

Behavioural Observations as Objective Measures of Trust in Child-Robot Interaction: Mutual Gaze

Anastasia Akkuzu

Department of Information Technology
Uppsala University
Uppsala, Sweden
beliz.akkuzu.0555@student.uu.se

ABSTRACT

Given the subjective nature of trust as a phenomenon and its unified multifaceted contributions for every individual context, the development of a computational model of trust proves to be a difficult endeavour. In this study, we investigate mutual gaze as a behavioural measure of social trust and liking in child-robot interaction. Developing on a prior user study involving 52 children interacting with a robot with variable human-likeness and lexical alignment in two interaction contexts (task-based and dialogue-based), we investigate the effects of human-likeness and lexical alignment on mutual gaze, associations and correlations between metrics assessing social trust and liking, and the development of mutual gaze as an objective measure of social trust and liking. We achieve this through several statistical analyses between the percent of mutual gaze in each interaction, human-likeness, lexical alignment, scores from social trust and liking metrics, self-disclosure content, age, and time. The main findings of our study support the use of mutual gaze as an objective measure for liking, but there is still not sufficient evidence to support the use of mutual gaze as an objective measure to identify and capture social trust as a whole. Furthermore, we found that human-likeness and lexical alignment do not significantly affect mutual gaze in an interaction, but the interaction context does. Moreover, it seems that age plays a role in the amount of mutual gaze in an interaction, where older participants engage in less mutual gaze compared to the younger participants. Alongside this, the amount of mutual gaze the participant engages in is stable across periods when they are not interacting with the robot, changing more towards the first half of the first interaction and the second half of the second interaction. Based on the study, our findings suggest using different objective behavioural measures for social trust compared to its related concepts such as liking. Also, our results have found that there may be other constructs intertwined with

liking, such as attention and interest, which may need to be addressed with separate metrics.

Author Keywords

Child-Robot Interaction, Mutual Gaze, Objective Measures of Trust, Computational Model of Trust, Behavioural Measures

1. INTRODUCTION

As technology in the realm of human-robot interaction (HRI) develops and becomes highly enmeshed in daily life, it becomes more and more necessary to identify and measure trust between humans and machines. From work environments, to schools, to private homes, robots are already beginning to exist alongside humans, and in the human social space. Necessarily, when in a social situation with a human, the robot is expected to act and present itself in a certain way as a means of developing a social relationship with the human [21]. Of course, as in any social relationship, there are elements of trust that are developed during interactions. While trust itself as a construct is highly elusive and subjective, there have been studies towards developing and understanding of the role of trust during social interactions in the field of HRI in various contexts. The definitions, measures, and models of trust in a work environment with adults differs greatly to the definitions, measures, and models of trust in an educational environment with children, even between individual studies [15, 37]. Furthermore, most of these metrics are designed for adult interactions, which may not accurately capture child-robot interaction (cHRI) due to differences in the stages of cognitive development [6].

Looking at the current state of identifying and measuring trust in child-robot interaction, quantitative self-reported metrics are prevalent in the methodologies, with some data being extracted from interviews or observations [37]. One of the limits of self-reported metrics are the intentionality that may affect the results, especially in a situation where an adult is interacting with a child. Due to the imbalance in the power dynamic between adults and children in an experimental setting, the child may seek social acceptance through withholding their true thoughts and feelings [6], which may lead to ceiling effects in the resulting data. To combat this, there is a push towards understanding unconscious behaviour as an indicator and measure of trust in cHRI. Due to the multifaceted and thoroughly interwoven nature of trust, it is currently not possible to

study all the factors of trust at once; the standardisation of objective measure of trust follows a bottom-up approach. Thus, trust can be split into domains to identify the context of the interaction, which necessarily affects the behaviour associated with it.

One behaviour that is quite crucial to building trust in interpersonal relationships is gaze, specifically mutual gaze [21]. Mutual gaze can also be defined as eye contact between two social agents in an interaction, and can be engaged both unconsciously or consciously. In having this clear definition, mutual gaze is a behaviour that can be identified using automated processes, which opens the possibility to developing a measure of trust that can be utilised in cHRI.

2. RELATED WORK

2.1 Defining Trust

Firstly, to measure trust within the context of child-robot interaction (cHRI), an operational definition of trust should be developed to provide a basis from which researchers can infer the occurrence of trust. This has proven to be a difficult task, given the subjective nature of trust as a phenomenon and requiring the unified contributions of many facets for every individual context.

In prior work, [22] establishes that trust is born from an ordered bidirectional willingness to cooperate, defined as the display of trusting behaviour by one agent in one direction and the assessed trustworthiness of another agent in the other direction. This can be contrasted with the notion of vulnerability from the definition of trust in [23], where willingly lacking the ability to monitor or control gives rise to trust. The notion of ‘willingness’ drives both definitions, which leads to the understanding of trust as an intentional behaviour in an interaction. However, in later works trust is defined as “a form of affiliation or credit” that may not always be intentionally expressed through behaviour [18]. This definition directly counteracts the notion that trust is an intentional behaviour. In addition to this, in [38], trust is characterised as a belief, neither a choice nor obligation, related to decreasing perceived risk which guides the behaviour of an agent. From these definitions, it is evident that trust cannot be defined in strictly general terms.

To combat this, we must identify the context in which trust occurs, which allows for the introduction of context-dependent parameters of trust. [20] identifies a common trend amongst these definitions of trust which applies to the realm of human-robot interaction (HRI), which states that trust is the correspondence between robot and human actions and behaviours. From this domain-specificity understanding of trust, we can further develop an understanding of which robot actions/behaviours and human actions/behaviours entail trust and how to measure these actions and behaviours.

2.2 Domains and Factors of Trust in HRI

From within HRI, there are two clear domains of trust that can be identified: social trust and competency trust; integrity appears as a third pseudo-domain closely related to social trust [32]. While competency trust is related to reliability,

competence, and performative skill [32], social trust can be defined by the interpersonal aspects of an interaction such as keeping secrets and promises [37]. As such, competency trust is more closely aligned with the mechanical capabilities of robots in industrial or warfighting contexts [29, 17, 8], where its facets are in direct relation to its function. With the introduction of social robots into various parts of society, there is a necessity to measure social trust more rigorously alongside competency trust due to the introduction of interpersonal facets relating to trust, by definition of a ‘social robot’. While many studies are already underway in understanding social trust with adult participants, some being [10, 38, 28], there is relatively little attention towards child-robot interaction, especially in an educational context where social robots are already being deployed.

In developing an understanding of social trust, or any kind of trust, we must look at the related factors, in a top-down fashion. By analysing these related factors, we can find indicators of trust, trustworthiness, and other constructs on a behavioural level. These factors can be classified into 3 groupings: human-related factors, robot-related factors, and environmental factors from the meta-analysis in [17]. Furthermore, [17] also shows that robot-related factors have the largest impact on trust, followed by environmental factors, and finally, human-related factors. As an example of the connection between factors and behavioural data, within [20], the robot-related factors can be further categorised into three more groups, namely performance-related factors, behaviour-related factors, and appearance-related factors. If we revisit the domains of trust outlined by [32], we can see that the performance-related factors relate with the domain of competency trust, behaviour-related factors relate to social trust, and appearance-related factors can be loosely linked to integrity. In addition to this, [19] demonstrates that eye gaze information (mutual gaze, gaze duration, pupil dilation, and distance of eye movement) can be used to develop a model to indicate and estimate engagement.

And since engagement is a behaviour-related factor under robot-related factors of trust in [20], the links between behaviour, factor, domain, and phenomenon are illustrated.

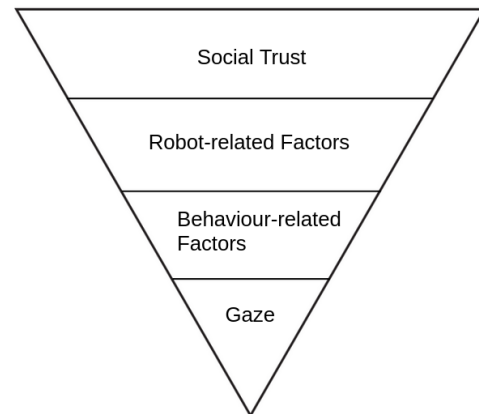


Figure 1. Top-down visualisation of the link between social trust and gaze

2.3 Measuring Trust through Behaviour in HRI/cHRI

2.3.1 Human Behaviour

Measuring behavioural data as indicators of trust is newly becoming standardised, with interest towards developing models of trust, which could be used to objectively measure occurrences and/or levels of trust in HRI. Currently, trust in HRI is measured through either questionnaires/surveys/self-disclosure, or through behavioural data while completing tasks [13]. Moreover, [18] has begun the development of a standardised coding scheme for multimodal behavioural measures of trust in HRI. While this is great progress for cross-comparing experiment data for adults, there are still many child-related factors that are omitted such as age-related interpretation of trust and related constructs, and the effect of social judgments, such as liking and friendship, on perceptions of trust [32]. One way to target trust, and trust only, is to minimise interpretability through emphasising behavioural data in child-robot interaction (cHRI); this would mean less intentionality in expressing trust, which would mitigate some aspects of subjectivity in the data. Also, by targeting behavioural data, we can utilise the results in the development of a computational model of trust, which would increase the generalisability of the current models, such as the one presented in [22]. Currently, there are several studies focusing on multimodal behavioural measures of trust in cHRI [7, 30, 32, 37, 34]. Due to the multimodal measures of trust, disentangling the effects of individual behaviours may not be possible, which leads us to the current study.

2.3.2 Robot Behaviour

On the other hand, robot behaviour can also be manipulated to elicit trust during an interaction [20]. Within robot behaviour, we can identify three modulating components (embodiment, personality, and social presence) that affect the perception of the robot, and then, the expression of trust that is observed.

Firstly, modifying the embodiment of the robot in terms of the human-likeness of its appearance has significant effects in predicting trust during an interaction [26]. An industrial robot may not be seen as human-like as a humanoid robot, but a highly humanoid robot can also produce feelings of disgust or revulsion, also known as the uncanny valley phenomenon [25]; these feelings of disgust and revulsion are detrimental to interpersonal trust. Additionally, the embodiment of the robot also impacts the participation and physical abilities that it is capable of displaying, limiting the multimodality of the robot's behaviour and expression. However, in certain contexts such as emergency evacuations, limiting this multimodality is beneficial to task performance [20]. In socialising, people often rely on this multimodal behaviour as crucial information in an interaction, which can inform their future actions and reactions [16], and provide an understanding of the robot's personality [35].

Secondly, the perceived personality of the robot plays a critical role in human-robot interactions [35]. A common measure of personality traits across HRI studies is based on the Big Five personality traits [24], which groups personality traits into five main categories: Openness to experience, Conscientiousness, Extroversion, Agreeableness, and Neuroticism. By

manipulating these trait categories in the robot personality, the effects are identified and observed in the reported perceptions of robot behaviour. Since personality is used to explain and predict behaviour in a social setting [39], it is necessary to understand the relations between robot personality traits and social trust.

Lastly, the behaviour of a robot in a social interaction can also be situated by identifying other factors that affect the social presence of the robot, such as perceived gender through modulating the voice and appearance. While social psychology literature often points to same-gender preferences during social interactions, there are findings that also point to cross-gender (in a binary classification of genders) preferences when interacting with a robot [31]. In [36], it was also found that robots that conformed to existing social gender norms were more socially accepted.

2.4 Current Study

In this study, we investigate mutual gaze as an objective measure of interpersonal trust and related constructs in cHRI. The questions we aim to satisfy within the scope of this project are:

RQ 1a *What are the effects of human-likeness and lexical alignment on mutual gaze during an interaction?*

While previous research in HRI supports the effects of robot-related behaviours during an interaction [26], there are few studies implementing several behaviours at once, especially in cHRI, and measuring the effects through human behavioural changes. Our hypothesis for this research question is that human-likeness and lexical alignment will have an effect on the amount of mutual gaze during an interaction.

RQ 1b *Is there a correlation between mutual gaze and social trust and liking metrics?*

Looking at the current metrics in cHRI [37], there is a growing need to develop a standardised approach to identify and measure social trust and liking to ensure a broader application and improved replicability [22]. Our hypothesis for this research question is that the discrepancies in mutual gaze will align with the self-reported data.

RQ 2 *Can this correlation to mutual gaze be used as an objective measure of social trust and liking?*

Furthermore, we aim to identify another objective measure that can be utilised in this field by comparing the self-disclosure metric to identify and measure trust and liking to mutual gaze as a behavioural metric. The hypothesis for this research question is that mutual gaze can be used as an objective measure of social trust and liking, but may be limited in terms of how many contexts it can be applied to.

3. METHODS

3.1 Related Variables

Now that we have defined the questions we hope to answer in this study, we can look at the variables involved in this process. The independent variables that were manipulated are the human-likeness of the robot and lexical alignment to

3.3.2 Mutual Gaze

As we are working to identify fickle and minute behavioural data, we have chosen to utilise OpenFace [4] to identify the different parameters and dimensions involved in gaze angles. By using a machine learning algorithm to identify behavioural data, we seek to minimise human errors that may contaminate the data when identifying the occurrence and duration of mutual gaze. Also, we can improve the replicability of the current study by using a standardised approach, such as the machine learning algorithms provided in OpenFace.

The procedure of extracting the data begins with collecting the raw video data from the frontal camera, which includes the storytelling interaction and self-disclosure interaction of the experiment from the related study [12]. Using OpenFace, we then extracted all features related to gaze angles, head position, facial landmarks, and the facial action units (AUs) from the videos using the standard extraction command in OpenFace. Each video had both interaction contexts, which means that there is one CSV file per participant. From each of these files, we extracted fifteen columns of features which were relevant in identifying mutual gaze.

The first group of four features include [frame, timestamp, confidence, and success] and are used to identify frames, provide their exact time of occurrence in the video, and if OpenFace was able to successfully extract the features of a given frame. The second group of six features include [gaze_0_x, gaze_0_y, gaze_0_z, gaze_1_x, gaze_1_y, gaze_1_z] which indicate the x, y, and z values for the direction of the left eye gaze and right eye gaze, respectively. The following group of two features include [gaze_angle_x, gaze_angle_y] which are the x and y values for the average of the eye gaze angles for both eyes in radians. The final three features, [pose_Rx, pose_Ry, pose_Rz], are to identify the pitch, yaw, and roll of the head in radians. All values are relative to the frontal camera mounted on the touchscreen display as the origin. Other features, including the facial action units, landmarks, and pose measurements in millimetres, were excluded due to being outside of the scope of gaze angle identification and measurement during the interaction.

3.3.3 Mutual Gaze Threshold

In seeking a relation between the occurrences of mutual gaze and expressions of trust in the self-reported data, it is crucial to develop a consistent and replicable approach to defining and identifying mutual gaze in large amounts of data. As such, after removing unsuccessful frames and frames with nil confidence from the working data, the resulting data was used to generate a 2D heat map for each file using [gaze_angle_x] as the x-coordinates and [gaze_angle_y] as the y-coordinates. The heat maps were generated using numpy and the pyplot package from matplotlib. This was done to identify eye gaze angles that indicate the occurrence of mutual gaze during the two interactions for each participant. The values for [gaze_angle_x] and [gaze_angle_y] were aggregated into a separate file to generate an aggregated heat map which shows a generalised area of mutual gaze across all participants and interaction contexts, which provides a more consistent approximation and is more broadly applicable.

Following this, to validate the results of the aggregated heat map using the [gaze_angle_x] and [gaze_angle_y] coordinates, we individually mapped the x- and y-coordinates of the left and right eyes in two separate heat maps. Similar to the previous mapping, we look to the values [gaze_0_x] and [gaze_0_y] to map the gaze angles of the left eye onto a 2 dimensional heat map; for the right eye, the the x- and y-values are [gaze_1_x] and [gaze_1_y], respectively. The resulting aggregated heat maps for the left and right eyes were then compared to the heat map from the average of the two eyes (see Figure 3). In both the individual and averaged angles, there is a primary cluster that is centred on the origin. Taking into consideration that the camera was located on the touchscreen interface, this is most likely the collection of gaze angles where the children are looking at the interface. Apart from this, there is a secondary cluster which indicates the angles in which the child looked up at the robot (mutual gaze). This secondary cluster indicates one specific area of interest in both the individual and averaged gaze angle heat maps, which we accept to be the robot as there is no other visual constant present in the experiment setup. However, the secondary clusters from the heat maps do not overlap precisely, which means we must introduce a region that encompasses the two boxes as we cannot discern which boxed area is closest to the robot's eyes from the video data.

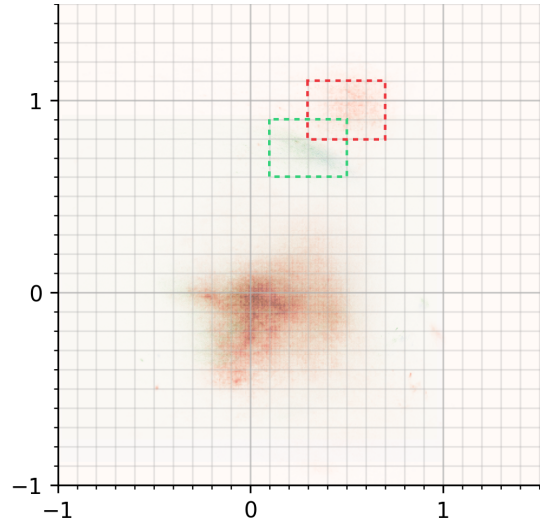


Figure 3. Layered heat map of averaged, left, and right eye gaze across all participants. The red box indicates the secondary cluster from [gaze_angle] values; the green box indicates the secondary cluster from [gaze_0] and [gaze_1] values

Drawing a boxed region to cover the areas of the red and green boxes may not capture the most instances of mutual gaze, especially if the front camera is shifted or the child moves, which is why we have opted to use a threshold approach where the angles should fall beyond $y = 0.4$, given that the experiment setup does not include other visual landmarks the child could look at behind or around the robot.

The threshold can be drawn starting at the top end of the main cluster that is where the touchscreen display was located, where any higher gaze can point to the child looking at the robot. By forming this threshold, we can find the frames in

which the children have mutual gaze with the robot, given that they fall within the bounds of this larger region (see Figure 4).

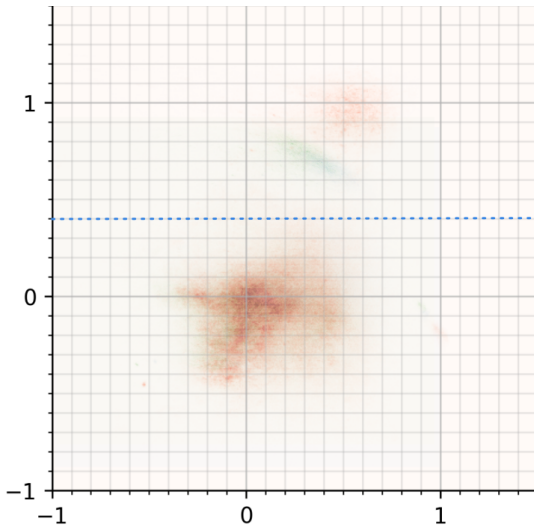


Figure 4. The blue line indicates the threshold ($y = 0.4$) between looking at the touchscreen display and looking at the robot (mutual gaze)

3.4 Data Analysis

3.4.1 Identifying and Extracting Frames with Mutual Gaze

Now that we have defined the region of the gaze angle that indicates mutual gaze, we can extract the frames that have their gaze angle values fall into that region. The three values bound by this region are $[gaze_angle_y, gaze_0_y, gaze_1_y]$. The values on the y-axis are set to be greater than the 0.4 threshold, since this is where the main cluster ends. These parameters are then applied to each of the filtered CSV files, which produces a file with only the rows of frames with mutual gaze. Excluding the files that have 0 frames with mutual gaze, we can then divide the frames that indicate mutual gaze by the total frames that were successfully identified by OpenFace and produce a ratio of mutual gaze per participant. Naturally, the files with no frames indicating mutual gaze are given a ratio of 0 percent.

At this point in the procedure, we split the combined interaction files into their individual interactions for each participant. Firstly, the combined interaction encompasses the entire duration the child was interacting with the robot. The mutual gaze in the combined interaction files were already identified and processed in the previous step.

The storytelling interaction only contains the task-based interaction with the robot and the self-disclosure interaction only contains the dialogue-based interaction with the robot. To separate these two interactions, we identified a timestamp between the first and second interaction, where the child is completing the post-interaction survey and interview. By converting these timestamps into seconds, we located the first frame with this timestamp and produced two separate CSV files. As per the procedural order of the experiment, the first half of the combined interaction was copied into the CSV file for the storytelling interaction, and the second half was copied into the CSV file for the self-disclosure interaction.

We now have three CSV files for each participant, one for the total combined interaction with the robot, one file for only the storytelling interaction, and then one file for only the self-disclosure interaction. To the storytelling and self-disclosure files, we applied the mutual gaze identification and extraction functions and produced a percentage ratio of mutual gaze for each interaction, per participant.

3.4.2 Validating the Frames with Mutual Gaze

The validation procedure for the frames that include mutual gaze is to manually view 5 randomly selected frames from each of the files to ensure mutual gaze is present. Files with less than 5 frames, but greater than 0 frames, with mutual gaze have all of the identified frames manually inspected for mutual gaze.

4. RESULTS

In this section, sections 4.1 and 4.2 cover the descriptive analyses and the statistical analyses involving mutual gaze as a dependant variable in the interactions. In these sections we are interested in understanding the effects of human-likeness and lexical alignment on mutual gaze (**RQ 1a**) and exploring the relationship between the social trust and liking metrics (**RQ 1b**). We evaluate mutual gaze as either per participant (storytelling and self-disclosure) or per interaction (storytelling or self-disclosure) in the context of these sections. Following this, section 4.3 compares the findings from both mutual gaze and self-disclosure as objective measures of trust to identify if mutual gaze can be similarly used in identifying and measuring social trust and liking (**RQ 2**). Finally, in section 4.4, we conduct exploratory analyses into two variables that may have a confounding effect on mutual gaze during different contexts; these variables are age and temporality of the interactions. The statistical software to analyse the interplay between the variables we will use is Jamovi, which is open-source and freely accessible [1].

4.1 Descriptive Analysis

Firstly, we conducted a descriptive analysis, including a Shapiro-Wilk test to understand the distribution of the data regarding the total frames, frames with mutual gaze, and the percentage of mutual gaze in the data from the combined interactions. The Shapiro-Wilk test showed that the data is not normally distributed for any dependent variable; based on this, we will be using non-parametric tests for the statistical data analyses.

Following this, when we compare the percentage of mutual gaze across the experimental conditions, there is a marked disparity in the average between the storytelling and self-disclosure sections. Going from storytelling to the self-disclosure sections, the HNA condition shows the greatest change while the MNA condition shows the least amount of change, with the changes in HA and MA falling in between these values.

When we examine the human-likeness by itself (HA/HNA and MA/MNA), we can see that there is a decrease in the average percentage of mutual gaze going from the human-like to machine-like conditions during the storytelling and self-disclosure sections. Additionally, looking at only lexical

alignment (HA/MA and HNA/MNA), we can also find a subtle decrease in the average percentage of mutual gaze in the storytelling and self-disclosure sections.

| | Storytelling Interaction | Self-disclosure Interaction | Δ |
|------------|--------------------------------------|--------------------------------------|----------|
| HA | 29.91% (SD = 20.97, Mdn = 30.73%) | 18.42% (SD = 17.71, Mdn = 15.30%) | 11.49 |
| HNA | 28.75% (SD = 20.57, Mdn = 25.83%) | 17.12% (SD = 17.20, Mdn = 14.10%) | 11.63 |
| MA | 21.46% (SD = 17.31, Mdn = 20.08%) | 11.79% (SD = 13.79, Mdn = 4.45%) | 9.67 |
| MNA | 17.36% (SD = 16.33, Mdn = 12.99%) | 8.68% (SD = 6.19, Mdn = 10.51%) | 8.68 |

Table 2. Percentage values of mutual gaze in each interaction and experimental condition

In a scatter plot of the percentage of mutual gaze across the four experimental conditions, we can see a trend of increasing values in the number of frames with mutual gaze as the duration of the interaction increases, in both combined and separated interaction.

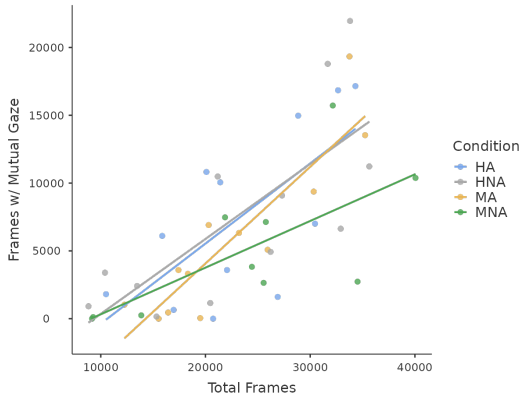


Figure 5. Scatter plot comparing the frames with mutual gaze to the total number of frames, with a regression line fitted to the average of the data points for the storytelling section

In both sections of the interaction, the machine-like and aligned condition (MA) produces the steepest slope while the machine-like and non-aligned condition (MNA) produces the flattest slope. In the story section, both conditions with lexical alignment (MA, HA) produce a steeper slope to the non-aligned conditions, which can indicate an effect of trust-worthiness that is developed by the lexical alignment; this is not the case for the self-disclosure section.

4.2 Associations with Mutual Gaze

4.2.1 Combined Interactions

Exploring the interactions between the experimental conditions and the average percent of mutual gaze in the combined interaction, we conducted a Kruskal-Wallis test. The null hypothesis for this test is that there is no difference in the median average percentage of mutual gaze across the experimental conditions. The results of the test suggest a

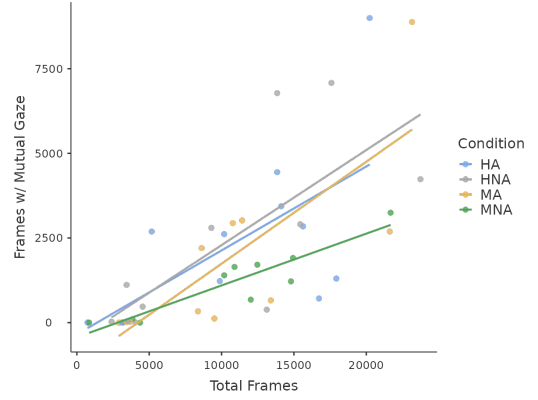


Figure 6. Scatter plot comparing the frames with mutual gaze to the total number of frames, with a regression line fitted to the average of the data points for the self-disclosure section

statistically insignificant acceptance of the null hypothesis ($H(3) = 3.56, p = 0.313$). What this signifies is that the median average percent of mutual gaze is identical across the experimental conditions for the combined interactions.

Additionally, we produced a Spearman correlation matrix to identify any relationships between the percent of mutual gaze and the social trust and liking scores taken from the post-interaction assessments. From this matrix, the comparisons were statistically insignificant, both for social trust ($\rho(44) = 0.175, p = 0.244$), and liking ($\rho(43) = 0.267, p = 0.077$) when compared to the average percent of mutual gaze in the combined interactions.

From these results, we can see that there is a need to delve deeper, to the level of the individual interactions, to identify the effects of mutual gaze in an interaction.

4.2.2 Storytelling Interaction

Looking at the storytelling interaction by itself, we conducted a Kruskal-Wallis test comparing the average percent of mutual gaze across the experimental conditions. The null hypothesis for this test is that there is no difference in the median average percentage of mutual gaze across the experimental conditions, as before. Results of the test are statistically insignificant, and suggest the acceptance of the null hypothesis ($H(3) = 2.94, p = 0.400$).

The Spearman correlation matrix that was produced to explore the association/relation between the percent of mutual gaze and the scores on social trust and liking exhibits a positive trend in all correlations. The results of this matrix are statistically insignificant for both the reported scores for social trust ($\rho(44) = 0.160, p = 0.290$) and liking ($\rho(43) = 0.219, p = 0.149$). Although, there seems to be greater relational strength between the percent of mutual gaze to the reported scores for liking than the correlation to the reported scores for social trust.

Interestingly, there is a strong positive correlation ($\rho(47) = 0.784, p < 0.001$) between the reported scores for social trust and liking, meaning that the children who tend to give higher scores for the questions pertaining to social trust (which

came first in the survey) also tend to give higher scores for the liking questions afterwards. Alongside this, there is also a statistically significant moderate positive correlation ($\rho(44) = 0.574, p = < 0.001$) between the amount of mutual gaze in the storytelling interaction and the self-disclosure interaction, which may indicate a personal tendency to engage in mutual gaze across interaction contexts.

| | Storytelling Interaction | | Self-disclosure Interaction | |
|--------------------|--------------------------|---------------|-----------------------------|---------------|
| | <i>Social Trust</i> | <i>Liking</i> | <i>Social Trust</i> | <i>Liking</i> |
| | <i>Scores</i> | <i>Scores</i> | <i>Scores</i> | <i>Scores</i> |
| Mutual Gaze | 0.160 | 0.219 | 0.176 | 0.214 |

*: $p = < 0.05$, **: $p = < 0.01$

Table 3. Spearman correlation matrix for the mutual gaze in each interaction

4.2.3 Self-disclosure Interaction

It is important to note that the participants of the experiment completed the post-interaction survey before the self-disclosure section, and were working on a collaborative task with the robot prior to the survey.

As per the previous sections, we conduct a Kruskal-Wallis test to determine if there are any differences between the median average percent of mutual gaze across the experimental conditions. The null hypothesis is, again, that there is no difference in median average percent of mutual gaze across the four experimental conditions. The results of the test are statistically insignificant and indicate towards accepting the null hypothesis ($H(3) = 1.82, p = 0.611$).

From the Spearman correlation matrix for this section of the interaction, we can see a similar outcome to the storytelling section, with two statistically insignificant weak positive correlations between the percentage of mutual gaze, the reported scores for social trust ($\rho(44) = 0.176, p = 0.241$), and the reported scores for liking ($\rho(43) = 0.214, p = 0.158$). There seems to be a stronger correlation between the percent of mutual gaze to the reported scores for liking than the reported scores for social trust, which is very similar to the storytelling section.

4.3 Comparing Objective Measures

Since the self-disclosure section is also a measure of trust in disclosing personal information, we compared the results of the disclosure section to the percent amount of mutual gaze during the interaction to assess how the findings may be interacting with one another. The disclosure elements were labelled as 'Disclosure-Good' for any ability the child was reportedly good at, and 'Disclosure-Bad' for any ability the child could identify a space for improvement. In a Spearman correlation test, both the percent of mutual gaze ($\rho(42) = -0.002, p = 0.990$) and 'Disclosure-Good' ($\rho(44) = 0.183, p = 0.223$) were found to have a statistically insignificant and weak correlation to 'Disclosure-Bad'. However, there is a statistically significant moderate negative correlation between the percent of mutual gaze in the interaction to 'Disclosure-Good' ($\rho(42) = -0.372, p = 0.013$). This can indicate that as positive self-disclosure increases, there is less

mutual gaze in the interaction; the content of the interaction is important in deciding to use mutual gaze as an objective measure.

4.4 Exploratory Analyses

4.4.1 Age

In exploring the age of the participants, we expect to gather an understanding of how the childhood developmental process may affect the amount mutual gaze in an interaction. In a scatter plot comparing the ages of the children to the amount of mutual gaze in the storytelling section, we can see a decline in mutual gaze as age increases; this is also apparent in a similar scatter plot comparing age to the percent of mutual gaze in the self-disclosure section.

Alongside this, a Spearman correlation matrix provides us with significant results concerning the percent of mutual gaze in each interaction. There is a weak negative correlation between age and the percent of mutual gaze during the storytelling section ($\rho(44) = -0.363, p = 0.013$). When we move onto the self-disclosure section, the correlational strength increases to a moderate correlational strength ($\rho(44) = -0.572, p = < 0.001$). From this, it is apparent that there is a significant relationship between mutual gaze and age, and that mutual gaze decreases as the age of the child increases.

| | Storytelling Mutual Gaze | Self-disclosure Mutual Gaze |
|------------|--------------------------|-----------------------------|
| Age | -0.363* | -0.572** |

*: $p = < 0.05$, **: $p = < 0.01$

Table 4. Spearman correlation matrix comparing age and mutual gaze during the two interactions

4.4.2 Temporal Effects

Looking at the temporal effects during the interactions, we might expect the novelty effect to play some role in changing the amount of mutual gaze between the beginning and the end of the interaction. For each interaction, we split them into two sections, the first half (1/2) and the second half (2/2) being roughly the same duration as each other. A Spearman correlation matrix shows statistically significant correlations between all four parts of the interactions.

Within the interactions, we can see significant and strongly positive correlations between each half in the storytelling ($\rho(44) = 0.734, p = < 0.001$) and self-disclosure interactions ($\rho(44) = 0.716, p = < 0.001$). When comparing across interactions, there is a moderate positive correlation between the first halves of the storytelling and self-disclosure sections ($\rho(44) = 0.488, p = < 0.001$), and another moderate correlation between the second halves of both interactions ($\rho(44) = 0.436, p = 0.002$). Furthermore, the moderate correlation between the first half of the storytelling section and the second half of the self-disclosure ($\rho(44) = 0.587, p = < 0.001$) is weaker than the moderate correlation between the second half of the storytelling section and the first half of the self-disclosure section ($\rho(44) = 0.620, p = < 0.001$). This may be indicative of some residual tendency to engage in mutual gaze carrying over to the following interaction.

| | | Self-Disclosure | |
|--------------|-------------|-----------------|-------------|
| | | First half | Second half |
| Storytelling | First half | 0.488** | 0.587** |
| | Second half | 0.620** | 0.436** |

*: $p = <0.05$, **: $p = <0.01$

Table 5. Spearman correlation of the average mutual gaze for the two halves of each interaction to study temporal effects

5. DISCUSSION

Now that we have analysed and explained the data, we can proceed to defining an answer for the research questions guiding this study.

5.1 Research Questions

Firstly, in **RQ 1a**, we explored the effects of human-likeness and lexical alignment on mutual gaze during an interaction. From the descriptive analyses, there is an indication towards an increased number of frames with mutual gaze over the duration of the interaction, where the MA condition shows the greatest increase in both the storytelling and self-disclosure interactions, compared to the other conditions. Furthermore, only in the storytelling interaction, the lexically aligned conditions (HA, MA) show a greater increase compared to the lexically non-aligned conditions. This emerging relationship between mutual gaze and lexical alignment also validates the previous findings in [27]. Apart from this, from the analyses for the combined and individual interactions, there were no statistically significant conclusions that can be drawn about the effects of the experimental conditions on the mutual gaze.

Additionally, in **RQ 1b**, we investigate the possibility of a correlation or association between the amount mutual gaze in an interaction and the social trust and liking metrics. While the combined interaction does not indicate anything of significance, the individual sections for the storytelling interaction and the self-disclosure interaction provide more information on this question. In these individual interactions, there are two significant findings; there is a moderate positive correlation between the amount of mutual gaze in the storytelling section and the self-disclosure section, and a positive correlation between the total social trust scores and the total liking scores. Interestingly, both of these measures interact only within themselves across different interaction contexts, instead of the expected relationship between measures; any interaction between mutual gaze and the social trust and liking scores is statistically insignificant and negligible. This may point to a lack of an association between mutual gaze and the scores, but more work needs to be done in this direction to definitively declare this.

In addition to this, it would be useful to state that the survey results all tend to collect towards the maximal end, which signals a ceiling effect that may be affecting the data concerning the social trust and liking scores. This is further discussed in the paper belonging to the related study [12], but for the intents and purposes of the current study, this may mean that some associations and correlations may not be detected.

Secondly, for **RQ 2**, we look into the use of mutual gaze as an objective measure of social trust and liking. We compared mutual gaze as a measure to the content of the self-disclosure interaction, which produced mainly insignificant results except for a statistically significant weak negative correlation between the amount of mutual gaze during the self-disclosure interaction and positive self-disclosure content. Succinctly put, as the amount of positive self-disclosure content increased, the amount of mutual gaze decreased. From [2], mutual gaze as a mechanism contributing to cognitive processing and conversational turn-taking is discussed, and also points out that gaze aversion is used to decrease discomfort and direct confrontation in a conversation. How this applies to the correlation we have found is by providing another dimension of mutual gaze that may be involved: social expectations. These can be avoiding staring, to not seem 'rude', or to hold the conversational turn while constructing an answer to avoid interrupting the other interlocutor later on. In the context of identifying the strength of mutual gaze as an objective measure of social trust and liking, exploring the effects of this dimension could provide powerful insights which may answer the question at hand.

5.2 Exploratory Analyses

Our strongest findings, perhaps, were from the exploratory analyses that investigated age and temporality as confounding variables in using mutual gaze as an objective measure.

For the interaction between the amount of mutual gaze and the ages of the participants, there is a significant negative correlation between these two variables. In the storytelling interaction, the correlational strength is weak, while in the self-disclosure interaction the correlational strength increases to a moderate degree. Overall, we see that the younger children tend to look at the robot more than the older children, and that this relationship is strengthened in the dialogue-based interaction. This finding is in direct opposition with the related finding in [5], which argues that younger children struggle to engage in and sustain eye contact compared with older children. The study in [5] is conducted on children between the ages of 3 and 4, whereas we have conducted this analysis on the behaviour of children between the ages of 7 and 10. This difference may be an explanation to the discrepancy between findings, however, there are also findings supporting a 'peak' in mutual gaze during an interaction, which may also explain this inconsistency. In [3], the 'peak' of eye contact during a dyadic interaction between an adult and child is established at kindergarten-age, with either side of this peak indicating a decreased amount of mutual gaze. With the findings in [5] and this study, there is the opportunity to explore the validity of the findings in [3] for applications in a human-robot interaction context.

Additionally, the increased social awareness stemming from childhood cognitive development can be attributed to the increased negative correlational strength during the self-disclosure section. Moving from task-related discussion to the disclosure of personal information necessarily changes the intimacy dynamics of the interaction, which means that the younger children may engage in more mutual gaze in response

to the increased cognitive load due to the increased intimacy during the interaction [14, 2].

Following this, we looked at how temporality may be affecting mutual gaze across the individual interactions and found that all of the findings generally indicate a moderate positive correlation between the various halves of the interactions. From this, the strongest correlational strength is between the second half of the storytelling interaction and the first half of the self-disclosure interaction. What this means that the amount of mutual gaze tends to 'carry over' from one interaction to the other, regardless of the change in interaction context from task-based to dialogue-based. Within the interactions themselves, the halves exhibit a strong positive correlation, which means that the amount of mutual gaze remains fairly stable throughout the interaction. Interestingly, these findings demonstrate that a change in the interaction context does not strongly impact the amount of mutual gaze exhibited during the interaction itself.

5.3 Design Implications

Taking into consideration the findings of this study, we propose three key takeaways that can help improve current practices in identifying and measuring social trust and liking in child-robot interaction. Firstly, when using mutual gaze as a measure of social trust and liking, the content and context seems to have a greater influence than the experimental conditions. By this, context can be described as the goal of the interaction, where an example would be that the interaction may be task-oriented or dialogue-oriented; content can be described by what is being evaluated during the interaction, where an example would be lexical content or self-disclosed skills. In comparing the survey scores related to social trust, liking, and mutual gaze, we found that these measures do not interact with mutual gaze as much as the context and content of the interactions. Alongside [12], we suggest high awareness of the context and content of the interaction if mutual gaze is used as a behavioural measure in an interaction. Following this, the findings in the exploratory analyses show us that outside factors such as age of the human-interlocutor and duration of the interaction, greatly influence mutual gaze. This means that researchers must be aware of these attributes of their participants and of the interaction when choosing mutual gaze as a behavioural metric.

Lastly, it has come to our attention that while mutual gaze as an objective measure may not fully capture social trust or liking, it seems to detect changes in intimacy during an interaction, which can be classified as a liking-related construct [9]. And so, our suggestion is to use several metrics, both intentional and unintentional, to capture constructs that may be intertwined with highly abstract notions like trust.

5.4 Limitations and Future Work

From the technical side, one major limitation we have encountered in this study is that our analysis of mutual gaze is highly dependent on irregular data; each participant had different behaviours when interacting with the robot, which may have affected the analysis quality since these behaviours cannot be normalised. Also, as we have discussed previously, the ceiling effects influencing the survey data may have affected the

evaluated relationship with mutual gaze. Apart from this, the ability to engage in mutual gaze can be affected by human-side factors such as visual impairments and relevant differences relating to the brain and nervous system. This greatly weakens the applicability of mutual gaze as a behavioural metric across a diverse population.

From this, we would like to promote future work towards other behavioural metrics that may transcend the aforementioned limitations. As we have discussed before, there are ample opportunities to explore missed correlations, other dimensions of social behaviour that may apply to mutual gaze, and the applicability of findings stemming from human-human interaction towards human-robot interaction.

6. CONCLUSION

From the findings of this project, we can identify the benefit of using mutual gaze as an objective measure for constructs related to social trust, namely for constructs relating to liking. For social trust itself, there is not sufficient evidence to conclude that using mutual gaze is better in identifying and measuring social trust over other objective measures of trust, such as self-disclosure, that were used in the related study. Among the findings related to mutual gaze, we have identified that mutual gaze is not significantly affected by the human-likeness or lexical alignment of the robot but is more affected by the context of the interaction, such as task-based (storytelling) versus dialogue-based (self-disclosure) interactions. The amount of mutual gaze is also affected by the content of the interaction during dialogue-based self-disclosure, where more positive self-disclosure resulted in significantly less mutual gaze; this association was not mirrored in any self-disclosure with negative content. In the exploratory analyses on age, we found that the amount of mutual gaze decreased as the age of the participants increased. Regarding temporal effects on mutual gaze, there is also evidence of engaging in similar amounts of mutual gaze at the end of one interaction and the start of the next interaction; 'carrying over' the amount of mutual gaze from one interaction to another. The amount of mutual gaze may change towards the 'ends' of the interactions, but appears to remain stable across times where the participant is not interacting with the robot..

Thus, we encourage further research into the multifaceted nature of social trust, liking, and intimacy, and developing objective measures to identify and capture these constructs through behavioural observations.

ACKNOWLEDGMENTS

I would like to express my deep gratitude and appreciation to my supervisors, colleagues, and friends for their discussion, guidance, and support throughout the entirety of this project.

REFERENCES

- [1] 2022. Jamovi - open statistical software for the desktop and cloud. <https://www.jamovi.org>. (2022). Accessed: 2023-4-23.
- [2] Sean Andrist, Xiang Zhi Tan, Michael Gleicher, and Bilge Mutlu. 2014. Conversational gaze aversion for humanlike robots. In *Proceedings of the 2014*

ACM/IEEE international conference on Human-robot interaction. 25–32.

- [3] Victor Ashear and John R Snortum. 1971. Eye contact in children as a function of age, sex, social and intellectual variables. *Developmental Psychology* 4, 3 (1971), 479.
- [4] Tadas Baltrušaitis, Peter Robinson, and Louis-Philippe Morency. 2016. Openface: an open source facial behavior analysis toolkit. In *2016 IEEE winter conference on applications of computer vision (WACV)*. IEEE, 1–10.
- [5] Peta Baxter, Chiara De Jong, Rian Aarts, Mirjam de Haas, and Paul Vogt. 2017. The effect of age on engagement in preschoolers’ child-robot interactions. In *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*. 81–82.
- [6] Tony Belpaeme, Paul Baxter, Joachim De Greeff, James Kennedy, Robin Read, Rosemarijn Looije, Mark Neerinx, Ilaria Baroni, and Mattia Coti Zelati. 2013. Child-robot interaction: Perspectives and challenges. In *Social Robotics: 5th International Conference, ICSR 2013, Bristol, UK, October 27-29, 2013, Proceedings 5*. Springer, 452–459.
- [7] Tony Belpaeme, Paul Baxter, Robin Read, Rachel Wood, Heriberto Cuayáhuítl, Bernd Kiefer, Stefania Racioppa, Ivana Kruijff-Korabayová, Georgios Athanasopoulos, Valentin Enescu, and others. 2012. Multimodal child-robot interaction: Building social bonds. *Journal of Human-Robot Interaction* 1, 2 (2012).
- [8] Deborah R Billings, Kristin E Schaefer, Jessie YC Chen, and Peter A Hancock. 2012. Human-robot interaction: developing trust in robots. In *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*. 109–110.
- [9] Jason Borenstein and Ronald Arkin. 2019. Robots, ethics, and intimacy: the need for scientific research. *On the Cognitive, Ethical, and Scientific Dimensions of Artificial Intelligence: Themes from IACAP 2016* (2019), 299–309.
- [10] Jürgen Brandstetter, Clay Beckner, Eduardo Benitez Sandoval, and Christoph Bartneck. 2017. Persistent lexical entrainment in HRI. In *Proceedings of the 2017 ACM/IEEE international conference on human-robot interaction*. 63–72.
- [11] Frank Broz, Hagen Lehmann, Yukiko Nakano, and Bilge Mutlu. 2012. Gaze in HRI: from modeling to communication. In *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*. 491–492.
- [12] Natalia Calvo-Barajas, Anastasia Akkuzu, and Ginevra Castellano. 2023. Balancing Human Likeness in Social Robots: Impact on Children’s Trust and Interaction in a Storytelling Context. (2023). unpublished.
- [13] Meia Chita-Tegmark, Theresa Law, Nicholas Rabb, and Matthias Scheutz. 2021. Can you trust your trust measure?. In *Proceedings of the 2021 ACM/IEEE international conference on human-robot interaction*. 92–100.
- [14] Lorenzo Desideri, Paola Bonifacci, Giulia Croati, Angelica Dalena, Maria Gesualdo, Gianfelice Molinaro, Arianna Gherardini, Lisa Cesario, and Cristina Ottaviani. 2021. The mind in the machine: Mind perception modulates gaze aversion during child–robot interaction. *International Journal of Social Robotics* 13 (2021), 599–614.
- [15] Friederike Eyssel. 2017. An experimental psychological perspective on social robotics. *Robotics and Autonomous Systems* 87 (2017), 363–371.
- [16] Oriel FeldmanHall and Amitai Shenhav. 2019. Resolving uncertainty in a social world. *Nature human behaviour* 3, 5 (2019), 426–435.
- [17] Peter A Hancock, Deborah R Billings, Kristin E Schaefer, Jessie YC Chen, Ewart J De Visser, and Raja Parasuraman. 2011. A meta-analysis of factors affecting trust in human-robot interaction. *Human factors* 53, 5 (2011), 517–527.
- [18] Marc Hulcelle, Giovanna Varni, Nicolas Rollet, and Chloé Clavel. 2021. TURIN: A coding system for Trust in hUman Robot INteraction. In *2021 9th International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, 1–8.
- [19] Ryo Ishii, Yuta Shinohara, T Nakano, and Toyoaki Nishida. 2011. Combining multiple types of eye-gaze information to predict user’s conversational engagement. In *2nd workshop on eye gaze on intelligent human machine interaction*.
- [20] Zahra Rezaei Khavas, S Reza Ahmadzadeh, and Paul Robinette. 2020. Modeling trust in human-robot interaction: A survey. In *Social Robotics: 12th International Conference, ICSR 2020, Golden, CO, USA, November 14–18, 2020, Proceedings 12*. Springer, 529–541.
- [21] Kyveli Kompatsiari, Vadim Tikhonoff, Francesca Ciardo, Giorgio Metta, and Agnieszka Wykowska. 2017. The importance of mutual gaze in human-robot interaction. In *Social Robotics: 9th International Conference, ICSR 2017, Tsukuba, Japan, November 22-24, 2017, Proceedings 9*. Springer, 443–452.
- [22] Jin Joo Lee, Brad Knox, Jolie Baumann, Cynthia Breazeal, and David DeSteno. 2013. Computationally modeling interpersonal trust. *Frontiers in psychology* (2013), 893.
- [23] Nikolas Martelaro, Victoria C Nneji, Wendy Ju, and Pamela Hinds. 2016. Tell me more designing hri to encourage more trust, disclosure, and companionship. In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 181–188.
- [24] Robert R McCrae and Paul T Costa Jr. 2008. The five-factor theory of personality. (2008).

- [25] Masahiro Mori, Karl F MacDorman, and Norri Kageki. 2012. The uncanny valley [from the field]. *IEEE Robotics & automation magazine* 19, 2 (2012), 98–100.
- [26] Manisha Natarajan and Matthew Gombolay. 2020. Effects of anthropomorphism and accountability on trust in human robot interaction. In *Proceedings of the 2020 ACM/IEEE international conference on human-robot interaction*. 33–42.
- [27] Bert Oben. 2018. Gaze as a predictor for lexical and gestural alignment. *Eyetracking in interaction: Studies on the role of eye gaze in dialogue* (2018), 233–262.
- [28] Julian B Rotter. 1967. A new scale for the measurement of interpersonal trust. *Journal of personality* (1967).
- [29] Nathan E Sanders and Chang S Nam. 2021. Applied quantitative models of trust in human-robot interaction. In *Trust in Human-Robot Interaction*. Elsevier, 449–476.
- [30] Giuseppina Schiavone, Domenico Formica, Fabrizio Taffoni, Domenico Campolo, Eugenio Guglielmelli, and Flavio Keller. 2011. Multimodal ecological technology: From child’s social behavior assessment to child-robot interaction improvement. *International Journal of Social Robotics* 3, 1 (2011), 69–81.
- [31] Mikey Siegel, Cynthia Breazeal, and Michael I Norton. 2009. Persuasive robotics: The influence of robot gender on human behavior. In *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2563–2568.
- [32] Rebecca Stower, Natalia Calvo-Barajas, Ginevra Castellano, and Arvid Kappas. 2021. A meta-analysis on children’s trust in social robots. *International Journal of Social Robotics* 13, 8 (2021), 1979–2001.
- [33] Rebecca Stower and Arvid Kappas. 2020. " Oh no, my instructions were wrong!" An Exploratory Pilot Towards Children’s Trust in Social Robots. In *2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 641–646.
- [34] Caroline L van Straten, Rinaldo Kühne, Jochen Peter, Chiara de Jong, and Alex Barco. 2020. Closeness, trust, and perceived social support in child-robot relationship formation: Development and validation of three self-report scales. *Interaction Studies* 21, 1 (2020), 57–84.
- [35] Adriana Tapus and Maja J Mataric. 2008. Socially Assistive Robots: The Link between Personality, Empathy, Physiological Signals, and Task Performance.. In *AAAI spring symposium: emotion, personality, and social behavior*. 133–140.
- [36] Benedict Tay, Younbo Jung, and Taezoon Park. 2014. When stereotypes meet robots: the double-edge sword of robot gender and personality in human–robot interaction. *Computers in Human Behavior* 38 (2014), 75–84.
- [37] Caroline L van Straten, Jochen Peter, and Rinaldo Kühne. 2020. Child–robot relationship formation: A narrative review of empirical research. *International Journal of Social Robotics* 12 (2020), 325–344.
- [38] Alan R Wagner, Paul Robinette, and Ayanna Howard. 2018. Modeling the human-robot trust phenomenon: A conceptual framework based on risk. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 8, 4 (2018), 1–24.
- [39] Sarah Woods, Kerstin Dautenhahn, Christina Kaouri, Renete Boekhorst, and Kheng Lee Koay. 2005. Is this robot like me? Links between human and robot personality traits. In *5th IEEE-RAS International Conference on Humanoid Robots, 2005*. IEEE, 375–380.