Memory-Based Personalization for Fostering a Long-Term Child-Robot Relationship

Ligthart, Mike E.U.; Neerincx, Mark A.; Hindriks, Koen V.

DOI
10.1109/HRI53351.2022.9889446

Publication date
2022

Document Version
Final published version

Published in
HRI 2022 - Proceedings of the 2022 ACM/IEEE International Conference on Human-Robot Interaction

Memory-Based Personalization for Fostering a Long-Term Child-Robot Relationship

Mike E.U. Ligthart
Vrije Universiteit Amsterdam
The Netherlands
m.e.u.ligthart@vu.nl
0000-0002-0768-9977

Mark A. Neerincx
Delft University of Technology & TNO Soesterberg
The Netherlands
mark.neerincx@tno.nl
0000-0002-8161-5722

Koen V. Hindriks
Vrije Universiteit Amsterdam
The Netherlands
k.v.hindriks@vu.nl
0000-0002-5707-5236

2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI) | 978-1-6654-0731-1/22/$31.00 ©2022 IEEE | DOI: 10.1109/HRI53351.2022.9889446

***Abstract*—After the novelty effect wears off children need a new motivator to keep interacting with a social robot. Enabling children to build a relationship with the robot is the key for facilitating a sustainable long-term interaction. We designed a memory-based personalization strategy that safeguards the continuity between sessions and tailors the interaction to the child's needs and interests to foster the child-robot relationship. A longitudinal (five sessions in two months) user study (N = 46, 8-10 y.o) showed that the strategy kept children interested longer in the robot, fosters more closeness, elicits more positive social cues, and adds continuity between sessions.**

***Index Terms*—child-robot interaction, social robots, personalization, memory, long-term, longitudinal user study**

I. Introduction

Many social robot applications have the intention to interact on multiple occasions with people over an extended period of time. This is especially true for socially assistive robots. The need for social support does not stop after one interaction, nor is it likely that the often complex goals are achieved. Simply repeating the same interaction every time is not enough, because most children will lose interest after a short while [1], [2].

A more substantial approach is necessary. One that includes adapting the behavior of the robot over time, to not only introduce novel behaviors [2], [3], but to create a common ground between the child and the robot [1]. A two-month field study showed that children who felt they could be friends with the robot kept on interacting, while peers who held a more mechanistical view of the robot lost interest [4]. Fostering a relationship between the child and the robot appears to be a key step in keeping the interaction compelling long-term. The big question is how to foster the child-robot relationship?

A logical step in the right direction is to tailor the interaction to children's interests, preferences, and needs [5]–[8]. Reviewing interpersonal psychology literature we have identified the need for novelty and narrative development, continuity, and similarity and familiarity as important factors for children to develop a relationship with the robot. To address these needs we designed a novel serial narrative dialog structure that enables the robot to autonomously engage in a multi-session conversation. Each session has an overarching narrative that adds continuity, familiarity, and a development to the conversation. Each session has its own narrative arc to add something new every time and to leave room for personalization. A key feature to personalize the conversation is a robot memory [3], [9]–[11].

With this work we contribute by providing a novel memory-based personalization strategy (in sections II and III) that is explicitly designed for fostering the child-robot relationship. It personalizes the narrative conversation on a personal level, rather than on a task level. Another novel contribution is the longitudinal (five episodes in two months) user study (N = 46, 8-10 y.o school children) with which we demonstrate that the strategy fosters the child-robot relationship and is able to keep children interested to interact long-term (in sections IV - VIII).

II. Related Work

To understand how to design robot behaviors and interaction content that facilitate the development of a child-robot relationship we need to understand what children are looking for in a relationship with the robot. We can learn from how children form relationships with people. Not in an attempt to copy those behaviors, but rather to identify needs the children might have. In this section we discuss which needs we identified as the focus for the current study.

How the robot addresses those needs does not need to be the same as how people do it. We do know that using a memory to personalize each new interaction based on the previous interactions is key for addressing the needs long term [3], [6], [7], [9]–[11]. Our goal is not to mimic the function of human memory, but to personalize a conversation using past information. We reviewed different child-robot applications that used this type of memory-based personalization.

A. Children's Needs for a Long-Term Child-Robot Relationship

There are many interpersonal needs at play when children develop their relationships with people [12]. We made a selection based on how suitable memory-based personalization likely is to address them. The needs are summarized into three categories: novelty and narrative development, continuity, and similarity and familiarity.

1) Novelty and narrative development: A straightforward need, which has a clear impact on how viable a longer interaction is, is the need for novelty. When children get a sense of what the robot can and cannot do (i.e. when the novelty wears off) their interest wavers [1], [13], [14]. A direct way to address this problem is to keep introducing novel interaction possibilities and content [3].

It is not only a matter of adding new content to the interaction, the quality of that content is also important for keeping children interested. They want to be gripped by the interaction [15], [16]. Adding a narrative development to the interaction content provides children with opportunities to be invested in the interaction [17]. In other words, by including a story for children to unravel, they will have to come back and interact again to discover how the story progresses.

2) Continuity: Continuity during multiple interactions is important on multiple levels. Firstly, continuity is an important factor for managing the quality of the conversation [18]. Serving conversational content in line with the narrative development or previously discussed topics leads to a more grounded and relevant conversation [19]. This level of continuity is important for children to keep track of the conversation and remain engaged by it [20]. Secondly, continuity is an important social aspect. Children want to be remembered by the robot [21]. They come to expect it as well, when the robot is presented as a companion with whom they can repeatedly interact [22].

3) Similarity and familiarity: People form relationships more easily if they have similar values [23] and attitudes [24]. The more children can relate to the robot and the more similarities (e.g. shared interests) they feel they have with the robot, the easier it is for children to view the robot as a friend [25]. What the robot discloses about itself shapes how similar children perceive the robot.

Besides feeling similar to the robot children also have a need for familiar content and patterns of interaction [26]. Besides enabling children to navigate the conversation more smoothly [18], familiar content allows them to construct a consistent image of the robot's identity and supports forming a relationship [26]. A balance needs to be found between including novel and familiar content.

B. Memory-based Personalization in Human-Robot Interaction

Memory is necessary for an efficient and pleasant conversation [19]. It also has an important social function in a long-term conversation [27] by facilitating a common ground [19] and a shared experience [28] between interlocutors. Both aspects are important for a human-robot relationship as well [11], [29], [30]. Memory is more than the passive storage of semantic information [10] and thus the question is not only what do we need to remember, but also how is the memory going to tailor the interaction to satisfy the child's needs?

We focus on the things the robot can elicit (e.g. like [31]), store, and recall via the conversation. In related work we find that memory systems in child-robot interaction typically store information about the child, the conversation, and the task. For example, information about the child include their name (e.g. [26], [32]), interests (e.g. [31], [33]), and opinions (e.g. [11]). Information about the conversation includes which questions were asked before (e.g. [26]). Task information include performance metrics (e.g. game scores [32] or adherence to a diabetes regimen [26]) and task decisions (e.g. [11]).

From the literature we identified three strategies to personalize the conversation using the information that is stored in the memory: memory references, content selection, and content augmentation. Memory references are a direct way for the robot to show that it remembers the child and make the conversation more personal. Using templated dialog stored information can be included in the conversation [26]. Using the child's name (e.g. [26], [32]), recalling something the child previously disclosed (e.g. [11]), or recalling the score of a previous game (e.g. [32]) are examples of memory references.

When selecting content the robot can use a memory select topics that match with the child's known interests (e.g. [33]). Finally, conversational content can also be augmented with memory information. For example, by adding an opinion congruent with the child's known opinion about a subject [11].

Children perceive a robot more as a friend and are more interested to keep interacting when that robot uses memory references [26]. Older children (7-10) show more positive affect when the robot uses a memory, while younger children (4-6) show more positive affect without a robot memory. Furthermore, especially the older children preferred the robot with memory augmentation and viewed it as more intelligent [11].

Most studies evaluate memory-based personalization strategies in the context of a task (e.g. diabetes management [26] or vocabulary learning [15]) and study effects on task adherence and performance or enjoyment [8]. How it influences a social interaction and relationship formation is not yet systematically studied. As far as we know no longitudinal studies (covering more than a few sessions in a longer time frame) exist that evaluate the long-term effects of memory-based personalization on the child-robot relationship. By doing a longitudinal study outside the lab with an autonomous robot we gain a better insight into how the robot would perform in the real world [34].

III. Design Rationale

The two key element of the design are the serial narrative dialog structure and the memory-based personalization strategy. Both design elements were implemented in an autonomously operating robot (i.e. the researcher only has to start a session). In this section we will specify and motivate the design elements.

A. Serial Narrative Dialog Structure

The main activity of the interaction is a conversation with the robot. To accommodate a multi-session conversation we adopted a serial narrative dialog structure. A proper metaphor would be a serial TV-show. It matches the rhythm of regular sessions that often occur in application contexts like healthcare (e.g. weekly treatments [35]) and education (e.g. tutoring). All the dialog content was developed in collaboration with three professional writers.

The building blocks of the conversation are mini-dialogs. A mini-dialog is a self-contained unit of discourse in the conversation. It is the dialog equivalent of a paragraph in written text. There are three types of mini-dialogs: narrative, chitchat, and functional.

Narrative mini-dialogs are part of a multi-session thread and deliver the overarching storylines of the conversation. Typically there are more narrative threads that are interwoven into the conversation. For example, one narrative thread in our implementation are the Robot Olympics Hero is participating in. Each episode Hero asks the child to help it prepare for the next event and in the final session Hero reveals it won a silver medal. With the narrative threads we aim to provide development, continuity, and familiarity between sessions.

Chitchat mini-dialogs are oriented around a topic. There are topic openers that help identify topics that children are interested in and topic follow-ups to dive deeper into a topic. For example, if the robot asks "what wild animal would you like as a pet?" and the child answers with "a panda", the robot can include a follow-up mini-dialog about pandas next session. With chitchats we aim to introduce novelty and similarity.

Finally, there are functional mini-dialogs that help manage the conversation. For example, a greeting. A fun interaction element we included as part of the greetings is a 'secret handshake'. Children could co-create their secret handshake by choosing between two well known 'tik tok' songs and choreograph a dance move or gesture with the robot's arms (similar to the co-creation process discussed in [36]).

In the user study each 15 minute conversation included about 7 or 8 mini-dialogs (on average 2 minutes per mini-dialog). The first and last were a greeting and goodbye. The main narrative thread was about the Robot Olympics with multiple mini-dialogs per session. There was a secondary thread about dreaming with one mini-dialog per session. The narrative mini-dialogs alternated with chitchats. Each session typically had one topic opener, a topic follow-up, and a stand-alone chitchat. To protect children's privacy only topics were covered that elicit low sensitive information, for example, children's favorite animals, food, color, and differences between robots and children.

Google's Dialogflow was used for automated speech recognition and an artificial cognitive agent[1] (implemented in GOAL [37]) was used for dialog management. The whole pipeline of streaming audio to Google (using the robot's onboard microphone), receiving data, and selecting a response takes 600ms on average. To protect children's privacy the microphone was only turned on to register an answer (average 3s and max. 10s per attempt).

The agent followed the conversational design patterns developed by Ligthart et al. (2019) [31]. Children have two speech attempts. When unsuccessful they have two touch repair attempts. The robot lists a number of answer options and by pressing the button on its foot children can select an answer. Over five session the robot asked 2094 questions. In 82% speech was sufficient and if a touch repair was necessary 94% of the repairs were successful. In less than 1% no answer was recognized. This facilitated a robust and fast yet autonomous and speech-focused interaction.

[1]Code: https://bit.ly/3sdiQsd and video examples: https://bit.ly/3F0TOQF

B. Memory-based Personalization Strategy

We opted for a multifaceted memory-based personalization strategy primarily aimed at manipulating the dialog. The strategy combines memory references, content selection, and content augmentation (as discussed in section II-B). Per sub-strategy we designed multiple manipulations (see Table I for a full overview). This section will cover which information is stored and discuss each manipulation and how it contributes to the overall strategy.

A Redis database is used to persistently keep an interaction history. The main way for the robot to populate the history is to ask questions about children's interests, preferences, and ask them to make narrative decisions. For example, the robot would ask if the child has a pet, their favorite color, or if they want to coach the robot during a Robot Olympics event. The relevant entities are extracted from the answers and stored in the history. Furthermore, the robot would store which mini-dialogs it included in the conversation and the secret handshake created by the child. The writers designed the questions and relationships between mini-dialogs. The GOAL agent reasons about these relationships and history items to select the next mini-dialog. The history items are available to the writers as placeholder variables with which they created templated dialogs. During the conversation the robot will retrieve the values (i.e. children's processed answers) of the placeholder variables from history at run time and insert them in the dialog.

The first substrategy are memory references. Children want to be remembered by the robot [21] and a powerful way for the robot to do that is by using their names [22] and recalling things the child shared with the robot [3], [33]. To protect the child's privacy their name was entered in advanced instead of sending it to Google. The latter manipulation, a personal reference, can be made explicitly (e.g. "I know you like [sheep]") or implicitly (e.g. "I'll pet the cat and of course you can pet the [sheep]"). Using the child's name and personal references are two manipulations that aim to add a level of (personal) continuity. A third memory reference manipulation aims to add (conversational) continuity and familiarity by referring to a past conversational topic. For example, "last time you mentioned [ice cream]". Memory references can also be chained, for example "wouldn't it be cool to have a [purple] [sheep]?".

The second substrategy is content selection. The narrative mini-dialogs offer children choices that influence the narrative immediately (e.g. "should I go left or right in the forest?") and the overarching narrative (e.g. "do you want to be my

Robot Olympics coach?”). Each narrative mini-dialog has a branching structure and using the persistently stored choices the branches are resolved. A second content selection manipulation is to select chitchat mini-dialogs that match with children’s interests that are stored in the user model. These two manipulations aim to increase the similarity between the child and the robot.

The third substrategy is content augmentation. The secret handshake is stored to augment the greeting and goodbye. Repeating iconic parts of the conversation creates more familiarity [38] and additionally a unique secret handshake reinforces the relationship with the child [39]. The second and third augmentations aim to add continuity and familiarity by embedding new content in the context of previous content [3]. The second augmentation motivates the inclusion of a mini-dialog by relating it to the child. For example, “Let’s talk about your favorite sport, [taekwondo]”. The third uses the knowledge of the content selection to foreshadow future mini-dialogs. For example, “Because you like [risotto], I’ll try to make my own tonight” to subsequently tell about its risotto cooking adventure the next time.

In the user study the memory-based personalization strategy was used from the second session onward. Memory references, content motivations, and foreshadowing manipulations all affect one utterance at a time. Once every two minutes one utterance was manipulated with one of these strategies. In the control condition a generic alternative of equal length is included instead. The duration of the conversation in both conditions was the same each session. In each 15 minute conversation the secret handshake (versus a wave in the control condition) was used at the start and at the end. The choices made during the narrative mini-dialogs persisted and influenced the narrative branches (versus a default path). The topics of two (out of seven) mini-dialogs were selected using the memory (versus a random selection). This amounts to approximately one manipulation every minute.

TABLE I
COMPARISON OF ALL THE MANIPULATIONS THAT ARE PART OF THE MEMORY-BASED PERSONALIZATION STRATEGY AND A CONTROL CONDITION.

Manipulation	**Personalization**	**Control**
Child’s name	Used	Not used
	Hi [name], nice to see you	*Hi, nice to see you*
Personal	Explicit and implicit	Generic
	I know you like [sheep]	*I learned about dogs*
Conversational	Explicit	Generic
	Last time you mentioned [ice cream]	*I saw someone eat pizza*
Narrative choices	Persist across sessions	Only immediate
Topic selection	Child’s interests	Random
Greeting	Secret handshake	Default wave
Motivation	Related to child	Generic
	Let’s talk about your favorite sport, [taekwondo]	*Let’s talk about a cool sport, football*
Foreshadowing	Specific	Generic
	Let’s talk about [Risotto] next time	*I hope to make pizza someday*

IV. Research Questions and Hypotheses

The main goal of the memory-based personalization strategy is to facilitate a sustainable long-term interaction. The main means to achieve that goal is to foster the child-robot relationship so that it can replace the novelty effect as a core motivator for children to interact with the robot. Addressing the identified needs with the serial narrative interaction structure and the personalization strategy are in turn the main mechanisms to foster the relationship. The research questions revolve around these three steps.

What is the effect of memory-based personalization on

RQ1. *children’s willingness to continue interacting with the robot?*

RQ2. *the child-robot relationship?*

RQ3. *children’s feeling of continuity, familiarity, and similarity?*

We expect that (hypothesis $H1$) with memory-based personalization children will be more willing to continue the interaction over time than when there is no memory-based personalization.

We measured the child-robot relationship with a closeness self-report questionnaire, by counting the self-disclosures, and annotating (the valence of) social cues. We expect that with memory-based personalization strategy over time children will ($H2_a$) feel closer to the robot, ($H2_b$) self-disclose more, and ($H2_c$) show more positive social cues towards the robot than without personalization.

Finally, we expect that with memory-based personalization children will experience the conversation with the robot as more ($H3_a$) continuous and ($H3_b$) familiar and ($H3_c$) rate the robot as more similar to them.

V. Method

To answer the research questions and evaluate the memory-based personalization strategy we ran a longitudinal user study. 46 school children interacted five times with the robot, roughly on a weekly basis for approximately fifteen minutes in a two-month period. We compared a robot that used the memory-based personalization strategy (personalization condition) with a robot not using any memory-based personalization (control condition). In this section we discuss the methods used to run the longitudinal study.

A. Experimental Design

To structure the study in general we used a mixed experimental design. The first independent variable was the inclusion or exclusion of the memory-based personalization strategy as a two-level between-subjects factor. The difference between the resulting ‘personalization’ and ‘control’ condition are listed in Table I. The second independent variable were the five sessions (time) as a within-subjects factor. To address the third question only a between-subjects comparison was performed. The dependent variables are the willingness to continue ($RQ1$), closeness, self-disclosure count, and valence of social cues ($RQ2$), and continuity, familiarity, and similarity ($RQ3$).

B. Participants

46 Dutch school children (27 girls and 19 boys; 8-10 and 1x 11 y.o.) started the study. Participants were recruited by their teachers from the participating school. The participants represented a diverse population with a mix of different ethnicities and social-economic backgrounds. Most participants had experience with interactive technologies (e.g. voice assistants or toy robots), but none had a conversation with a robot before. This study (ECIS-2021-03) was approved by the Ethical Committee for Information Sciences of the institution of the first author. The participants and their legal guardians signed an informed consent form before participating. The age and gender of the participants were kept balanced while assigning participants to a condition. Participants with the same age and gender were randomly paired. The pair was randomly split between the personalization (N=24) and control condition (N=22). Two participants from the control condition dropped out after three sessions. They indicated they rather wanted to focus on their school work.

C. Measures and Instruments

Measuring a complex concept like the child-robot relationship is not straightforward. The outcomes of the HRI 2021 workshop on Interdisciplinary Research Methods for Child-Robot Relationship Formation [40] showed that we need a solid theoretical grounding for the measures and not limit ourselves to self-report measures [41].

We aim to facilitate a companionship between the child and the robot. A companionship is an interpersonal relationship [42] that is operationally defined as "having fun while involved in complementary and reciprocal social interaction" [43]. It's considered one of the most important functions of an early friendship [44]. Closeness is a concept that characterizes a companionship and is described as the sense of connectedness to others [12], [45]. Van Straten et al. (2020) developed and validated a self-report scale for child-robot closeness [45].

The next step is to extend the measurement toolkit with behavioral measures. To get a closer relationship people self-disclose increasingly intimate information to each other [46], [47]. Counting children's self-disclosures can be indicative of how the relationship progresses [31], [48]. Another, more basic, behavioral measure is to register the type and valence of the social cues the children use during the interaction. The more positive the social cues (e.g. smiling) are, the more expeditious the relationship formation process likely goes [49].

The measures and their instruments are briefly discussed in the remainder of this section[2].

[2]A full overview of the measures: https://bit.ly/3sdiQsd

1) Willingness to continue: To measure participants willingness to continue the interaction the robot acceptance 5-point self-report scale developed and validated by de Jong et al. (2020) was used. The four items (statements) all focused on participant's willingness to interact again with the robot. The scale was developed for a Dutch child-robot interaction with a Nao robot [50]. We added a fifth item asking about participant's willingness to recommend the conversation with the robot to their friends. This measure was used after each session.

2) Closeness: To measure the closeness of the child-robot relationship a 5-point self-report scale developed and validated by van Straten et al. (2020) was used. It contains five items that directly ask about participants feelings of friendship towards the robot [45]. It uses the same rating scale as de Jong et al. (2020) [50]. This measure was used after the first, third (middle) and fifth (last) session.

3) Self-disclosure: Following the work by Kory-Westlund et al. (2018) on using self-disclosure to measure relationship formation [48] we included three open questions as self-disclosure opportunities each session. The prompts were the same in both conditions. Google's Dialogflow was used for speech recognition. The resulting transcripts were used to calculate a combined word count that was used as the self-disclosure score for each session. The next fictional dialog illustrates one of these opportunities.

R What is your favorite pizza topping?
C Chili pepper
R Why is chili pepper your favorite? (prompt)
C Because I like to spice up my pizza. (self-disclosure score of 8)

4) Social cues: We used a focused and feasible approach to measure the valence of social cues displayed by participants similar to the approach used by Serholt and Barendregt (2016) [49]. A probing strategy to collect the social cues, similar to Corrigan et al. (2014) [51], was used. During each session three personalization manipulations were selected as a probe (secret handshake, personal reference, and motivation). With a front-facing video camera short clips were made recording the response of the participant to the manipulation. The direction of the participants gaze (directed at, near, or away from the robot), facial expression (e.g. smiling, concentrated, neutral stare, or bored), and additional gestures (e.g. waving or fidgeting) were annotated by one non-expert coder. Based on these annotations a valence score (-2 to 2) was computed for to each probe. For example, looking at the robot while smiling gets a score of +2, looking concentrated at the robot +1, starting neutrally or starting concentrated near the robot 0, looking away -1, looking away bored or restless -2. An average valence scored was calculated for each session. This measure allows to directly assess the impact of the personalization manipulations.

5) Continuity, familiarity, and similarity: Custom self-report items were made to measure children's feeling of continuity between interaction (4 items), their familiarity during the conversation (3 items), and their similarity to the robot (4 items). The continuity measure focuses on directly assessing the influence of various memory-based personalization features (e.g. “I had the feeling that the conversation with Hero started anew each time”). The familiarly measure focused on assessing the robot’s contribution to the familiarity of the conversation (e.g. “While talking Hero mentioned things we discussed in a previous conversation”). The similarity measure focused on assessing how similar the child feels the robot is to them (e.g. “Hero and I like the same things”). These items were assessed with the same ‘rising bar’ rating scale as [45], [50]. These measures were only used after the last session.

D. Set-up and Procedure

The study was conducted in an empty classroom at the school. The robot was standing on the ground, the children were asked to sit in front of it on a rug. The researcher remained in the room, but was positioned behind the participant, at an appropriate distance, to avoid unnecessary contact between the researcher and participant. The study was run during normal school days and participants came in one-by-one during their lessons.

There were five sessions with the robot. At the start of the first session participants were briefed about what to expect and their rights as participants. When ready, participants were introduced to the robot. This included a five minute tutorial on how to talk to the robot. For the other four interactions, the interaction started right away with the conversation. The conversation took approximately 15 minutes per session.

After the conversation the participants filled in the questionnaires and were interviewed (approx. 10 minutes). When finished with a session, the participant went back to their classroom and called the next participant.

VI. Results

To answer the main research questions and to test the hypothesis formulated in section IV a number of analyses were performed. Only the participants that participated with all the five sessions were included in these analyses ($N_{per} = 24$, $N_{con} = 20$). Data are mean $\pm$ standard error or median [quartiles].

A. Willingness to continue

After each session the willingness to continue to interact with Hero was measured. The scores were generally high across the board (see top left in Figure 1), as is not uncommon for these types of measurements in child-robot interaction research. This negatively skewed the distribution of scores. Therefore, the method of Brunner et al. (2002) [52], and the nparLD R-package [53], for non-parametric analysis of longitudinal data in factorial experiments (with the Wald-Type Statistic [WTS]) is followed to investigate if there is an interaction effect of condition and session on the willingness to continue.

There was no statistically significant interaction between the use of memory-based personalization and the sessions on the willingness to continue, $WTS = 8.59$, $p = .07$. No main effect of condition on the willingness to continue was found, $WTS(1) = 1.88$, $p = 0.17$. However, the main effect of session showed that there was a statistically significant difference in the willingness to continue between different sessions, $WTS = 9.57, p = .04$. Friedman’s tests were ran post hoc to determine if the main effect of session was present in both conditions. A Bonferroni correction was applied to account for multipe testing.

The willingness to continue did not significantly change between sessions in the personalization condition, $\chi^2 = 5.24$, $p = .26$. The willingness to continue did significantly decreased between sessions in the control condition, $\chi^2 = 22.24$, $p < .0005$. The differences were significant between the last (4.3 [3.6, 4.7]) and the first and second (both 4.6 [4.4, 5.0]) sessions, $p's < .04$.

Looking at scores and 95% confidence intervals the scores in both conditions appear to remain the same for the first three sessions and seem to diverge in the last two sessions (due to the decrease in the control condition). To explore this trend Mann-Whitney U tests were ran on the last two sessions. In session 4 the scores in the personalization condition (4.8 [4.4, 5.0]) and control condition (4.5 [3.8, 5.0]) were not statistically significantly different, $U = 300$, $z = 1.45$, $p = 0.15$. In session 5 participants were statistically significantly more willing to continue in the personalization condition (4.8 [4.6, 5.0]) than in the control condition (4.3 [3.6, 4.7]), $U = 350.5$, $z = 2.68$, $p = .007$, Cohen’s $d = 0.85$.

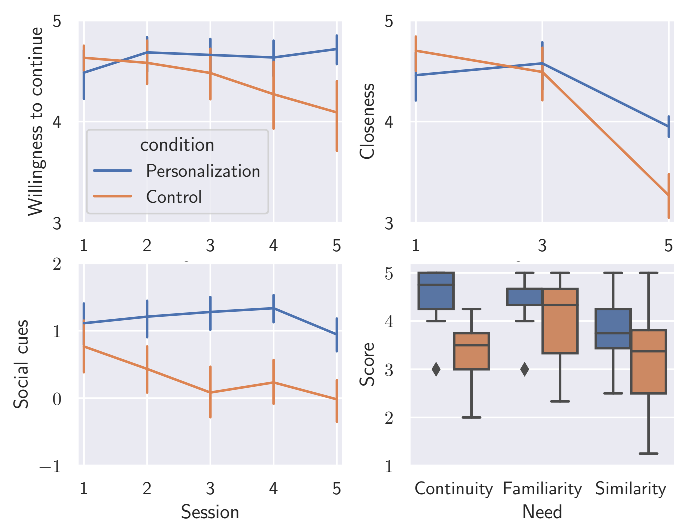

Fig. 1. Scores with 95% CI of the willingness to continue (top left), closeness (top right), and valence of social cues (bottom left) over time per condition and the scores for continuity, familiarity, and similarity (bottom right) per condition after the last session. All scores range from 1-5 except social cue valence that ranges from -2 to +2.

B. Closeness

Closeness was measured after the first, third (middle), and fifth (last) session (see top right in Figure 1). A two-way mixed ANOVA was used to determine if there was an interaction effect of condition and session on the scores. Post hoc ANOVA's were ran to test simple main effects for condition and session on the scores. Bonferroni corrections were applied to account for multiple testing. There was a statistically significant interaction between the use of personalization and the sessions on closeness, $F(2, 76) = 16.33$, $p < 0.0005$, partial $\eta^2 = .30$.

Similarly to the willingness to continue, the closeness scores remain fairly constant between the first and middle session. Between the middle and last session closeness declines in both conditions. However, the decline in the personalization condition is less steep than in the control condition. This is shown by the following results. For both the control condition and the personalization condition there was no statistically significant difference in closeness between the first and middle session, $\delta_{con} = .02 \pm .09$, $p_{con} = 1.0$ and $\delta_{per} = -.17 \pm .09$, $p_{per} = .22$. There was a statistically significant difference between the first and last and the middle and last session for both conditions (all $\delta's > .55 \pm .11$, all $p's < .0005$). There was no statistical significant difference in closeness between conditions after the first ($\delta = .222 \pm .17$, $p = .20$) and middle ($\delta = .03 \pm .12$, $p = .78$) session. Closeness was, however, statistically significantly higher after the last session with personalization ($M = 3.95 \pm .07$) compared to the control ($M = 3.41 \pm .08$), $F(1.38) = 25.32$, $p < .0005$, partial $\eta^2 = .40$.

C. Self-disclosure

A two-way mixed ANOVA was performed. No statistically significant interaction between the use of personalization and the sessions on the amount of self-disclosure was found, $F(4, 139.28) = .89$, $p = .46$. No main effect of session on self-disclosure was found, $F(4, 139.28) = 2.19$, $p = .083$. No main effect of condition on self-disclosure was found, $F(1, 40) = .058$, $p = .81$.

D. Social cues

To analyse the valence of social cues the method of Brunner et al. (2002) [52] was used again. There was a statistical significant condition and session interaction effect on social cues, $WTS(4) = 11.27$, $p = 0.02$. Also a main effect of condition, $WTS(1) = 34.80$, $p < 0.0005$, and session, $WTS(4) = 25.73$, $p < 0.0005$ on social cues were found. Post hoc testing were done with a Bonferroni correction. See bottom left in Figure 1.

Mann-Whitney U tests revealed that only during the first session there was no statistical significant difference between the personalization (1.33) and the control (0.67) condition, $U = 294.00$, $p = .19$. In all other sessions the social cues were more positive in the personalization condition (≥ 1.0) than in the control condition ($\leq .67$), all $U's > 391.50$ and $p's < .0005$. A Friedman test was run for both conditions to determine if there were differences in social cues during the five sessions. Pairwise comparisons were performed. In both conditions no pair of sessions statistically significantly differed from each other, all $\chi^{2\prime}s \leq 1.27$, all $p's \geq .054$.

The probes covered three types of personalization manipulations: personal memory references ($\delta_{valence} = .75$), content motivation ($\delta_{valence} = 1.0$), and the secret handshake ($\delta_{valence} = 1.2$). All three manipulations statistically significantly elicited more positive social cues, all $U's \geq 413.5$, all $p's \leq .0005$, Cohen's $d's \geq 1.6$. There was no statistical significant difference in valence scores between the three manipulations, $\chi^2 = 5.48$, $p = .07$. Smiling responses are a big factor that contributes to the valence scores. On average 44% of the participants (compared to 14% in the control) smiled during the secret handshake. A personal memory reference resulted in 36% (vs. 6%) of the participants smiling and the content motivation in 31% (vs. 9%).

E. Continuity, familiarity, and similarity

Mann-Whitney U tests were ran to determine if there were differences in continuity, familiarity and similarity scores between participants interacting with the robot in the personalization or control condition. A Bonferroni correction was applied to account for the repeated testing. See bottom right in Figure 1.

The continuity scores in the personalization condition (4.75 [4.3, 5.0]) were statistically significantly higher than in the control condition (3.50 [3.0, 3.8]), $U = 462$, $z = 5.27$, $p < .0005$, Cohen's $d = 2.57$. The familiarity scores in the personalization condition (4.67 [4.3, 4.7]) and control condition (4.33 [3.3, 4.7]) were not statistically significantly different, $U = 302$, $z = 1.48$, $p = .14$. The similarity scores in the personalization condition (3.75 [3.3, 4.3]) and control condition (3.38 [2.5, 3.9]) were also not statistically significantly different, $U = 314$, $z = 1.75$, $p = .08$.

VII. Discussion

A. Sustainable Long-Term Interaction

Overall the children were very willing to continue the interaction. The mean scores were all between the 4 and 5 on a 5-point scale. This is not uncommon for self-report measures that focus on the subjective experience of the interaction. The novelty in the beginning and the quality of the narrative dialogs in general are two likely factors that contribute to these high scores. However, after four sessions the value of the memory-based personalization strategy starts to show. Children's interest in the interaction remains at the same high level when the personalization strategy is used. In the control condition children's interest starts to decrease over time. Children significantly are more interested to continue the interaction after the last session with personalization than without. These results confirm hypothesis $H1$. Memory-based personalization is an investment that pays off after a few sessions and contributes to facilitating a sustainable long-term interaction.

B. Fostering the Child-Robot Relationship

Children experienced the same levels of closeness in their friendship with the robot regardless of condition at the start and half way. After the last sessions children significantly felt less close to the robot in both conditions. However, children felt significantly closer to the robot that used the memory-based personalization strategy. The overall decrease in closeness does not necessarily mean that the relationship with the robot deteriorated. It could also reflect a shift in how children evaluated the robot. At first children might have rated the idea of a relationship with the robot ('as what could be') and later switched to rate it based on their experienced closeness ('as is'). This is one of the key reasons why it is important to run a longitudinal study. The results did show that after a while memory-based personalization fosters more closeness between the child and the robot than without it, confirming hypothesis $H2_a$.

Children did not self-disclose more to a robot with a personalization strategy. As a result we have to reject hypothesis $H2_b$. This might be a result of the fairly restrictive way children could participate in the conversation. The self-disclosure prompts were all in relation to explaining a choice the child made (e.g. 'Why is chili pepper your favorite?'). Although children were free to elaborate as much as they wanted, they resorted to giving basic factual explanations. It is likely that their self-disclosures did not reflect how they experienced their relationship with the robot. Whether self-disclosure is a suitable measure is also up for discussion, because it is likely affected by several social processes and is subject to a lot of interpersonal variation, making it an imprecise measure [54].

The results finally showed that the personalization manipulations elicited more positive social cues compared to the control condition. For example, children smiled more when the robot displayed the secret handshake or made a personal memory reference. This effect persisted over time. We can accept hypothesis $H2_c$. All manipulations that were probed contributed equally to this effect. This shows the value of using a multifaceted strategy.

C. Addressing Children's Needs

The design of the memory-based personalization strategy revolves around addressing certain needs children have for their relationship with the robot. We measured for the need for continuity, familiarity, and similarity. Only the need for continuity is significantly supported by the personalization strategy. No significant differences were found for familiarity and similarity.

In hindsight, how me measured these needs might not reflect their actual need satisfaction. Continuity was directly linked to the memory manipulations and can be better interpreted as a manipulation check. Familiarity and similarity are more tied to the overall conversation. And although the manipulations explicitly highlight familiar content and similarities, the general narrative design has probably a bigger impact on those aspects. It likely equalizes any difference that might exist. Looking at the scores children generally experienced that the conversation contained familiar elements. Children felt neither similar or dissimilar to the robot, although it is unclear what level of similarity is fitting. More research is needed to get insight into what kind of similarities are beneficial to highlight and how to best highlight them.

D. Limitations and Future Work

In our longitudinal study we mostly observed differences in the last session. To get a better picture of the development of children's interest and relationship with the robot more sessions and participants are helpful. It can give insight in how long the current personalization strategy remains effective and to further ensure children evaluate the robot 'as is' and not 'as what could be'.

In our current study a more detailed examination of the effects and satisfaction of specific personalization manipulations as well as the effect of the serial narrative dialog on the child-robot relationship is lacking. We collected qualitative data to facilitate that examination. That analysis is left for future work. We furthermore used only one coder for annotating the social cues. More coders would increase the reliability of the valence scores.

The memory-based personalization strategy can be improved by incorporating best-practices from the cognitive architecture community.

VIII. Conclusion

Most social robot applications for children, especially those that offer social support, are meant to operate long-term. However, when the novelty of the robot wears off children lose interest. To fully benefit from the support the robot has to offer children need a new motivation to keep interacting. Enabling children to form a relationship with the robot is key for facilitating a sustainable long-term interaction.

To tackle that long standing challenge we share our designs for a serial narrative dialog and memory-based personalization strategy. The serial narrative dialog enables children to robustly have a series of (speech) conversations with an autonomous robot. The main challenge was to create enough high quality content. We employed the help of three professional writers. Evaluating that process and the created mini-dialogs is left for future work.

The memory-based personalization strategy populates a persistent interaction history by eliciting the children to self-disclose about their interests and preferences. That information is used to make the conversation more personal by making memory references and to tailor the conversation to the child's interests via content selection and augmentation.

With a longitudinal (five sessions in two months) user study (N = 46, 8-10 y.o) we successfully demonstrated that the memory-based personalization strategy provides continuity between sessions. Children feel closer to the robot and they smile more after a personalization manipulation. It keeps children more interested to interact with the robot. With this paper we contribute by proving a set of eight concrete, reusable, and validated memory manipulations that long-term foster the child-robot relationship and facilitate a sustainable long-term interaction.

REFERENCES

[1] T. Kanda, T. Hirano, D. Eaton, and H. Ishiguro, "Interactive robots as social partners and peer tutors for children: A field trial," *Human–Computer Interaction*, vol. 19, no. 1-2, pp. 61–84, 2004.

[2] I. Leite, C. Martinho, A. Pereira, and A. Paiva, "As time goes by: Long-term evaluation of social presence in robotic companions," in *RO-MAN 2009 - The 18th IEEE International Symposium on Robot and Human Interactive Communication*, 2009, pp. 669–674.

[3] I. Leite, C. Martinho, and A. Paiva, "Social robots for long-term interaction: A survey," *International Journal of Social Robotics*, vol. 5, no. 2, pp. 291–308, Apr 2013. [Online]. Available: https://doi.org/10.1007/s12369-013-0178-y

[4] T. Kanda, R. Sato, N. Saiwaki, and H. Ishiguro, "A two-month field trial in an elementary school for long-term human—robot interaction," *IEEE Transactions on Robotics*, vol. 23, no. 5, pp. 962–971, Oct 2007.

[5] B. Irfan, A. Ramachandran, S. Spaulding, D. F. Glas, I. Leite, and K. L. Koay, "Personalization in long-term human-robot interaction," in *Proceedings of the 14th ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '19. IEEE Press, 2019, p. 685–686.

[6] J. Saunders, D. S. Syrdal, K. L. Koay, N. Burke, and K. Dautenhahn, ""teach me–show me"—end-user personalization of a smart home and companion robot," *IEEE Transactions on Human-Machine Systems*, vol. 46, no. 1, pp. 27–40, 2016.

[7] M. K. Lee, J. Forlizzi, S. Kiesler, P. Rybski, J. Antanitis, and S. Savetsila, "Personalization in hri: A longitudinal field experiment," in *Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '12. New York, NY, USA: Association for Computing Machinery, 2012, p. 319–326.

[8] M. Ahmad, O. Mubin, and J. Orlando, "A systematic review of adaptivity in human-robot interaction," *Multimodal Technologies and Interaction*, vol. 1, no. 3, 2017. [Online]. Available: https://www.mdpi.com/2414-4088/1/3/14

[9] G. Castellano, R. Aylett, K. Dautenhahn, A. Paiva, P. W. McOwan, and S. Ho, "Long-term affect sensitive and socially interactive companions," in *Proceedings of the 4th International Workshop on Human-Computer Conversation*. Citeseer, 2008, pp. 1–5.

[10] P. Baxter and T. Belpaeme, "Pervasive memory: The future of long-term social hri lies in the past," in *Third international symposium on new frontiers in human-robot interaction at AISB*, 2014.

[11] I. Leite, A. Pereira, and J. F. Lehman, "Persistent memory in repeated child-robot conversations," in *Proceedings of the 2017 Conference on Interaction Design and Children*, ser. IDC '17. New York, NY, USA: Association for Computing Machinery, 2017, p. 238–247. [Online]. Available: https://doi.org/10.1145/3078072.3079728

[12] E. S. Berscheid and P. C. Regan, *The psychology of interpersonal relationships*. Psychology Press, 2005.

[13] Y. Fernaeus, M. Håkansson, M. Jacobsson, and S. Ljungblad, "How do you play with a robotic toy animal? a long-term study of pleo," in *Proceedings of the 9th International Conference on Interaction Design and Children*, ser. IDC '10. New York, NY, USA: Association for Computing Machinery, 2010, p. 39–48.

[14] T. Salter, K. Dautenhahn, and R. te Bockhorst, "Robots moving out of the laboratory - detecting interaction levels and human contact in noisy school environments," in *13th IEEE International Workshop on Robot and Human Interactive Communication (RO-MAN)*. Kurashiki, Okayama, Japan: IEEE, 2004, pp. 563–568.

[15] M. I. Ahmad, O. Mubin, S. Shahid, and J. Orlando, "Robot's adaptive emotional feedback sustains children's social engagement and promotes their vocabulary learning: a long-term child–robot interaction study," *Adaptive Behavior*, vol. 27, no. 4, pp. 243–266, 2019. [Online]. Available: https://doi.org/10.1177/1059712319844182

[16] M. Díaz, N. Nuño, J. Saez-Pons, D. E. Pardo, and C. Angulo, "Building up child-robot relationship for therapeutic purposes: From initial attraction towards long-term social engagement," in *2011 IEEE International Conference on Automatic Face Gesture Recognition (FG)*, 2011, pp. 927–932.

[17] S. M. Lwin, "Capturing the dynamics of narrative development in an oral storytelling performance: A multimodal perspective," *Language and Literature*, vol. 19, no. 4, pp. 357–377, 2010. [Online]. Available: https://doi.org/10.1177/0963947010373029

[18] H. P. Grice, *Logic and Conversation*. Leiden, The Netherlands: Brill, 1975, pp. 41 – 58. [Online]. Available: https://brill.com/view/book/edcoll/9789004368811/BP000003.xml

[19] G. L. McKinley, S. Brown-Schmidt, and A. S. Benjamin, "Memory for conversation and the development of common ground," *Memory & Cognition*, vol. 45, no. 8, pp. 1281–1294, 2017. [Online]. Available: https://doi.org/10.3758/s13421-017-0730-3

[20] G. Skantze, "Turn-taking in conversational systems and human-robot interaction: A review," *Computer Speech & Language*, vol. 67, p. 101178, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S088523082030111X

[21] O. A. B. Henkemans, V. Hoondert, F. Schrama-Groot, R. Looije, L. L. Alpay, and M. A. Neerincx, ""i just have diabetes": children's need for diabetes self-management support and how a social robot can accommodate their needs," *Patient Intelligence*, vol. 4, pp. 51–61, 2012.

[22] R. Ros, M. Nalin, R. Wood, P. Baxter, R. Looije, Y. Demiris, T. Belpaeme, A. Giusti, and C. Pozzi, "Child-robot interaction in the wild: Advice to the aspiring experimenter," in *Proceedings of the 13th International Conference on Multimodal Interfaces*, ser. ICMI '11. New York, NY, USA: Association for Computing Machinery, 2011, p. 335–342.

[23] B. R. Burleson, W. Samter, and A. E. Lucchetti, "Similarity in communication values as a predictor of friendship choices: Studies of friends and best friends," *Southern Communication Journal*, vol. 57, no. 4, pp. 260–276, 1992. [Online]. Available: https://doi.org/10.1080/10417949209372873

[24] D. Byrne, "Interpersonal attraction and attitude similarity," *The Journal of Abnormal and Social Psychology*, vol. 62, no. 3, pp. 713–715, 1961.

[25] M. Selfhout, J. Denissen, S. Branje, and W. Meeus, "In the eye of the beholder: Perceived, actual, and peer-rated similarity in personality, communication, and friendship intensity during the acquaintanceship process," *Journal of Personality and Social Psychology*, vol. 96, no. 6, pp. 1152–1165, 2009.

[26] I. Kruijff-Korbayová, E. Oleari, A. Bagherzadhalimi, F. Sacchitelli, B. Kiefer, S. Racioppa, C. Pozzi, and A. Sanna, "Young users' perception of a social robot displaying familiarity and eliciting disclosure," in *Social Robotics: 7th International Conference, ICSR 2015, Paris, France, October 26-30, 2015, Proceedings*, A. Tapus, E. André, J.-C. Martin, F. Ferland, and M. Ammi, Eds. Cham: Springer International Publishing, 2015, pp. 380–389.

[27] H. H. Clark, *Using language*. Cambridge university press, 1996.

[28] D. Edwards and D. Middleton, "Joint remembering: Constructing an account of shared experience through conversational discourse," *Discourse Processes*, vol. 9, no. 4, pp. 423–459, 1986. [Online]. Available: https://doi.org/10.1080/01638538609544651

[29] S. Kiesler, "Fostering common ground in human-robot interaction," in *ROMAN 2005. IEEE International Workshop on Robot and Human Interactive Communication, 2005.*, 2005, pp. 729–734.

[30] T. Matsumoto, S. Satake, T. Kanda, M. Imai, and N. Hagita, "Do you remember that shop? computational model of spatial memory for shopping companion robots," in *Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '12. New York, NY, USA: Association for Computing Machinery, 2012, p. 447–454.

[31] M. Ligthart, T. Fernhout, M. A. Neerincx, K. L. A. van Bindsbergen, M. A. Grootenhuis, and K. V. Hindriks, "A child and a robot getting acquainted - interaction design for eliciting self-disclosure," in *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems*, ser. AAMAS '19. International Foundation for Autonomous Agents and Multiagent Systems, 2019, pp. 61–70.

[32] M. I. Ahmad, O. Mubin, and J. Orlando, "Adaptive social robot for sustaining social engagement during long-term children–robot interaction," *International Journal of Human–Computer Interaction*, vol. 33, no. 12, pp. 943–962, 2017. [Online]. Available: https://doi.org/10.1080/10447318.2017.1300750

[33] N. Mattar and I. Wachsmuth, "Let's get personal," in *Human-Computer Interaction. Advanced Interaction Modalities and Techniques*, M. Kurosu, Ed. Cham: Springer International Publishing, 2014, pp. 450–461.

[34] Belpaeme, Tony, "Advice to new human-robot interaction researchers," in *Human-robot interaction : evaluation methods and their standardization*, ser. Springer Series on Bio- and Neurosystems, Jost, Céline and Le Pévédic, Brigitte and Belpaeme, Tony and Bethel, Cindy and Chrysostomou, Dimitrios and Crook, Nigel and Grandgeorge, Marine and Mirnig, NicolE, Ed. Springer, 2020, vol. 12, pp. 355–369. [Online]. Available: {http://dx.doi.org/10.1007/978-3-030-42307-0_14}

[35] M. Ligthart, K. Hindriks, and M. A. Neerincx, "Reducing stress by bonding with a social robot: Towards autonomous long-term child-robot interaction," in *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '18. New York, NY, USA: ACM, 2018, pp. 305–306. [Online]. Available: http://doi.acm.org/10.1145/3173386.3176904

[36] M. E. Ligthart, M. A. Neerincx, and K. V. Hindriks, "Co-creation as a facilitator for co-regulation in child-robot interaction," in *Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '21 Companion. New York, NY, USA: Association for Computing Machinery, 2021, p. 298–302. [Online]. Available: https://doi.org/10.1145/3434074.3447180

[37] K. V. Hindriks, *Programming Rational Agents in GOAL*. Boston, MA: Springer US, 2009, pp. 119–157. [Online]. Available: https://doi.org/10.1007/978-0-387-89299-3_4

[38] N. R. Norrick, "Functions of repetition in conversation:," *Text - Interdisciplinary Journal for the Study of Discourse*, vol. 7, no. 3, pp. 245–264, 1987. [Online]. Available: https://doi.org/10.1515/text.1.1987.7.3.245

[39] L. W. Friedman and H. H. Friedman, "I-get-it as a type of bonding humor: The secret handshake," *Available at SSRN 913622*, 2003.

[40] R. Stower, M. Ligthart, C. van Straten, N. Calvo-Barajas, E. Velner, and T. Beelen, "Interdisciplinary research methods for child-robot relationship formation," in *Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '21 Companion. New York, NY, USA: Association for Computing Machinery, 2021, p. 700–702.

[41] ——, "Outcomes of the workshop on interdisciplinary research methods for child-robot relationship formation," May 2021. [Online]. Available: https://child-robot-interaction.github.io/outcome.html

[42] D. Buhrmester and W. Furman, "The development of companionship and intimacy," *Child Development*, vol. 58, no. 4, pp. 1101–1113, 1987. [Online]. Available: http://www.jstor.org/stable/1130550

[43] N. Bauminger-Zviely and G. Agam-Ben-Artzi, "Young friendship in hfasd and typical development: Friend versus non-friend comparisons," *Journal of Autism and Developmental Disorders*, vol. 44, no. 7, pp. 1733–1748, 2014. [Online]. Available: https://doi.org/10.1007/s10803-014-2052-7

[44] C. Howes, *The earliest friendships*, ser. Cambridge studies in social and emotional development. New York, NY, US: Cambridge University Press, 1996, pp. 66–86.

[45] C. L. v. Straten, R. Kühne, J. Peter, C. de Jong, and A. Barco, "Closeness, trust, and perceived social support in child-robot relationship formation: Development and validation of three self-report scales," *Interaction Studies*, vol. 21, no. 1, pp. 57–84, 2020.

[46] I. Altman and D. Taylor, "Social penetration theory," *New York: Holt, Rinehart & Winston*, 1973.

[47] A. Aron, E. Melinat, E. N. Aron, R. D. Vallone, and R. J. Bator, "The experimental generation of interpersonal closeness: A procedure and some preliminary findings," *Personality and Social Psychology Bulletin*, vol. 23, no. 4, pp. 363–377, 1997.

[48] J. M. K. Westlund, H. W. Park, R. Williams, and C. Breazeal, "Measuring young children's long-term relationships with social robots," in *Proceedings of the 17th ACM Conference on Interaction Design and Children*, ser. IDC '18. New York, NY, USA: Association for Computing Machinery, 2018, p. 207–218.

[49] S. Serholt and W. Barendregt, "Robots tutoring children: Longitudinal evaluation of social engagement in child-robot interaction," in *Proceedings of the 9th Nordic Conference on Human-Computer Interaction*, ser. NordiCHI '16. New York, NY, USA: Association for Computing Machinery, 2016.

[50] C. de Jong, R. Kühne, J. Peter, C. L. van Straten, and A. Barco, "Intentional acceptance of social robots: Development and validation of a self-report measure for children," *International Journal of Human-Computer Studies*, vol. 139, p. 102426, 2020.

[51] L. J. Corrigan, C. Basedow, D. Küster, A. Kappas, C. Peters, and G. Castellano, "Mixing implicit and explicit probes: Finding a ground truth for engagement in social human-robot interactions," in *Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '14. New York, NY, USA: Association for Computing Machinery, 2014, p. 140–141.

[52] E. Brunner, S. Domhof, and F. Langer, *Nonparametric analysis of longitudinal data in factorial experiments*. Wiley-Interscience, 2002, vol. 373.

[53] K. Noguchi, Y. R. Gel, E. Brunner, and F. Konietschke, "nparld: An r software package for the nonparametric analysis of longitudinal data in factorial experiments," *Journal of Statistical Software, Articles*, vol. 50, no. 12, pp. 1–23, 2012.

[54] R. Stower, N. Calvo-Barajas, G. Castellano, and A. Kappas, "A meta-analysis on children's trust in social robots," *International Journal of Social Robotics*, 2021. [Online]. Available: https://doi.org/10.1007/s12369-020-00736-8