Does What Users Say Match What They Do? Comparing Self-Reported Attitudes and Behaviours towards a Social Robot

Rebecca Stower, Karen Tatarian, Damien Rudaz, Marine Chamoux, Mohamed Chetouani, Arvid Kappas

Does what users say match what they do? Comparing self-reported attitudes and behaviours towards a social robot

Rebecca Stower[1], Karen Tatarian[2], Damien Rudaz[2,3], Marine Chamoux[2], Mohamed Chetouani[4] and Arvid Kappas[5]

Abstract— **Constructs intended to capture social attitudes and behaviour towards social robots are incredibly varied, with little overlap or consistency in how they may be related. In this study we conduct an exploratory analysis between participants self-reported attitudes and behaviour towards a social robot. We designed an autonomous interaction where 102 participants interacted with a social robot (Pepper) in a hypothetical travel planning scenario, during which the robot displayed various multi-modal social behaviours (gaze, gesture, proxemics, and social dialogue). Several behavioural measures were embedded throughout the interaction, followed by a self-report questionnaire targeting participant's social attitudes towards the robot (social trust, liking, rapport, competency trust, technology acceptance, mind perception, social presence, and social information processing). Several relationships were identified between participant's behaviour and self-reported attitudes towards the robot. Implications for how to conceptualise and measure interactions with social robots are discussed.**

I. Introduction

Among human-robot-interaction (HRI) research, there have been varied attempts at capturing people's social attitudes towards robots, e.g., the Godspeed Questionnaire [1], Almere Questionnaire [2], The Robot Social Attributes Scale [3] and the Negative Attitudes Towards Robots Scale (NARS) [4]. These questionnaires cover concepts such as anthropomorphism, animacy, liking, intelligence, anxiety, acceptance, social presence, and trust. More recently, a systematic review conducted by [5] aimed to assess attitudes, anxiety, acceptance, and trust towards social robots and found that although attitudes towards robots were generally positive, the overall effect size was quite small and with high heterogeneity, reflective of inconsistent effects reported throughout the field.

A Perceived Social Intelligence Scale for Robots [6] has also been proposed, with two main dimensions; social competence and social presentation. However, this scale has so far only been validated for video evaluations of social robots, rather than real-life interactions. In addition, how such perceptions relate to peoples actual behaviour towards the robot is yet to be explored. Consequently, there is currently limited understanding of how participants attitudes and behaviours towards robots interact, especially in real-life, autonomous interactions.

Inspired by human-human interaction, many behavioural models have also been implemented with the goal of making human-robot interaction appear as 'natural' as possible. These include models of gaze behaviour (drawing inspiration from [7]), proxemics [8], gesture/kinesics [9] and social or interpersonal dialogue (e.g., politeness theory [10], social penetration theory [11]). However, the translation of human(like) social behaviours to robotic agents is not always straightforward. Robots can embody a variety of different physical forms, many of which make it physically challenging to mimic human behaviours (e.g., facial expressions) [12]. Conversely, in some cases robotic embodiment goes beyond human-like behaviours, for example using LEDs in eyes to communicate emotion [13]. The same behaviour in different robot embodiments can also lead to differing effects (e.g., [14]). Consequently, understanding how social behaviours of robots are interpreted by users is often difficult, leading to a large variety of implementations and measurements.

There is also ongoing debate within the field of HRI about the use of Wizard of Oz (WoZ) paradigms, where a teleoperator controls the behaviour of the robot, versus fully autonomous interactions. Although WoZ can be useful to study robot behaviours and interactions which may currently be technically out of reach, it has been criticised as potentially acting only as a simulation of human-human interaction.

Thus, although there are increasing studies which focus on autonomous interactions with robots (e.g., [15], [16], [17]), there is still work to be done towards understanding of participants attitudes and behaviour towards genuinely autonomous social robots. In addition, there is a need to include multiple assessment measures to fully capture how users attitudes towards robots translate into actual behaviours. To address this gap, we designed an interaction with an autonomous social robot displaying various multi-modal social behaviours and captured participants self-reported attitudes and behaviours towards an autonomous social robot (Pepper, SoftBank Robotics).

II. The Current Study

In this study, we aimed to analyse how participants self-reported attitudes towards a social robot correspond with

*This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 765955

[1]Rebecca Stower is with the Department of Psychology, CHArt-LUTIN laboratory, Paris 8 University, Paris, France rebecca.stower@univ-paris8.fr

[2]Karen Tatarian is with SoftBank Robotics Europe, Paris, France

[2,3]Damien Rudaz is with SoftBank Robotics Europe and l'Institut Polytechnique de Paris, Paris, France

[2]Marine Chamoux is with SoftBank Robotics Europe, Paris, France

[4]Mohamed Chetouani is with Institut Systèmes Intelligents et de Robotique, CNRS 7222, Sorbonne Université, Paris, France

[5]Arvid Kappas is with the Department of Psychology and Methods, Jacobs University, Bremen, Germany

their actual behaviour during the interaction. Participants interacted with Pepper in a hypothetical travel planning scenario. During the interaction, the Pepper robot operated autonomously to display different combinations of multi-modal social behaviours (gaze, gesture, proxemics, and social dialogue). For a full description of the design and evaluation of the multi-modal behaviours see [18].

III. Method

A. Scenario

The interaction was structured in the form of a travel planning scenario for a hypothetical holiday within Europe, with Pepper acting as a travel agent (See Figure 1). In order to minimise voice recognition issues, the interaction was structured using both verbal cues and interacting with Pepper's chest tablet. Pepper began the interaction by greeting the participant, introducing itself, and offering water placed on a nearby table. Participants had the option to either accept or refuse the water before proceeding with the interaction. Pepper then asked a series of dichotomous questions about participants travel preferences. Participants could indicate their choice by selecting the corresponding option on Pepper's tablet. Pepper then asked the participant to describe more about themselves (self-disclosure) with the explanation that it was to personalise the travel recommendations. For each participant, Pepper generated two different travel locations (e.g., if the participant said they preferred cities and travelling by train, Pepper would suggest London or Amsterdam), and randomly expressed a preference for one. Participants could then either endorse or reject the recommendation by Pepper. The interaction was designed to last between 5-10 minutes. The full script for the interaction is available on GitHub.[1]

Fig. 1. Example of participant interacting with Pepper (picture taken with permission)

ADD DIAGRAM WITH PHASES OF INTERACTION?

[1]Multi-modal Social Cues System Implementation GitHub Repository; url redacted for peer review

B. Participants

Participants were recruited from the INSEAD-Sorbonne Université Behavioural Lab. Ethics approval was obtained from the INSEAD Institutional Review Board. The experiment took approximately 20 minutes and participants were paid €6. Upon entering the study, participants were explained the purpose of the study and asked to indicate their consent to participate and to the use of their video data. 115 participants completed the experiment, of which 13 were excluded due to technical issues with the robot, leaving 102 eligible participants. All participants were French-Speaking.

C. Design

Participants interactions with the robot were measured in two forms; behavioural data either extracted from the robot logs or coded from the videos, and self-report questionnaires. A full description of the coding scheme, variables, questionnaires, and data is available online.[2]

D. Measures

1) Behavioural Data: Using the logs recorded by Pepper, we extracted a number of behavioural outcomes. All other behavioural outcomes were coded based on the video data. Outcomes were either continuous, categorical, or binary, see Table I .

2) Self-Report Questionnaires: We measured participants perceptions of the robot using a series of self-report scales targeting perceived agency, social trust, competency trust, liking, rapport, acceptance, social presence, and social information processing. We modified some items to better fit our study context and to be consistent between scales (e.g., replacing "person" or "system" with"Pepper"). All items were also translated into French. The final questionnaire consisted of 86 items, the order of which was randomised for each participant. We modified all answer scales to a 7-point Likert scale ranging from 1 - *Strongly Disagree* to 7 - *Strongly Agree*.

a) Mind Perception: Perceived agency was measured using a modified version of 18-item Dimensions of Mind Perception Scale [19], with good reliability, $\alpha = 0.89$.

b) Social and Competency Trust: Competency Trust was measured using the Perceived Reliability, Perceived Technical Competence, and Perceived Understandability subscales from [20]. The overall reliability for the combined 3 subscales was $\alpha = 0.86$. Social Trust was measured using a combination of items from [21], [22], $\alpha = 0.75$.

c) Liking/Rapport: Liking and Rapport were measured using modified versions of both the rapport scale from [23], $\alpha = 0.83$ and the Reysen Likeability Scale [24], $\alpha = 0.88$.

d) Technology Acceptance: For acceptance, we used the 'Perceived Ease of Use'and 'Intention to Use' subscales of the Technology Acceptance Model [25], [26]. We excluded the 'Perceived Usefulness' subscale as it did not apply to the current interaction. The combined reliability for both subscales was $\alpha = 0.71$.

[2]url redacted for peer review

TABLE I
BEHAVIOURAL OUTCOMES

	Continuous	Categorical	Binary
Robot Logs	• Distance: The distance participants maintained from Pepper during the interaction • Self-Disclosure: How long participants spoke for during the self-disclosure sequence • Time-to-Recommendation: How long it took participants (in seconds) to decide on a travel destination		• Follow Recommendation: Whether participants followed the recommendation of the robot for their chosen travel destination • Talk-and-Tablet: Whether participants used both the tablet and voice command to interact with Pepper
Video Coding	• Number of voice recognition errors	• Mode of Address towards Pepper: formal ('tu'), informal ('vous'), using Pepper's name ('Pepper'), or no mode of address ('nothing') • Consistency of Social Behaviour: Both social greeting and social closing behaviours, neither social greeting nor closing, social greeting but non-social closing, or non-social greeting but social closing	• Accept Water: Whether participants accepted or rejected the water offered • Social Behaviour: opening wave, closing wave, back-channeling

e) Social Presence: Social presence was measured using the scale proposed by [27]. This scale is comprised of two sub-scales; self-reported co-presence, and perceived-other co-presence, $\alpha = 0.79$.

f) Social Information Processing: Finally, Social Information Processing was measured using the 5-item social information processing sub-scale from the Tromsø Social Intelligence Scale [28], as well as the 'Identifies Humans' and 'Social Competence' sub-scales from the Perceived Social Intelligence (PSI) Scales Test Manual [6]. The combined reliability for the three subscales was $\alpha = 0.76$.

E. Procedure

After filling out the required consent forms, participants were informed they would have a short interaction with a robot acting as a travel agent to plan a hypothetical holiday, after which they would be asked to complete some questionnaires assessing their perception of the robot. They were encouraged to answer the questions from the robot as if they were planning a real holiday. We also explained that they could interact with Pepper both by voice and by using the tablet on Pepper's chest. To aid voice recognition, participants were advised to speak loudly and clearly, and that if they did not receive a response or saw a question mark on the robot's tablet, to repeat their answer again.

Participants were then led to a room with the robot, and instructed to stand in front of the robot. The experimenter then stated that the interaction would begin shortly and left the room. The interaction with Pepper then began as described in section III-A. Two cameras were positioned in the room, one from a front angle, and one from a side angle. A webcam was also set up to allow the experimenters to view the progress of the interaction from a control room.

At the end of the interaction with Pepper, participants were led to a separate room where they completed the questionnaires. Afterwards, participants were debriefed and given their compensation for participating.

IV. RESULTS

Although in the original study participants saw one of 5 different behavioural profiles (see [18]), the analyses reported here focus only on the relationship between participant's self-reported attitudes and behaviours, and as such are collapsed over conditions. As these analyses were exploratory, we did not formulate any *a-priori* hypotheses regarding the association between participant's attitudes and behaviour. We also chose a less conservative approach to post-hoc testing [29].

A. Self-Report Questionnaires

First, we created a rank-ordered correlation matrix using Spearman's ρ between all self-report questionnaires and the continuous behavioural measures (see Table IV-B). There are significant moderate correlations between all self-report variables, indicating these tap into separate but related constructs. In addition, there are significant positive correlations between number of disclosures and rapport, and speaking time and social presence. Between the behavioural variables, there is a significant positive correlation between total speaking time during the self-disclosure phase and number of pieces of information disclosed.

B. Dichotomous Behavioural Outcomes

Next, we counted the number of participants who displayed each dichotomous social behaviour (accept water, follow recommendation, opening and closing waves, talk-and-tablet, and back-channels), see Figure 2. We then calculated Phi correlation coefficients between them.

There was a medium positive association between producing opening and closing waves ($\phi = 0.53$) and weak positive associations between back-channels and opening

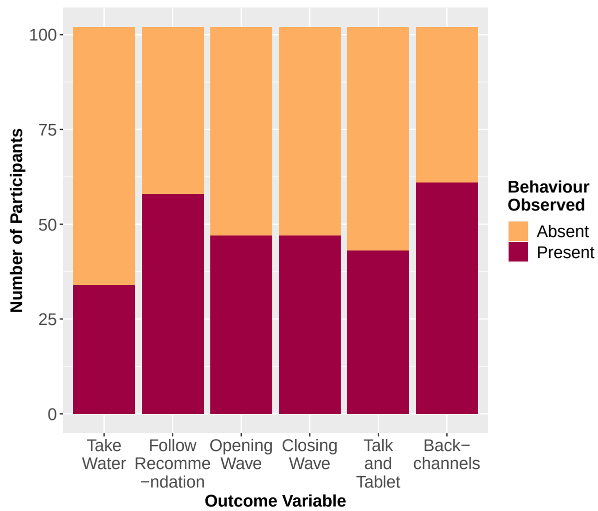

Fig. 2. Number of participants who displayed each social behaviour

wave ($\phi = 0.20$), and back-channels and talking and using the tablet ($\phi = 0.37$)

We then used Mann-Whitney U tests to compare self-reported responses on the questionnaires between participants where a behaviour was present vs absent (see Table III). Participants who took the water offered by Pepper reported significantly higher social trust, social presence and social information processing than participants who did not accept the water. Participants who produced an opening wave reported significantly higher technology acceptance, liking, rapport, and social presence than those who did not wave. Similarly, participants who produced a closing wave reported significantly higher perceived agency, social trust, technology acceptance, liking, rapport and social presence. They also maintained a significantly closer distance to Pepper throughout the interaction. Participants who produced back-channels throughout the interaction reported significantly higher technology acceptance than those who did not. They also spoke for a longer time and disclosed more pieces of information. Participants who used both the tablet and continued talking to Pepper throughout the interaction maintained a closer distance to Pepper than those who used only the tablet.

C. Categorical Behavioural Outcomes

For categorical behavioural variables with >2 outcomes, Kruskal-Wallis H-Tests were conducted to compare differences between groups on self-reported attitudes towards Pepper. There was a significant effect of the consistency of participants behaviour on social presence, $\chi^2(3) = 9.53$, $p = 0.02$, $\epsilon^2 = 0.09$, and total speaking time $\chi^2(3) = 12.09$, $p = 0.007$, $\epsilon^2 = 0.11$. Follow up tests revealed participants who were consistently social towards the robot spoke significantly longer than participants who were consistently non-social, $W = 4.09$, $p = .02$. Participants who were consistently social reported marginally significantly higher ratings of social presence than participants who became less social throughout the interaction, $W = 1.88$, $p = .07$.

D. Voice Recognition Errors

Although we attempted to limit the amount of autonomous voice recognition, the introductory phase included some reciprocal interaction between Pepper and the participant (e.g., asking ”how are you”). Thus, there was still some potential for voice recognition errors to occur. To explore if these (naturally occurring) errors had any effect on participants attitudes or behaviour we calculated spearman's ρ correlations between the number of voice recognition errors and all other outcomes. Significant negative correlations were identified between the number of voice recognition errors, technology acceptance $\rho = -0.23$, $p < .05$, the total time the participant spent talking during the self-disclosure phase $\rho = -0.20$, $p < .05$ and the number of pieces of information they disclosed $\rho = -0.27$, $p < .01$.

V. Discussion

The goal of this research was to explore how self-report and behavioural measures of social intelligence relate to each other. Through evaluating participants interactions with an autonomous social robot, we aimed to develop a deeper understanding of if and how these attitudes relate to participants actual behaviour towards the robot.

Several patterns were identified between behavioural and self-report outcomes. Social presence was one of the most commonly affected outcomes, with participants who took the water offered from Pepper and produced an opening wave or closing wave all reporting higher ratings of social presence than participants who did not display those behaviours. These behaviours all involve interacting in the same physical space as Pepper, as opposed to the other behavioural variables such as following the recommendation from Pepper or producing back-channels, which rely more on verbal interactivity. As such, it is possible that greater physical interaction within the space leads to higher ratings of social presence.

Technology acceptance was also higher in participants who produced an opening wave or closing wave than those who did not, as well as those who produced back-channels. Potentially, these participants were more predisposed to accept the interaction with Pepper, which subsequently influenced their behaviour throughout the interaction. In order to test this hypothesis, pre and post tests regarding participants acceptance of robots before and after an interaction, and how this relates to behaviour towards robots would be useful.

Liking and rapport were both higher among participants who produced opening or closing waves. This finding is not particularly surprising, as findings in psychology show that interpersonal synchrony (i.e. mimicking behaviour such as waving) is associated with higher rapport [30], [31]. As such, participants which liked the robot more (either because of the robots own behaviour, or some other factor) may have expressed this affiliation through imitating the robots waves. Rapport was also positively related to the number

TABLE II

SPEARMAN'S ρ RANK-ORDERED CORRELATIONS BETWEEN SELF-REPORT ITEMS AND CONTINUOUS BEHAVIOURAL MEASURES

Measure	Agency	Social Trust	Competency Trust	TAM	Liking	Rapport	Social Presence	Social Intelligence	Time to Recommen -dation	Time to Button	Speaking Time	Number of Disclosures	Distance Maintained
Self-Report													
Agency	1.00												
Social Trust	0.32**	1.00											
Competency Trust	0.29**	0.53****	1.00										
TAM	0.33***	0.51***	0.51***	1.00									
Liking	0.47***	0.67***	0.6***	0.71***	1.00								
Rapport	0.44***	0.52***	0.55***	0.66***	0.72***	1.00							
Social Presence	0.42***	0.68***	0.43***	0.67***	0.73***	0.75***	1.00						
Social Information Processing	0.58***	0.41***	0.54***	0.39***	0.6***	0.49***	0.42***	1.00					
Behavioural													
Time to Recommen -dation	-0.045	0.031	-0.012	-0.0042	-0.02	-0.0074	-0.031	0.071	1.00				
Time to Button	-0.036	-0.05	-0.13	0.03	-0.12	0.024	0.087	-0.078	-0.0053	1.00			
Speaking Time	0.011	0.089	0.055	0.087	0.029	0.14	0.23*	0.046	0.11	0.05	1.00		
Number of Disclosures	0.12	0.052	0.07	0.17	0.094	0.24*	0.17	0.17	0.18	0.055	0.67***	1.00	
Distance Maintained	-0.0053	-0.013	-0.14	-0.069	0.0084	-0.071	-0.11	0.11	-0.067	0.052	-0.052	-0.038	1.00

* Significant at the $p < .05$ level, ** Significant at the $p < .01$ level, *** Significant at the $p < .001$ level

TABLE III

MANN-WHITNEY U COMPARISONS FOR THE EFFECT OF DICHOTOMOUS VARIABLES ON CONTINUOUS OUTCOMES

Measure	Take Water			Follow Recommendation			Opening Wave			Closing Wave			Back-Channels			Talk and use Tablet		
	U	*p*	*d*	*U*	*p*	*d*	*U*	*p*	*d*	*U*	*p*	*d*	*U*	*p*	*d*	*U*	*p*	*d*
Self-Report																		
Agency	967.5	.182	-0.27	1210.0	.658	-0.08	1119.0	.245	-0.25	936.5	.016*	-0.47	1184.5	.655	-0.10	1193.5	.614	-0.06
Social Trust	658.0	.000***	-0.60	1215.5	.685	0.03	1037.5	.087	-0.28	945.5	.020*	-0.42	984.0	.068	-0.34	1132.0	.355	-0.08
Competency Trust	987.5	.233	-0.33	1220.5	.710	-0.07	1149.5	.338	-0.18	1153.0	.350	-1.19	1138.5	.446	-0.12	1059.0	.200	0.25
TAM	921.5	.096	-0.36	1272.5	.984	0.02	919.5	.012*	-0.45	756.5	.000***	-0.64	880.0	.011*	-0.54	1131.0	.352	-0.14
Liking	923.0	.099	-0.35	1168.5	.469	0.06	925.0	.014*	-0.48	787.0	.001**	-0.69	1046.0	.163	-0.31	1241.5	.857	0.00
Rapport	1033.0	.384	-0.17	1154.0	.413	-0.13	931.5	.015*	-0.51	901.5	.009**	-0.52	1101.0	.309	-0.21	1227.0	.781	0.06
Social Presence	823.0	.018*	-0.59	1237.0	.795	0.03	903.0	.009**	-0.61	704.5	.000***	-0.86	971.5	.057	-0.384	1054.0	.147	-0.26
Social Information Processing	850.5	.030**	-0.43	1253.5	.882	-0.01	1163.0	.386	-0.24	1125.0	.262	-0.17	1137.0	.440	-0.13	1133.00	.360	0.16
Behavioural																		
Time to Recommendation	922.0	.097	-0.47	1077.0	.180	-0.12	1059.0	.118	0.006	1223.0	.643	-0.06	1144.0	.469	-0.03	1182.0	.560	0.02
Time to Button	1131.5	.865	-0.15	1118.5	.289	0.22	1207.5	.571	-0.08	1148.5	.335	-0.37	1134.5	.431	-0.04	991.5	.061	-0.52
Speaking Time	1034.0	.388	-0.18	1104.5	.248	-0.14	1046.5	.099	-0.25	1106.5	.213	-0.17	962.5	.050*	-0.41	1080.5	.204	-0.21
Number of Disclosures	1132.5	.868	0.00	1104.5	.238	-0.20	1142.0	.304	-0.25	1197.5	.517	-0.06	882.0	.010*	-0.51	1181.5	.550	-0.11
Distance Maintained	1119.0	.796	0.22	1273.0	.238	-0.20	1227.0	.663	0.11	958	.049*	0.26	1249.0	.995	-0.01	780.0	.067	0.47

* Significant at the $p < .05$ level, ** Significant at the $p < .01$ level, *** Significant at the $p < .001$ level

of disclosures made, which aligns with social penetration theory.

Social trust, social intelligence, and perceived agency were the only other self-report variables to show relationships with behavioural outcomes (taking the water offered and/or opening or closing waves). Interestingly, these variables mainly relate to the more social or "warmth" dimension of social cognition [32]; perceived competence was not affected at all by any of the behaviours. This is potentially explained by the more social context of the task, as well as the fact that participants (mostly) had no reason to doubt Pepper's competency in the interaction.

That being said, the structure of the interaction also allowed for investigation into how naturally-occurring errors, such as those which are likely to eventuate in real-life interactions, relate to participants attitudes and behaviours. Findings from these analyses indicate voice recognition errors are related to the length of time people spoke to Pepper for, the number of pieces of information they were willing to disclose, and their technology acceptance ratings.

This suggests that voice recognition errors are related to participants attitudes and behaviours towards Pepper.

However, since the voice recognition errors occurred naturally, rather than being systematically manipulated, it is difficult to assess the direction of these relationships. It is possible that some third, unrelated factor could lead to both voice recognition errors being more frequent and to less willingness to engage from the participants (e.g., participants which are less comfortable with the robot may mumble or speak less clearly, leading to more voice recognition errors). This is again something which could be addressed through the use of a pre-post design. It would also be useful to understand whether in some circumstances errors of the robot are perceived more negatively than others (i.e., if a user already has a negative impression of the robot, errors may have a larger detrimental effect than for users which have a generally positive impression).

VI. Conclusion

We report the results from a study targeted at understanding the relationship between self-report measures of social attitudes towards robots, and participants own behaviour towards the robot. Significant positive relationships were identified between the self-report measures, indicating multiple constructs could interact to form a general impression of the robot. In addition, several relationships were identified between behavioural and self-report measures of social intelligence, indicating participants behaviour towards a robot may reflect perceptions of the robot's social intelligence.

In sum, this study suggests that participants attitudes towards a robot can be related to how they behave during an interaction with an autonomous social robot. Several potentially interesting directions for future research studies were identified which would benefit from the continued development of such research.

ACKNOWLEDGMENT

We would like to thank Hugues Pellerin for advice regarding the statistical analyses for the paper. Also thanks to the team at the INSEAD behavioural lab for their help with data collection.

References

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

[2] M. Heerink, B. Kröse, V. Evers, and B. Wielinga, "Assessing Acceptance of Assistive Social Agent Technology by Older Adults: the Almere Model," *International Journal of Social Robotics*, vol. 2, no. 4, pp. 361–375, dec 2010.

[3] C. M. Carpinella, A. B. Wyman, M. A. Perez, and S. J. Stroessner, "The Robotic Social Attributes Scale (RoSAS): Development and Validation," *2017 12th ACM/IEEE International Conference on Human-Robot Interaction (HRI*, pp. 254–262, 2017.

[4] T. Nomura, T. Kanda, and T. Suzuki, "Experimental investigation into influence of negative attitudes toward robots on human–robot interaction," *AI & SOCIETY*, vol. 20, no. 2, pp. 138–150, mar 2006.

[5] S. Naneva, M. Sarda Gou, T. L. Webb, and T. J. Prescott, "A Systematic Review of Attitudes, Anxiety, Acceptance, and Trust Towards Social Robots," *International Journal of Social Robotics*, jun 2020.

[6] K. A. Barchard, L. Lapping-Carr, R. S. Westfall, A. Fink-Armold, S. B. Banisetty, and D. Feil-Seifer, "Measuring the Perceived Social Intelligence of Robots," *ACM Transactions on Human-Robot Interaction*, vol. 9, no. 4, pp. 1–29, oct 2020.

[7] A. Kendon, "Some functions of gaze-direction in social interaction." *Acta Psychologica*, vol. 26, no. 1, pp. 22–63, 1967.

[8] E. T. Hall, R. L. Birdwhistell, B. Bock, P. Bohannan, A. R. Diebold, M. Durbin, M. S. Edmonson, J. L. Fischer, D. Hymes, S. T. Kimball, W. La Barre, J. E. McClellan, D. S. Marshall, G. B. Milner, H. B. Sarles, G. L. Trager, and A. P. Vayda, "Proxemics [and Comments and Replies]," *Current Anthropology*, vol. 9, no. 2/3, pp. 83–108, apr 1968.

[9] D. McNeill, *Hand and Mind: What Gestures Reveal about Thought*. Chicago: University of Chicago Press, 1992.

[10] S. C. Levinson, P. Brown, S. C. Levinson, and S. C. Levinson, *Politeness: Some universals in language usage*. Cambridge university press Cambridge, 1987, vol. 4.

[11] I. Altman and D. A. Taylor, *Social penetration: The development of interpersonal relationships.* Holt, Rinehart & Winston, 1973.

[12] S. Andrist, X. Z. Tan, M. Gleicher, and B. Mutlu, "Conversational gaze aversion for humanlike robots," in *ACM/IEEE International Conference on Human-Robot Interaction*. ACM Press, 2014, pp. 25–32.

[13] M. Westhoven, T. van der Grinten, and S. Mueller, "Perceptions of a Help-Requesting Robot - Effects of Eye-Expressions, Colored Lights and Politeness of Speech," in *Proceedings of Mensch und Computer 2019 on - MuC'19*. New York, New York, USA: ACM Press, 2019, pp. 43–54.

[14] D. Zanatto, M. Patacchiola, J. Goslin, and A. Cangelosi, "Priming Anthropomorphism: Can the credibility of humanlike robots be transferred to non-humanlike robots?" in *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, vol. 2016-April. IEEE, mar 2016, pp. 543–544.

[15] S. Andrist, B. Mutlu, and A. Tapus, "Look like me: Matching robot personality via gaze to increase motivation," in *Conference on Human Factors in Computing Systems - Proceedings*, vol. 2015-April, no. 1. New York, New York, USA: ACM Press, 2015, pp. 3603–3612.

[16] R. Mead and M. J. Matarić, "Autonomous human–robot proxemics: socially aware navigation based on interaction potential," *Autonomous Robots*, vol. 41, no. 5, pp. 1189–1201, 2017.

[17] K. Tatarian, M. Chamoux, A. K. Pandey, and M. Chetouani, "Robot Gaze Behavior and Proxemics to Coordinate Conversational Roles in Group Interactions," in *2021 30th IEEE International Conference on Robot Human Interactive Communication (RO-MAN)*, 2021, pp. 1297–1304.

[18] K. Tatarian, R. Stower, D. Rudaz, M. Chamoux, A. Kappas, and M. Chetouani, "How does Modality Matter ? Investigating the Synthesis and Effects of Multi-modal Robot Behavior on Social Intelligence," *International Journal of Social Robotics*, 2021.

[19] H. M. Gray, K. Gray, and D. M. Wegner, "Dimensions of Mind Perception," *Science*, vol. 315, no. 5812, pp. 619–619, feb 2007.

[20] M. Madsen and S. Gregor, "Measuring Human-Computer Trust," *Proceedings of Eleventh Australasian Conference on Information Systems*, pp. 6–8, 2000.

[21] B. J. Fogg and H. Tseng, "The elements of computer credibility," *Conference on Human Factors in Computing Systems - Proceedings*, no. May, pp. 80–87, 1999.

[22] K. E. Schaefer, "Measuring Trust in Human Robot Interactions: Development of the "Trust Perception Scale-HRI"," in *Robust Intelligence and Trust in Autonomous Systems*, R. Mittu, D. Sofge, A. Wagner, and W. F. Lawless, Eds. Boston, MA: Springer US, 2016, no. April 2016, pp. 191–218.

[23] A. M. von der Pütten, N. C. Krämer, J. Gratch, and S.-H. Kang, ""It doesn't matter what you are!" Explaining social effects of agents and avatars," *Computers in Human Behavior*, vol. 26, no. 6, pp. 1641–1650, nov 2010.

[24] S. Reysen, "Construction of a new scale: The Reysen likability scale," *Social Behavior and Personality: an international journal*, vol. 33, no. 2, pp. 201–208, jan 2005.

[25] F. D. Davis, "Perceived usefulness, perceived ease of use, and user acceptance of information technology," *MIS quarterly*, pp. 319–340, 1989.

[26] V. Venkatesh, M. G. Morris, G. B. Davis, and F. D. Davis, "User acceptance of information technology: Toward a unified view," *MIS quarterly*, pp. 425–478, 2003.

[27] F. Biocca, C. Harms, and J. K. Burgoon, “Toward a More Robust Theory and Measure of Social Presence: Review and Suggested Criteria,” *Presence: Teleoperators and Virtual Environments*, vol. 12, no. 5, pp. 456–480, 2003.
[28] D. Silvera, M. Martinussen, and T. I. Dahl, “The TromsøSocial Intelligence Scale, a self-report measure of social intelligence,” *Scandinavian Journal of Psychology*, vol. 42, no. 4, pp. 313–319, sep 2001.
[29] R. Bender and S. Lange, “Adjusting for multiple testing—when and how?” *Journal of Clinical Epidemiology*, vol. 54, no. 4, pp. 343–349, apr 2001.
[30] J. E. Grahe and F. J. Bernieri, “The importance of nonverbal cues in judging rapport,” *Journal of Nonverbal Behavior*, vol. 23, no. 4, pp. 253–269, 1999.
[31] E. Delaherche, M. Chetouani, A. Mahdhaoui, C. Saint-Georges, S. Viaux, and D. Cohen, “Interpersonal Synchrony: A Survey of Evaluation Methods across Disciplines,” *IEEE Transactions on Affective Computing*, vol. 3, no. 3, pp. 349–365, jul 2012.
[32] S. T. Fiske, A. J. C. Cuddy, and P. Glick, “Universal dimensions of social cognition: warmth and competence,” *Trends in Cognitive Sciences*, vol. 11, no. 2, pp. 77–83, feb 2007.