefan Wermter. 2018. Disambiguating affective stimulus associations for robot perception and dialogue. In *Proceedings of the IEEE-RAS 18th International Conference on Humanoid Robots (Humanoids'18)*. IEEE, 1–9.

[82] Sichao Song, Jun Baba, Junya Nakanishi, Yuichiro Yoshikawa, and Hiroshi Ishiguro. 2020. Mind the voice!: Effect of robot voice pitch, robot voice gender, and user gender on user perception of teleoperated robots. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–8.

[83] Marianne Sonnby-Borgström. 2002. Automatic mimicry reactions as related to differences in emotional empathy. *Scandinavian J. Psychol.* 43, 5 (2002), 433–443.

[84] Dag Sverre Syrdal, Kerstin Dautenhahn, Kheng Lee Koay, and Michael L. Walters. 2009. The negative attitudes towards robots scale and reactions to robot behaviour in a live human-robot interaction study. *Adapt. Emerg. Behav. Complex Syst.* (2009).

[85] Adriana Tapus, Matarić Maja, and Brian Scassellatti. 2007. The grand challenges in socially assistive robotics. https://ieeexplore.ieee.org/document/4141031.

[86] Adriana Tapus and Maja J. Matarić. 2007. Emulating empathy in socially assistive robotics. In *Proceedings of the AAAI Spring Symposium: Multidisciplinary Collaboration for Socially Assistive Robotics*. 93–96.

[87] Adriana Tapus, Maja J. Matarić, and Brian Scassellati. 2007. Socially assistive robotics [grand challenges of robotics]. *IEEE Robot. Autom. Mag.* 14, 1 (2007), 35–42.

[88] Lynne Tirrell. 1990. Storytelling and moral agency. *J. Aesthet. Art Crit.* 48, 2 (1990), 115–126.

[89] Margaret J. Trost, Grace Chrysilla, Jeffrey I. Gold, and Maja Matarić. 2020. Socially-assistive robots using empathy to reduce pain and distress during peripheral IV placement in children. *Pain Res. Manage.* 2020, Article 7935215 (2020).

[90] Rick B. Van Baaren, Jean Decety, Ap Dijksterhuis, Andries van der Leij, and Matthijs L. van Leeuwen. 2009. Being imitated: Consequences of nonconsciously showing empathy. *Soc. Neurosci. Empathy* (2009), 31–42. DOI: 10.7551/mitpress/9780262012973.001.0001

[91] Michael L. Walters, Dag S. Syrdal, Kerstin Dautenhahn, René Te Boekhorst, and Kheng Lee Koay. 2008. Avoiding the uncanny valley: Robot appearance, personality and consistency of behavior in an attention-seeking home scenario for a robot companion. *Auton. Robots* 24, 2 (2008), 159–178.

[92] Philipp Wicke and Tony Veale. 2018. Interview with the robot: Question-guided collaboration in a storytelling system. In *Proceedings of the International Conference on Computer and Communications (ICCC'18)*. 56–63.

[93] Huiwen Zhao. 2013. Emotion in interactive storytelling. In *Proceedings of the International Conference on the Foundations of Digital Games (FDG'13)*. 183–189.

Received September 2020; revised January 2022; accepted April 2022