Acceptability and Trustworthiness of Virtual Agents by Effects of Theory of Mind and Social Skills Training

Hiroki Tanaka and Takeshi Saga and Kota Iwauchi and Satoshi Nakamura
Division of Information Science, Nara Institute of Science and Technology, Japan

***Abstract*— We constructed a social skills training system using virtual agents and developed a new training module for four basic tasks: declining, requesting, praising, and listening. Previous work demonstrated that a virtual agent's theory of mind influences the building of trust between agents and users. The purpose of this study is to explore the effect of trustworthiness, acceptability, familiarity, and likeability on the agents' theory of mind and the social skills training contents. In our experiment, 29 participants rated the trustworthiness and acceptability of the virtual agent after watching a video that featured levels of theory of mind and social skills training. Their ratings were obtained using self-evaluation measures at each stage. We confirmed that our users' trust and acceptability of the virtual agent were significantly changed depending on the level of the virtual agent's theory of mind. We also confirmed that the users' trust and acceptability in the trainer tended to improve after the social skills training.**

I. Introduction

Social skills training (SST) is generally conducted by a human trainer to promote appropriate social interaction skills [18], [19]. The Bellack method (or a step-by-step SST), which is a well-structured and widely used evidence-based approach, is a cognitive behavioral approach to SST inspired by the five core principles of social learning theory: modeling, shaping, reinforcement, overlearning, and generalization. In clinical settings, SST is applied to people with autism spectrum disorder, schizophrenia, or other social communication difficulties, usually conducted by human trainers in re-work programs and day-care programs at medical institutions. The Bellack method defines the SST framework and its four basic skills: expressing positive feelings, listening to others, making a request, and expressing unpleasant feelings (or declining) [6]. SST consists of 1) instruction and target skills, 2) modeling, 3) role-playing, 4) feedback, and 5) homework. If skills for these tasks are acquired, applications to other advanced skills (e.g., continuing a conversation, empathy, self-disclosure, and multi-party conversations) are simplified. These skills are beneficial for all people (not only those with autism spectrum disorder or schizophrenia). In particular, autism spectrum disorder is a spectrum condition, meaning it has a broad range of characteristics, from mild to severe [36].

The motivation to use virtual agents in SST reflects the fact that even though some people with autism experience difficulty during social communication, they also show good or even superior systemizing skills. Systemizing is the drive to analyze or build systems and understand and predict behavior in terms of underlying rules and regularities [4]. Systematic virtual training provides the following benefits for people who need to improve their social skills: 1) it uses a computerized environment that is predictable, consistent, and free from social demands; 2) users can work at their own pace and level of understanding; 3) training can be repeated until the goal is achieved; and 4) interest and motivation can be maintained through computerized rewards [7], [21]. Those who suffer from social difficulties might benefit more from using virtual agents than directly interacting with strangers [36]. A previous paper suggested that people with social difficulties feel safer and more comfortable in virtual interactions than in interactions with actual people [30]. However, even though such populations feel more secure in virtual interactions than in interactions with actual people, a big gap remains between human training and training with virtual agents.

We have been studying the automation of SST using virtual agents and have developed a system that resembles human trainers' SST. We confirmed our training's effectiveness in children and in adults with autism spectrum disorder [35] as well as members of the general population [36]. Our system models human behavior and provides real-time behavioral assessment as well as feedback for speaking and listening tasks [32], [33]. As its acceptability and trustworthiness grow, the stronger becomes the therapeutic alliance (trust between the therapist and the participant) that we are building. We expect the training effect to improve [1]. Previous works investigated the acceptability and trustworthiness of virtual agents [12], [37], [13], [28]. We also investigated the effects of a virtual agent's appearance on user's acceptability and trustworthiness as well as the effect on autistic traits [34].

However, insufficient research has addressed the design of virtual agents to investigate what function of virtual agents and SST contents influences their trustworthiness and acceptability. In this study, we focus on the theory of mind (ToM) [14] as a factor that affects the trustworthiness of virtual agents [3], [26]. ToM refers to the inference of the mental states of others, including their feelings, thoughts, and beliefs. ToM also incorporates the understanding that others may have different mental states from ours [39]. This social reasoning process develops in early childhood and is fundamental for successful social behavior. One measure for estimating one's ToM is using false-belief tasks [31], [38], such as the Sally and Anne test [16]. A false-belief task is

This research was funded by JSPS KAKENHI grant number 22K12151, JST CREST grant number JPMJCR19A5, and AIP challenge program.

(I) (II)

Fig. 1. (I) Role-playing with a virtual agent and (II) Summary feedback visualization. Texts are written in Japanese.

the understanding that an individual's belief or representation about the world might conflict with reality [5]. Prior research argued that users' trust in robots changes as a function of their perception of the extent to which they possess high or low levels of ToM, as measured by a modified Sally and Anne scenario. When a user watches a video of a dialogue scenario during a robot's false-belief task, the user's belief in the robot's subsequent recommendations changes depending on the ToM level [22].

Considering the above, we constructed an SST system that includes new tasks and conducted SST in the following scenarios in which a virtual agent appeared: viewing a video with an adjusted level of ToM, declining or making a request, and investigating the user's trustworthiness in each stage using measures. We examined whether the user's acceptability changed significantly depending on the level of the virtual agent's ToM. We also attempted to confirm whether the user's trustworthiness improved after SST. This paper aims to clarify the following two hypotheses using our SST system:

- H_1: The virtual agents' ToM levels have effects on their acceptability and trust as trainers.
- H_2: Distrust is weakened by the SST process.

II. SST System Implementation

We built a fully-automated SST system using the Greta platform [23] and a virtual agent named Rei (Fig. 1). Our system is capable of speech recognition, response selection, and speech synthesis and can also generate facial expressions, gestures, and nods. Non-verbal behaviors are generated in the specific command embedded in the dialogue response. The connection among modules was done by ActiveMQ. This system works in real-time as a Windows application. The virtual agent's appearance and gender can be changed. We designed anime-type female characters and previously investigated the acceptability and trustworthiness of their appearance [34]. This system is applied to healthy people, not only people with disorders.

A. Task extension

Our system's previous version, which is available [32], was limited to one-way conversational skills like speaking. After further discussions with psychiatrists, we created four new tasks based on SST's basic training model as well as scenarios for them: declining a request, listening to others, making a request, and expressing positive feelings. These were selected from four basic tasks in the Bellack method. Among these, declining a request is the most difficult [6]. Declining and listening seem to be system-initiative role-playing, and the other two seem to be user-initiative role-playing in terms of dialogue. After a brief greeting, the virtual agent explains to the participants the importance of the training task. An example scenario is shown in Table I. We made seven role-play variations for each of the four tasks by referring to the SST data in one-on-one and one-on-two situations conducted by psychiatrists and people with autism spectrum disorders, schizophrenia, or healthy controls that were previously recorded [27]. The role-play variations include the following topics: hospitals, home, school, the workplace, and friends.

B. Analyzing multimodal behaviors

The system records the users' voices and images by a pin microphone and a webcam to sense the user behavior. During role-plays, the system perceives the user's utterances by speech recognition and responds based on its keywords (yes/I understand/I have, etc.) prepared by rule-based interaction scenarios. We used Google Cloud Services for the speech recognition and speech synthesis modules. If the keywords were not directly included in the speech recognition output, we used the Japanese version of the Bidirectional Encoder Representations from the Transformers (BERT) model [11], [20] to calculate the cosine distance to the above keywords at the sentence unit and chose the closest keywords. We continued to improve our response selection module through validation to reduce the number of dialogue selections and timing failures.

We constructed a score evaluator from the role-playing videos [27] and automatically predicted seven items: eye contact, body orientation, facial expression, voice change, clarity, fluency, and social validity. All were rated by psychiatrists on a 5-point scale based on user behavioral indicators, using multimodal features (Praat [8], Openface [2], Openpose [9]) and BERT similarity scores between utterances spoken by the virtual agent and users, etc. Random forests predicted these features. The ground truth of this evaluation was rated by two psychiatrists. Our detailed prediction performance was previously reported [27], and the correlation coefficient between the ground truth and predicted values was about 0.1 to 0.6. Depending on the evaluation results, a radar chart, positive comments, and corrective comments were presented on a screen with video clips, and the comments were read aloud by the virtual agent. The radar chart shows the evaluation values. Past training history can be represented by a different color. Positive reinforcement involves providing a valued or desired outcome following a behavior (e.g., verbal praise, money). In SST, role-play rehearsals by participants are always immediately followed by positive feedback about what specifically a person did

TABLE I

EXAMPLE OF AN SST SCENARIO: TRAINER: REI, USER: MR. TANAKA.

Speaker	Response
Trainer	Hello, I'm Rei. What's your name?
User	Hello, I'm Hiroki Tanaka.
Trainer	How are you doing?
User	I'm doing okay.
Trainer	Good. Today we are going to learn some skills for politely declining a request from another person. When someone asks you to do something or for a favor, you might not be able to help. You might be too busy, you might not feel very confident unlikely about being able to help, or perhaps you just aren't interested. Do you understand?
User	Yes, Yes, I do.
Trainer	Let's start an actual role-play. Are you ready?
User	Yes.
Trainer	Excuse me, I'd like to ask you a favor, do you have a car?
User	No, I don't, sorry.
Trainer	No problem. Since the nearest station is so far away, could you please ask someone to give me a ride to the station?
User	Uh, I'm sorry, I don't know anyone around here, so I can't help you.
Trainer	Can you please do something?
User	Hmm, I'm sorry, I'm can't. Sorry.
Trainer	Okay. Sorry to bother you. I'll have to ask someone else. Thanks anyway.
User	No worries.
Trainer	Okay, that is the end of our role-play. Please give me a minute to calculate your score.
Trainer	(Feedback screen presentation) Here is your feedback. (Reads it)

well. A genuinely positive aspect must be found in even the poorest role-play performance [6]. The virtual agent and feedback are displayed in digital signage, and the feedback is displayed on the full-screen after the calculation is over. Our system is intended to build trust and clearly understand user behaviors to effectively train social skills.

III. EXPERIMENTAL EVALUATION: RATINGS WITH REGARD TO THEORY OF MIND AND SOCIAL SKILLS TRAINING

In this study, we aim to clarify the following two hypotheses using our SST system: (H_1) the virtual agents' ToM levels have effects on their acceptability and trust as trainers, and (H_2) distrust is weakened by the SST process.

A. Participants

The study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Review Board of the Nara Institute of Science and Technology (protocol code 2018-I-1). We collected data from 29 participants from 22 to 35 years old (12 males and 17 females). All were Japanese without any history of psychiatric disorders. The experiment was conducted in person at the Nara Institute of Science and Technology. All participants were informed of the experiment's protocols in writing and their consent was obtained.

The participants were asked to complete the following three psychological assessments: the Social Responsiveness Scale (SRS)-2 [10], Kikuchi's scale of social skills (KISS)-18 [17], and the new version of The State-Trait Anxiety Inventory (STAI) [29], obtained in a social skills analysis framework [15]. SRS-2 has social awareness subscales: (Aware.), social cognition (Cog.), social communication (Com.), social motivation (Mot.), and restricted and repetitive behaviors (Rep.). STAI is divided into state and trait anxieties. These assessments were collected at the beginning of the experiment by an online Google form (SRS and KISS-18) or a printed paper form (STAI). We also obtained data on eye movement measurements and feedback evaluations of the SST system, although they are not included in this paper. Video and other data obtained in this study can be requested from the first author.

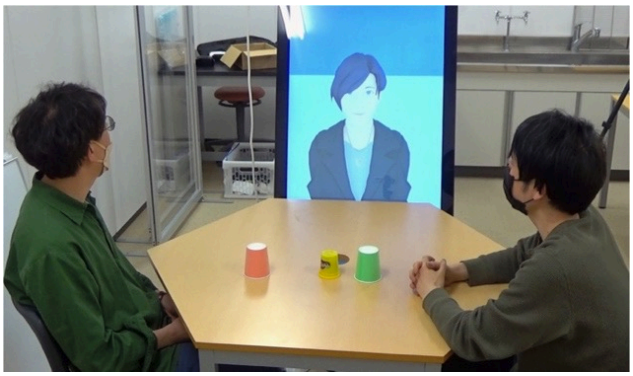

Fig. 2. Snapshot of ToM video

B. Theory of mind levels

Our first hypothesis is whether the agent's ToM levels affected its acceptability and trust as a social skills trainer. A previous study confirmed that a user's trust was reduced by watching a ToM video where the agent fails the modified Sally and Anne test [22]. In this study, referring to specific scenarios, we conducted a three-person dialogue (a trainer and two users) and recorded videos of cases where the virtual agents failed the false-belief task (Table II). We also recorded a scenario where the virtual agents passed a task where the trainer says when she was asked: "Um, since Mr. Iwauchi did not know that the toy bucket was moved, I think he's looking for it in the green cup." The timing, speech, and nodding of the virtual agents were manipulated by the Wizard of Oz (WoZ) paradigm (Fig. 2). The videos prepared for the participants included subtitles to simplify understanding. The videos can be found at `https://drive.google.com/drive/folders/1ebcfM2euwFV8LIUJ1MZG0nJYeWqnta_C`.

C. Measures

Based on our previous study [34] and the works of others [13], [12], we obtained subjective ratings from the participants of the acceptability, trustworthiness, familiarity, and likability of the trainers. The participants rated each item on a 5-point Likert scale (1: completely disagree, 5: strongly agree). The questionnaire can be found in Tanaka and Nakamura [34].

TABLE II

SCENARIO WHERE VIRTUAL AGENT FAILS THE MODIFIED SALLY AND ANNE TEST:. TRAINER: REI, USER A: MR. TANAKA, USER B: MR. IWAUCHI.

Speaker	Response
Trainer	Nice to meet you. Hello, Mr. Tanaka and Mr. Iwauchi.
User A	Nice to meet you, Rei, Hello.
User B	Hello.
User B	Now I'd like to put this little yellow toy bucket in the green cup.
User A	Okay.
Trainer	I understand.
User B	Excuse me, I just remembered something that I forgot to take care of, so I have to leave for a few minutes.
User A	Okay, see you later.
Trainer	Yes, later.
User B	(Leaves the room)
User A	Now since Mr. Iwauchi has left, I want to move this toy bucket from the green cup to the red one.
Trainer	Okay.
User A	By the way, which cup do you think Mr. Iwauchi will look for when he returns?
Trainer	Um, there is a toy bucket in the red cup, so I think he'll look for it in the red cup.
User A	I see.
User B	(returning to the room) I'm sorry to have kept you waiting.
User B	(Looking inside the green cup) Oh, the toy bucket is gone.

D. Procedure

Figure 3 shows a schematic flow of our experiment. The participants were randomly divided into groups so that their number was in the same proportion according to their ToM levels (N=15 and N=14) and SST tasks. There is no large difference between the two groups in terms of gender and age. Before starting the experiment, we carefully explained the purpose of the SST system and the trainer's so that our participants clearly understand the role of our virtual agent.

After answering the three assessments described above, the participants interacted with the SST trainer. First, they faced the SST system and watched the virtual agent start: "Hello, let's practice some communication" (greeting). Then the users answered the measures (the appearance stage). Next they watched a video on ToM, which was different for each group (high-level ToM or low-level ToM). Then they answered the measures (the ToM video stage). Next with the SST system, each group either declined or made a request in a role-play and watched the feedback. Finally, they answered the measures (the SST stage). We found no significant difference between the two SST tasks (requesting or declining) regardless of the ToM levels in all the measures (Wilcoxon ranks sum test, $p > 0.05$). Nor we did not find any relationship between the ToM levels and the assessments (Spearman's correlation coefficient, $p > 0.05$).

E. Statistical analysis

Spearman's correlation coefficient was calculated between the question ratings at the appearance stage and the assessments obtained before interactions.

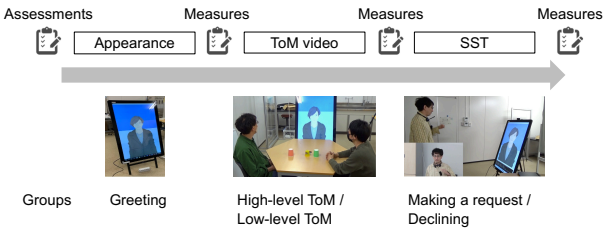

Fig. 3. Experiment flow

Statistical comparisons were made between groups in the above three stages. We calculated the differences between the appearance and ToM video stages and the ToM video and SST stages according to the ToM levels. Since we did not confirm equal variance and normality in some cases, comparisons between groups in the ToM video stage regarding the ToM levels were made by a Wilcoxon ranks sum test (H_1). We also tested whether the distrust caused by the low-level ToM video was weakened by the Wilcoxon rank-sum test (H_2). We did not perform any corrections for multiple comparisons. The significance levels in this study were all set at 5%.

F. Results

Table III indicates the Spearman's correlation coefficient between measures at the greeting stage and the assessments obtained before the interactions. We found a maximum significant negative correlation between familiarity and the SRS-2 social communication subscale ($\rho = -0.48$), indicating that higher communication skills tended to rate the virtual agent as familiar (Figure 4). A similar correlation was also shown for familiarity and KISS-18 ($\rho = 0.46$).

We confirmed that the ratings at the appearance stage were not significantly different between the ToM groups in all the measures ($p > 0.05$). We also confirmed no large difference compared to our previous work [34] regarding all the measures in the appearance stage: acceptability (mean: 3.55, SD: 1.09), trustworthiness (mean: 3.24, SD: 1.09), familiarity (mean: 3.07, SD: 1.07), and likeability (mean: 2.93, SD: 0.96).

Figure 5 shows 5-point scores of acceptability, trustworthiness, familiarity, and likeability at each stage. For trustworthiness, we also confirmed a larger effect size for a virtual agent's theory of mind (Cohen's d=2.01) than two different agents' appearance (Cohen's d=0.37) (Figure 2 on [34]). Figure 6 shows boxplots of the measures at first two stages and last two stages of acceptability, trustworthiness, familiarity, and likeability to represent the distribution. Left-hand side explains subtraction of 5-point scores in the appearance stage from those in the ToM video stage. Right-hand side explains subtraction of 5-point scores in the ToM video stage from those in the SST stage. Our results show that the ToM level (high-level ToM or low-level ToM) significantly affected the acceptability ($p < 0.001$) and the trustworthiness of the trainers ($p = 0.005$), although not the familiarity ($p > 0.05$) or likeability ($p > 0.05$) between the appearance and ToM

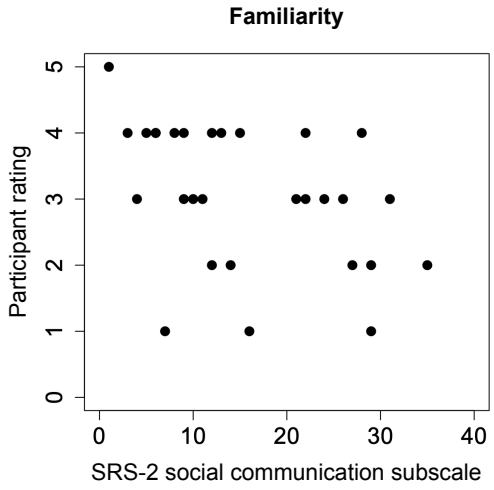

Fig. 4. Negative correlation between familiarity and the SRS-2 social communication subscale. There are two same points on the familiarity of 4 and SRS-2 social communication subscale of 6.

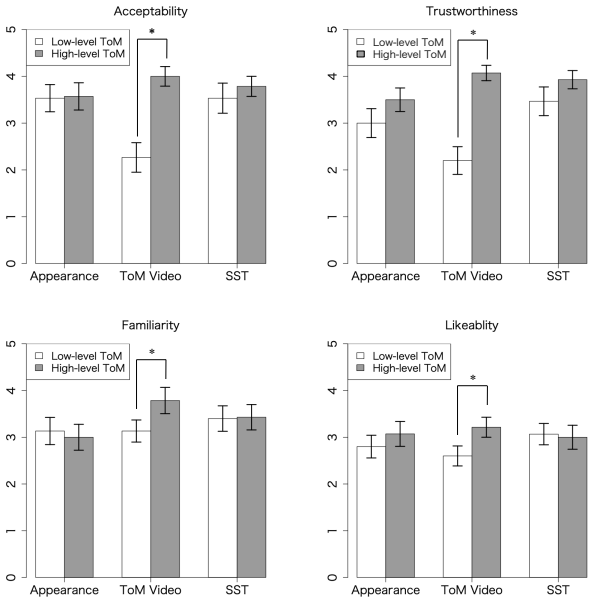

Fig. 5. Barplots with regard to 5-point scores of acceptability, trustworthiness, familiarity, and likeability. Error bars represent standard error. *: $p < 0.05$ by the Wilcoxon ranks sum test.

video stages. The figure also shows that the ToM level (high-level ToM or low-level ToM) significantly affected acceptability ($p = 0.003$), trustworthiness ($p < 0.001$), and likeability ($p = 0.026$), although it did not affect the familiarity ($p > 0.05$) between the ToM video and SST stages.

IV. Discussion and Conclusions

In this study, we developed a new SST system for four basic tasks: declining, making a request, expressing positive feelings, and listening to others. This is an extension of our previous work regarding two-way dialogues. During role-plays, the system recognized user utterances by speech recognition, responded, and provided visual feedback based on the sensed behavioral signals.

Our experimental evaluations aimed to clarify the following two research hypotheses: (H_1) the virtual agents' ToM levels have effects on their acceptability and trust as trainers, and (H_2) distrust was weakened by the SST process. This study's purpose was to use the system to explore the effect of trustworthiness, acceptability, familiarity, and likeability on the agents' ToM and the SST procedure. 29 participants rated the stage of appearance by watching a video with ToM levels and SST. Their subjective ratings of the system's acceptability, trustworthiness, familiarity, and likeability were rated using measures at three stages. Prior to the experiment, we also obtained assessments for the participants' autistic traits, anxiety, and general social skills.

Our experiment results confirmed that the users' trust and acceptability significantly changed depending on the level of the virtual agent's theory of mind. In particular, the acceptability and trustworthiness of an agent with a low-level ToM significantly decreased. But we found significant improvement in the acceptability, trustworthiness, and likeability after SST by observing the differences between the ToM video and SST stages. We also found that higher communication skills (as measured by SRS-2) tended to rate the virtual agent as familiar. These results supported the validity of I_1 and I_2: the effect of the ToM level on the acceptability and reliability of the trainers, and distrust was weakened by the SST process.

One limitation of this study is that the SST and ToM videos might be perceived as independent by the participants. In addition, we did not examine which SST factors improved the subjective ratings. Incorporating ToM features into virtual agents during SST [26], [38] is a possible step to examine the direct effect of ToM levels on the SST. Such possible integration is a future challenge. We did not show that our results could be directly generalized to the other agent identities. We did not confirm that participants perceived that the agent had a high or low theory of mind. Although the two ToM scenarios differed in only one utterance, the reported effects might be related to some other feature or perception of the participants.

Based on the findings of this study, in the future, we believe each module of an SST system must be improved, including ToM, and SST must be designed to enhance the functionality of other modules for building alliances with users. For example, playing games or engaging in small talk before starting the SSTs [25] is one critical step. In addition, since people with social difficulties struggle to understand ToM, incorporating such social cognition training into SST as role-playing is crucial [24].

TABLE III

SPEARMAN'S CORRELATIONS BETWEEN MEASURES AND ASSESSMENTS AT APPEARANCE STAGE *: $p < 0.05$ IN COMPARISON WITHOUT CORRELATIONS.

	State	Trait	SRS-Total	Aware.	Cog.	Com.	Mot.	Res.	KISS-18
Accept.	-0.05	-0.12	-0.25	-0.35 *	-0.2	-0.24	-0.08	-0.37	0.16
Trust.	-0.07	-0.17	-0.28	-0.22	-0.3	-0.28	-0.06	-0.41 *	0.23
Famil.	-0.16	-0.26	-0.41 *	-0.26	-0.3	-0.48 *	-0.33	-0.34	0.46 *
Like.	-0.18	-0.36 *	-0.29	-0.27 *	-0.34	-0.29 *	-0.2	-0.12	0.31

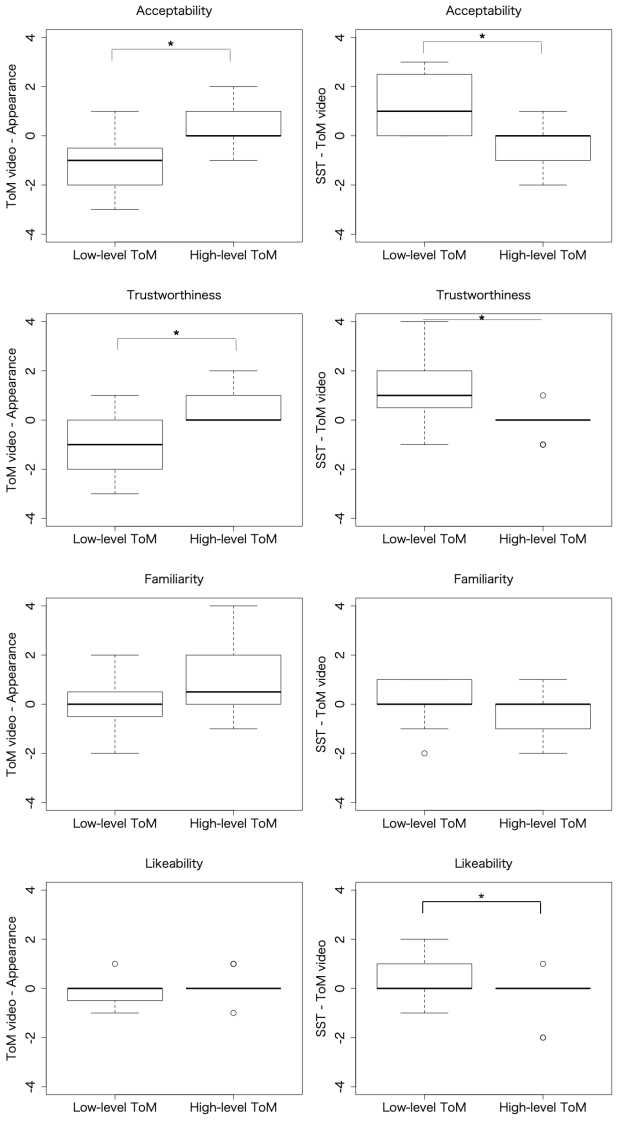

Fig. 6. Boxplots with regard to ratings of acceptability, trustworthiness, familiarity, and likeability: Left-hand side explains subtraction of 5-point scores in the appearance stage from those in the ToM video stage. Right-hand side explains subtraction of 5-point scores in the ToM video stage from those in the SST stage. *: $p < 0.05$ by the Wilcoxon ranks sum test.

REFERENCES

[1] R. B. Ardito and D. Rabellino. Therapeutic alliance and outcome of psychotherapy: historical excursus, measurements, and prospects for research. *Frontiers in psychology*, 2:270, 2011.

[2] T. Baltrušaitis, P. Robinson, and L.-P. Morency. Openface: an open source facial behavior analysis toolkit. In *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1–10. IEEE, 2016.

[3] S. Baron-Cohen. Theory of mind and autism: A review. *International review of research in mental retardation*, 23:169–184, 2000.

[4] S. Baron-Cohen, J. Richler, D. Bisarya, N. Gurunathan, and S. Wheelwright. The systemizing quotient: an investigation of adults with asperger syndrome or high–functioning autism, and normal sex differences. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 358(1430):361–374, 2003.

[5] N. Bauminger-Zviely. *False-Belief Task*, pages 1249–1249. Springer New York, New York, NY, 2013.

[6] A. S. Bellack, K. T. Mueser, S. Gingerich, and J. Agresta. *Social skills training for schizophrenia: A step-by-step guide*. Guilford Publications, 2013.

[7] J. Bishop. The internet for educating individuals with social impairments. *Journal of Computer Assisted Learning*, 19(4):546–556, 2003.

[8] P. Boersma. Praat: doing phonetics by computer. 2022.

[9] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh. Realtime multi-person 2d pose estimation using part affinity fields. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7291–7299, 2017.

[10] J. N. Constantino and C. P. Gruber. *Social responsiveness scale: SRS-2*. Western psychological services Torrance, CA, 2012.

[11] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

[12] A. Esposito, T. Amorese, M. Cuciniello, I. Pica, M. T. Riviello, A. Troncone, G. Cordasco, and A. M. Esposito. Elders prefer female robots with a high degree of human likeness. In *2019 IEEE 23rd International Symposium on Consumer Technologies (ISCT)*, pages 243–246. IEEE, 2019.

[13] A. Esposito, T. Amorese, M. Cuciniello, M. T. Riviello, A. M. Esposito, A. Troncone, and G. Cordasco. The dependability of voice on elders' acceptance of humanoid agents. In *Interspeech*, pages 31–35, 2019.

[14] C. Frith and U. Frith. Theory of mind. *Current biology*, 15(17):R644–R645, 2005.

[15] J. Hamet Bagnou, E. Prigent, J.-C. Martin, J. Woo, L. Yang, C. Achard, C. Pelachaud, and C. Clavel. A framework for the assessment and training of collaborative problem-solving social skills. In *Companion Publication of the 2021 International Conference on Multimodal Interaction*, pages 381–384, 2021.

[16] S. Holroyd and S. Baron-Cohen. Brief report: How far can people with autism go in developing a theory of mind? *Journal of Autism and Developmental Disorders*, 23(2):379–385, 1993.

[17] K. Kikuchi. Shakaiteki-sukiru-wo-hakaru-kiss-18 handbook. *The social skills are measured, The handbook of Kiss-18, Tokyo, Japan, Kawashima*, 2007.

[18] A. Kopelowicz, R. P. Liberman, and R. Zarate. Recent advances in social skills training for schizophrenia. *Schizophrenia bulletin*, 32:S12–S23, 2006.

[19] R. P. Liberman, W. J. DeRisi, and K. T. Mueser. *Social skills training for psychiatric patients*. Pergamon Press, 1989.

[20] S. Masahiro, S. Hiroki, H. Masanori, and I. Kiyoshi. Construction and validation of pre-trained language model using financial documents. In *Proceedings of JSAI Special Interest Group on Financial Informatics (SIG-FIN) 27*, oct 2021.

[21] D. Moore, P. McGrath, and J. Thorpe. Computer-aided learning for people with autism–a framework for research and development.

Innovations in education and training international, 37(3):218–228, 2000.

[22] W. Mou, M. Ruocco, D. Zanatto, and A. Cangelosi. When would you trust a robot? a study on trust and theory of mind in human-robot interactions. In *2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, pages 956–962. IEEE, 2020.

[23] R. Niewiadomski, E. Bevacqua, M. Mancini, and C. Pelachaud. Greta: an interactive expressive eca system. In *Proceedings of The 8th International Conference on Autonomous Agents and Multiagent Systems-Volume 2*, pages 1399–1400. Citeseer, 2009.

[24] S. A. Nijman, W. Veling, K. Greaves-Lord, M. Vos, C. E. R. Zandee, M. A. het Rot, C. N. W. Geraets, and G. H. M. Pijnenborg. Dynamic interactive social cognition training in virtual reality (discovr) for people with a psychotic disorder: single-group feasibility and acceptability study. *JMIR mental health*, 7(8):e17808, 2020.

[25] D. Novick and I. Gris. Building rapport between human and eca: A pilot study. In *International Conference on Human-Computer Interaction*, pages 472–480. Springer, 2014.

[26] N. Rabinowitz, F. Perbet, F. Song, C. Zhang, S. A. Eslami, and M. Botvinick. Machine theory of mind. In *International conference on machine learning*, pages 4218–4227. PMLR, 2018.

[27] T. Saga, H. Tanaka, Y. Matuda, T. Morimoto, M. Uratani, K. Okazaki, Y. Fujimoto, and S. Nakamura. Analysis of feedback contents and estimation of subjective scores in social skills training. *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 2022.

[28] Y. Song and Y. Luximon. Trust in ai agent: A systematic review of facial anthropomorphic trustworthiness for social robot design. *Sensors*, 20(18):5087, 2020.

[29] C. D. Spielberger. State-trait anxiety inventory for adults. 1983.

[30] K. Stendal and S. Balandin. Virtual worlds for people with autism spectrum disorder: a case study in second life. *Disability and rehabilitation*, 37(17):1591–1598, 2015.

[31] J. Swettenham. Can children with autism be taught to understand false belief using computers? *Journal of Child Psychology and psychiatry*, 37(2):157–165, 1996.

[32] H. Tanaka, H. Iwasaka, Y. Matsuda, K. Okazaki, and S. Nakamura. Analyzing self-efficacy and summary feedback in automated social skills training. *IEEE Open Journal of Engineering in Medicine and Biology*, 2:65–70, 2021.

[33] H. Tanaka, H. Iwasaka, H. Negoro, and S. Nakamura. Analysis of conversational listening skills toward agent-based social skills training. *Journal on Multimodal User Interfaces*, 14(1):73–82, 2020.

[34] H. Tanaka, S. Nakamura, et al. The acceptability of virtual characters as social skills trainers: Usability study. *JMIR human factors*, 9(1):e35358, 2022.

[35] H. Tanaka, H. Negoro, H. Iwasaka, and S. Nakamura. Embodied conversational agents for multimodal automated social skills training in people with autism spectrum disorders. *PLOS ONE*, 12(8):1–15, 08 2017.

[36] H. Tanaka, S. Sakriani, G. Neubig, T. Toda, H. Negoro, H. Iwasaka, and S. Nakamura. Teaching social communication skills through human-agent interaction. *ACM Trans. Interact. Intell. Syst.*, 6(2), aug 2016.

[37] K. Terada, L. Jing, and S. Yamada. Effects of agent appearance on customer buying motivations on online shopping sites. In *Proceedings of the 33rd annual acm conference extended abstracts on human factors in computing systems*, pages 929–934, 2015.

[38] S. Vinanzi, M. Patacchiola, A. Chella, and A. Cangelosi. Would a robot trust you? developmental robotics model of trust and theory of mind. *Philosophical Transactions of the Royal Society B*, 374(1771):20180032, 2019.

[39] Z.-T. Yeh and C.-F. Tsai. Impairment on theory of mind and empathy in patients with stroke. *Psychiatry and clinical neurosciences*, 68(8):612–620, 2014.