

The Effects of Robot's Facial Expressions on Children's First Impressions of Trustworthiness

Natalia Calvo-Barajas¹, Giulia Perugia¹, Ginevra Castellano¹

Abstract—Facial expressions of emotions influence the perception of robots in first encounters. People can judge trustworthiness, likability, and aggressiveness in a few milliseconds by simply observing other individuals' faces. While first impressions have been extensively studied in adult-robot interaction, they have been addressed in child-robot interaction only rarely. This knowledge is crucial, as the first impression children build of robots might influence their willingness to interact with them over extended periods of time, for example in applications where robots play the role of companions or tutors. The present study focuses on investigating the effects of facial expressions of emotions on children's perceptions of trust towards robots during first encounters. We constructed a set of facial expressions of happiness and anger varying in terms of intensity. We implemented these facial expressions onto a Furhat robot that was either male-like or female-like. 129 children were exposed to the robot's expressions for a few seconds. We asked them to evaluate the robot in terms of trustworthiness, likability, and competence and investigated how emotion type, emotion intensity, and gender-likeness affected the perception of the robot. Results showed that a few seconds are enough for children to make a trait inference based on the robot's emotion. We observed that emotion type, emotion intensity, and gender-likeness did not directly affect trust, but the perception of likability and competence of the robot served as facilitator to judge trustworthiness.

I. INTRODUCTION

Social robots are increasingly used as children's companions in schools and homes. This motivates researchers and developers across different fields to design and build robots that can interact with humans in a friendly and, at the same time, safe way. One of the robot-related features associated with social acceptance is the robot's capability to express emotions through body gestures or facial cues [1][2]. In Child-Robot Interaction (cHRI), facial expressions of emotions can influence children's judgments of trust, likability, and their attitudes towards robotic systems [1][3][4].

Facial expressions of emotion are evaluated from the moment we are exposed to a stranger's face for the first time [5]. This first impression gives us information on whether that person can be trusted or not [6][7]. In fact, people judge trustworthiness, competence, likability, aggressiveness, and attractiveness in a few milliseconds [8]. This inference is crucial as emotions are used to convey interpersonal information such as feelings, intentions, and non-verbal messages. In Human-Robot Interaction (HRI), trait inference helps adult

users in deciding whether a robot is to be trusted or not, and whether it is the case to start an interaction with it. While there is some evidence that facial cues (e.g., eyebrows, lips) can affect trust behavior and social cognition in young children [9], there is little evidence that this could extend to facial expressions [1].

The present study focuses on investigating the effect of facial expressions of emotions on children's perceptions of trust towards robots during first encounters. We exposed children to facial expressions of happiness and anger varying in intensity (i.e., low, medium, high). The facial expressions were displayed by a humanoid robot that was either female-like or male-like in appearance. We asked children to rate their perception of the robot in terms of trustworthiness, likability, and competence. We found out that the robot's facial expressions of emotion have an effect on children's impression formation. This first impression might guide children to understand the robot's intentions and therefore influence cHRI.

II. RELATED WORK

A. Education and Social Robots

In the educational field, social robots are used as companions and tutors to support children's learning in various subjects, such as second language acquisition, math, and computational thinking [4][10][11]. In a recent literature review, Belpaeme and colleagues described the benefits of robots in increasing children's cognitive (i.e., knowledge, comprehension, application, analysis, synthesis, and evaluation) and affective learning outcomes (i.e., attention, receptiveness, responsiveness, and reflectiveness) across different domains [12].

Studies in Social Psychology confirmed that teacher's non-verbal behavior affects attitudes and perceptions of students and hence their learning [13]. As non-verbal behavior contributes to building a learner-teacher trustworthy communication, it is crucial to know which kind of cues learners use to evaluate others. The understanding of child-robot interaction (cHRI) in education includes not only the examination of children's social interaction, acceptance, and emotional involvement with a robot [14], but also the impact of the robot's behavior and its physical-related features on the interaction itself [1][11]. We can assume that the first encounter with a robot is crucial to set a communication channel and therefore helps in making initial judgments of robots [11]. These initial judgments are based on the robot's type and appearance and persist over time affecting the long-term social acceptance of the robot [1][15]. For instance, the

*This work was supported by the European Commission Horizon 2020 Research and Innovation Program under Grant Agreement No. 765955.

¹Authors are with the Department of Information Technology, Uppsala University, 75237 Uppsala, Sweden. [natalia.calvo, giulia.perugia, ginevra.castellano@it.uu.se]

judgment of trustworthiness guides the decision of a child in choosing a robot teammate to complete a task [4] or serves as an indicator of conformity in the evaluation of a robot's performance [16]. Broadly, a first impression affects children's comfort and rapport with the robot and, consequently, trust development and learning outcomes [14][12].

B. First Impressions, Trust, and Emotions

Trust is a key factor that people evaluate when meeting someone for the first time. When the first impression relies on facial appearance, the perceived trust serves as an indicator to know whether to approach or avoid a person [8]. There is evidence that 100 ms are sufficient to judge someone's attractiveness, likability, trustworthiness, competence, and aggressiveness [8]. In this line, studies have demonstrated that expressions of emotion convey information about the affective state and interpersonal intent of a subject and are correlated with the judgment of trust and dominance [17]. For instance, in first encounters, happy faces are perceived as more trustworthy than faces expressing anger [6]. Children make trait inferences about others' facial appearance to guide their trust perception and behavior [9][18]. Ma et al. studied how young children's trust perception is influenced by the emotions expressed by potential informants [19]. They found that 4 and 5 year-old children trust more individuals expressing positive emotions.

Nevertheless, trust is challenging to define and measure. Psychology showed that trust is a multidimensional construct that can be measured along two main dimensions: affect- and cognition-based trust. The first relates to interpersonal trust, looking at benevolence, interpersonal care, and perceived warmth. The second focuses on perceived competence and reliability [20]. Similar to Psychology, also in cHRI, trust can be defined along two components *social trust*, relating to the "*belief that the robot will keep its word or promises*", and *competency trust*, defined as "*perceived competency and reliability of the robot*" [21]. There is evidence that children rely on the perceptions of competence and benevolence to determine whom to trust [22]. *If positive emotions are able to elicit higher competence and benevolence, then robots that display them might be perceived as more trustworthy.* Besides, several factors influence how children trust a robot. For instance, attribute factors — robot personality, expressiveness, embodiment, and anthropomorphism— have a greater impact on social trust, whereas performance factors —robot behavior, failure rates, interaction length— contributes more to the development of competency trust [21][2].

Non-verbal emotional expressions can elicit trust and closeness in the relationship between children and robots [11][4]. For instance, Tielman et al. [23] found that children reacted more positively towards an adaptive emotional expressive Nao robot than a flat Nao robot. Another study demonstrated that an iCat robot that expressed emotions was more effective in achieving a certain goal [24]. However, most of the studies found in the literature used robots that convey an emotional state based on gestures and limited facial movements [1][2][21]. Only a few numbers of cHRI

studies have used a human-like embodiment with the capability of exhibiting human-like facial expression of emotions [25][26][27]. These studies focused on addressing the impact of robot characteristics on emotional responsiveness rather than on trust perception. Therefore, *it is yet to understand whether the facial expression of a human-like robot might influence how trustworthy it is perceived.*

Literature suggests that people can better judge emotions that are high in intensity [28]. In [5], users could infer the emotions of the iCat robot and rate their intensity. This result suggested that users are more prone to detect the exaggerated expression of emotion in a robot. However, *there is not enough evidence that shows how humans, especially children, can make a first impression judgment based on the robot's emotional intensity level.*

C. The Role of Robot's Gender

Gender-likeness has been largely studied in HRI to understand how people respond to robots [29]. For instance, Siegel et al. found that humans perceived a robot of the opposite sex as more credible, trustworthy, and engaging. Also, they demonstrated that men are more likely to be persuaded by a female-like robot [30]. Other studies have been dedicated to assess the influence of gender-likeness on the perception of trust in robots. In [12], the upper-body shape of a robot was manipulated to attribute the robot a gender. The authors investigated trust ratings of female- and male-like robots and found that a female-like robot evoked more affective and cognitive trust.

Although the above studies demonstrated that people make different trait inferences based on the robot's gender-likeness, they all argued that such effect was present when accompanied by a gender stereotypical task. For instance, when a robot was perceived as agentic, participants rated it suitable for stereotypically male tasks, whereas, when a robot was rated as more communal, participants preferred it for stereotypically female tasks. [31]. To the best of our knowledge, *no study has investigated with children whether gender-likeness mediates the perception of trustworthiness of a robot expressing emotions in the first encounter.*

III. RESEARCH QUESTIONS

As we found out, children are able to make judgments of trustworthiness after brief exposures to faces and these judgments affect intentions and behaviors towards the other person [9]. Happiness signals that the observed person can be approached, anger that s/he should be avoided. However, the effect that these two emotions have on the perception of trust has not been investigated in cHRI. We hence pose the following research question (RQ):

RQ1: Does the type of emotion displayed by a robot affect children's perception of trustworthiness, likability, and competence?

Following up on the evaluation of the robot's facial expressions of emotion and emphasizing the lack of evidence on how intensity level might affect children's first impression of a robot, we aim to investigate the following RQ:

RQ2: Does the intensity level of happiness and anger influence children’s perception of trustworthiness, likability, and competence of a robot?

As there is evidence that female-gendered robots are perceived as significantly more likable and trustworthy than male-gendered robots in adults [30], we want to understand if gender-likeness affects perception of trustworthiness also in cHRI, specifically in first encounters. Hence, we address the following RQ:

RQ3: Does the gender-likeness of the robot affect children’s perception of trust, likability and competence?

Following findings from Psychology and cHRI, we hypothesize that:

H1: Children perceive the robot that expresses happiness as more trustworthy, likable, and competent than the one that expresses anger.

H2: Children perceive the robot that expresses higher levels of happiness as more trustworthy, likable, and competent than the one that expresses lower levels of happiness, and the robot that expresses higher levels of anger as less trustworthy, likable, and competent than the one that expresses lower levels of anger.

H3: Children perceive female-like robots as more trustworthy, likable and competent than male-like robots.

IV. METHODOLOGY

We constructed a user study with a 2x2x3 mixed design with Robot Gender (Female-like/Male-like) as within-subject factor, and Emotion Type (Happiness/Anger) and Intensity Level (Low/Medium/High) as between-subject factors.

A. Measures

Social trust was captured by measuring the likability in terms of appearance and friendliness [32] (Table 1). As competency trust is usually assessed when an interaction is present, we could not measure it in this study [10]. To investigate how children judge the perceived intelligence of the robot, and hence its competence, we instead measured smartness and helpfulness. Regarding trust, we took inspiration from the methods presented in [21]. We selected four items —secrets, truth-teller, influence, and trust-core— that contribute to understanding the way children perceived the robot’s social reliability and their conceptualization of trusting a robot. We implemented a modified version of the Godspeed questionnaire [33] for Likability and Competence. The 8 items presented in Table 1 are measured on a 5-point Likert scale.

TABLE 1: Questionnaire and Dependent Variables

Question	Item Measured	Dependent Variable
Do you like the appearance of the robot?	Appearance	Likability
Do you think the robot can be friendly?	Friendly	Likability
Do you think the robot can help others?	Helpfulness	Competence
Do you think the robot is smart?	Smartness	Competence
Do you think the robot can keep secrets?	Secrets	Trust
Do you think the robot is a truth teller?	Truth-teller	Trust
Can the robot influence others behavior?	Influence	Trust
Do you think the robot is trustworthy?	Trust-Core	Trust

B. Participants

We conducted the study at a science festival for children held in Uppsala (Sweden) called Scifest. Children who visited the Uppsala Social Robotics Lab booth were asked to participate in the experiment. We collected data in two of the three days of the festival. We did not collect personal data from the children’s interaction with the robots, but only asked them to fill out the questionnaires in Swedish. A total of 155 children participated in this study. We discarded data from 26 participants for reasons such as dropping the activity, having already encountered the robot, or incomplete questionnaires. The questionnaires from 129 children (46 female, 56 male, 27 No Answer) were considered in our analysis. They ranged in age from 9 to 14 years ($M = 11.29, SD = 0.85$). Table 2 shows the number of valid responses per stimuli.

TABLE 2: Valid number of samples per Stimulus

Gesture Stimulus	Male-Stimulus		Female Stimulus	
	Frequency	%	Frequency	%
High happiness	14	12.6	17	15.3
Medium happiness	18	16.2	25	22.5
Low happiness	16	14.4	19	17.1
High anger	16	14.4	19	17.1
Medium anger	21	18.9	14	12.6
Low anger	19	17.1	23	20.7

C. Apparatus and Stimuli

We designed a set of dynamic facial expressions to be projected onto the Furhat robot [34]. Furhat is a human-like robotic head equipped with a rigid mask in which a facial texture is projected from within. The robot has three degrees of freedom that allow it to orientate its head in the 3D space. As Furhat’s face can be fully animated and customized, we could easily manipulate the Action Units (AU) denoting happiness and anger and create female-like and male-like textures for the robot (Fig.1).

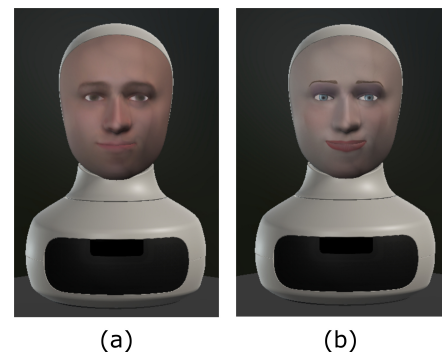


Fig. 1: (a) Male-like robot. (b) Female-like robot

Oosterhof and Todorov proposed a computational model to describe facial expressions [17]. They modified two dimensions —valence and dominance— to create a set of faces that elicited approach/avoidance and strength/weakness judgements. These set of faces are equivalent to happiness

and anger at different intensity levels [6]. We refer to their approach in this study. We used happiness and anger to signal approach or avoidance, and the intensity of happiness and anger to suggest the strength/weakness of the emotion (Fig.2).

When Ekman defined the set of six basic expressions of emotion (anger, disgust, surprise, fear, happiness, sadness), he also introduced the concept of Actions Units (AUs) for the measurement of facial movements [35]. We manipulated the AUs suggested by Ekman to animate the mouth and eyebrows regions of Furhat and create dynamic facial expressions of happiness and anger [36]. To give the impression of a change in intensity, we modified the AU activation in the apex in agreement with the Facial Action Coding System (FACS) [35]. We used an in-built domain-specific-language provided by Furhat in the Kotlin programming language to create the set of stimuli.

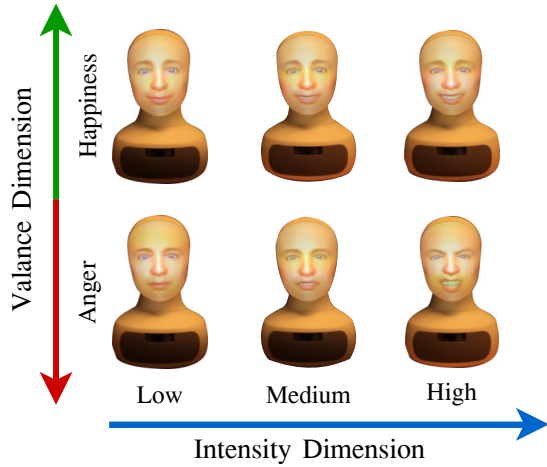


Fig. 2: Exemplar of male stimuli projected onto Furhat robot

We implemented the two emotions of happiness and anger and their levels of intensity on the female-like and male-like default textures of the Furhat platform (Fig.1).

1) *Validating the set of stimuli:* To validate the proposed set of stimuli, we conducted a study with a 2x2x3 within-subject design through Amazon Mechanical Turk. Users were exposed to the virtual version of the robot and watched snap videos of the designed facial expression lasting 5s per stimulus. For each expression, participants were asked to select which emotion was displayed and the level of intensity they perceived. We investigated if participants could correctly: 1) identify the level of emotional intensity (high/medium/low); and 2) recognize the emotion of the facial expression (anger/happiness). 42 participants (13 female, 29 male), ranging in age from 25 to 56 years ($M = 32.52, SD = 6.76$), took part in the validation study. 35 of them passed all the attention checks, only their responses were considered in the results.

We built a confusion matrix to present the goodness of the emotion and intensity recognition. In Figure 3.a-c: the rows indicate the label assigned by the participant and the columns indicate the actual label of the stimuli. Results revealed that

	Happiness	Anger	Surprise	Sadness	Disgust	Fear
Happiness	92	0	4	2	2	1
Anger	1	53	8	4	21	1

(a) Emotion Recognition

	High	Medium	Low
High	47	31	21
Medium	16	51	33
Low	14	38	46

(b) Intensity of Happiness

	High	Medium	Low
High	76	20	4
Medium	20	44	36
Low	14	35	51

(c) Intensity of Anger

Fig. 3: (a) Confusion matrix emotion recognition. (b) Confusion matrix intensity level of Happiness. (c) Confusion matrix intensity level of Anger

happiness was correctly identified by almost all participants [92%], while anger was more difficult to identify [53%], as participants mostly confused it with disgust [21%] (Figure 3.a). This result is similar to previous studies that showed that anger, disgust, and fear are more difficult to differentiate [37]. Concerning intensity, we found that high intensity of anger was the easiest to recognize [76%], followed by low anger and medium happiness [51%], and high happiness [47%] (Figures 3.b and 3.c). Overall, *the type of emotion and level of intensity were identified by the participants*. Considering the human-level recognition rates, our set of stimuli was considered valid for the experiment.

D. Experimental Setup and Procedure

We conducted the user study in two of the three days of a science fair. During data collection, we isolated and separated the experimental space from the other parts of the exhibition using curtains. We allowed two children to participate in the experiment at the same time, they were instructed not to talk to each other during the experiment. The robot was covered by a blanket to have genuine first impressions. The experimenter informed the participants that they were going to see a robot displaying two facial expressions, but did not give information about gender nor emotion type. Then, the experimenter uncovered the robot. At this point, the first stimulus was displayed for five seconds, and then the robot was covered again, so the children could fill in the questionnaire on first impressions through a tablet. The experimenter repeated the same process for the second stimulus. Each participant was asked to rate one facial expression for female-like and one for male-like stimuli. Children received candies at the end of the activity as a compensation for their time. The allocation of children to experimental conditions was randomized.

The experimenter took notes to collect information about the study and be able to exclude participants that did not meet the inclusion criteria at a later point (e.g., children that were exposed to the robot before the blanket was removed).

Also, she wrote down children's reactions to the robot when this was uncovered.

V. RESULTS

We employed a 2X2X3 MANOVA with Robot Gender (male-like/female-like) as within-subjects independent variable, and Emotion Type (happiness/anger) and Intensity Level (low/medium/high) as between-subjects independent variables. All p -values of the post-hoc analyses are adjusted with a Bonferroni correction. As the participation in the experiment was voluntary, children could drop the activity at any time. Because of this, each stimulus had a slightly different sample of participants (Table 2).

A. Emotion Type

RQ1 is dedicated to investigating the influence of happiness and anger on the judgment of trustworthiness, likability, and competence of the robot. The results of the MANOVA showed a significant main effect of Emotion Type ($F(8,191) = 2.869, p = .005$) on the dependent variables. Post-hoc tests revealed that children perceived the robot as significantly more likable in terms of *appearance* ($p = .001$) if it expressed happiness ($M = 3.38, SD = 1.01$) instead of anger ($M = 2.89, SD = 1.14$), and, in terms of *friendliness* ($p = .023$) if it expressed happiness ($M = 3.36, SD = 1.04$) instead of anger ($M = 2.98, SD = 1.21$). We also found that the robot was perceived as significantly more competent in terms of *smartness* ($p = .002$) if it expressed happiness ($M = 4.12, SD = .95$) instead of anger ($M = 3.64, SD = 1.25$). However, we did not find a significant effect of Emotion Type on the perception of trustworthiness. We can conclude that, in agreement with H1, *children perceived the robot that expressed happiness as more likable and competent than the one expressing anger*. However, this result did not extend to the perception of trustworthiness.

B. Intensity Level

RQ2 is devoted to investigating the influence of high, medium, and low levels of intensity on judgment of trustworthiness, likability, and competence of the robot for both happiness and anger. The MANOVA revealed that there was a significant main effect of the Intensity Level ($F(40,935) = 1.525, p = .021$) on the dependent variables. Post-hoc tests revealed that children perceived the robot as more likable in terms of *appearance* when it expressed low happiness ($M = 3.65, SD = 0.71$) instead of high anger ($M = 2.76, SD = 1.4, p = .007$) and low anger ($M = 2.92, SD = .98, p = .037$). Also the robot was perceived as more likable in terms of *friendliness* when it expressed low happiness ($M = 3.68, SD = .91$) instead of medium happiness ($M = 2.73, SD = 1.04, p = .011$), and high anger ($M = 2.67, SD = 1.59, p = .003$). Besides, the robot was perceived as more competent in terms of *smartness* when it expressed high happiness ($M = 4.36, SD = .82$) instead of high anger ($M = 3.48, SD = 1.17, p = .038$). Also in this case, we did not find a significant effect of Intensity Level on the perception of trustworthiness. We can conclude that H2 cannot be accepted

in its entirety. Indeed, *children perceived the robot that expressed lower level of happiness, instead of high levels of happiness, as more likable*. However, *they preferred the robot expressing high happiness to the one expressing high anger*. Similar to emotion type, these results do not extend to the judgement of trustworthiness.

C. Gender-Likeness

RQ3 investigates the effect of male-like and female-like stimuli on the judgment of trustworthiness, likability, and competence. Results revealed no significant main effect of Gender-likeness ($F(8,183) = 1.275, p = .259$) on the dependent variables, nor a significant interaction effect of Gender-likeness and Emotion Type ($F(8,191) = 1.252, p = .271$). However, they did highlight a significant interaction effect of Gender-likeness and emotion Intensity Level ($F(40,935) = 1.453, p = .036$). The interaction effect revealed that the male-like robot was perceived as more likable in terms of *appearance*, when it expressed high anger ($M = 3.62, SD = 1.4$) instead of medium anger ($M = 2.89, SD = .87, p = .037$), and when it expressed low happiness ($M = 3.75, SD = .57$) instead of medium anger ($M = 2.89, SD = .87, p = .015$), and low anger ($M = 3, SD = 1, p = .037$). We also found that the female-like robot was perceived as less likable in terms of *appearance*, when the robot expressed high anger ($M = 1.94, SD = .89$) instead of: medium anger ($M = 3.15, SD = 1.14, p = .002$), low anger ($M = 2.85, SD = 0.98, p = .008$), high happiness ($M = 3.55, SD = 1.03, p < .001$), medium happiness ($M = 2.88, SD = 1.2, p = .010$), and low happiness ($M = 3.57, SD = .81, p < .001$); also, when the robot expressed low anger ($p = .026$), and medium happiness ($p = .042$), instead of low happiness. Therefore, we can conclude that H3 is rejected. Indeed, *children did not perceive the female-like robot as more trustworthy, likable, and competent than the male-like robot*. Interestingly, however, they perceived the female-like robot as *less* likable when it expressed strong negative emotions such as high anger, whereas they perceived the male-like robot as *more* likable when displaying the same type of negative emotion. Similar to emotion type and intensity, also gender-likeness did not show a significant effect on trustworthiness.

Our results did not show an effect on children's trust in the robot. However, we presume that the robot's facial expression do not affect trustworthiness directly, but they rather do so indirectly, through the mediation of likability and competence. To test this claim, we performed four regression analyses with the items of Likability – appearance and friendliness – and the items of Competence – Helpfulness, and Smartness – as predictor variables, and the four items of Trustworthiness – Secrets, Truth-teller, Influence, Trust Core – as dependent variables (see Table 1). We found out that:

- The perceived *friendliness* ($\beta = .236, t(202) = 3.223, p = .001$) and *smartness* ($\beta = .320, t(202) = 4.543, p < .001$) of the robot were predictors of whether the robot was perceived as being able to keep *secrets*;
- The perceived *helpfulness* ($\beta = .203, t(205) = 3.108, p = .002$) and *smartness* ($\beta = .426, t(205) =$

6.425, $p < .001$) of the robot were predictors of whether the robot was perceived as a *truth-teller*;

- The perceived *helpfulness* ($\beta = .452$, $t(203) = 6.351$, $p < .001$) of the robot was a predictor of whether the robot was perceived as being able to *influence* others;
- The perceived *helpfulness* ($\beta = .314$, $t(202) = 4.580$, $p < .001$) and *smartness* ($\beta = .290$, $t(202) = 4.147$, $p < .001$) of the robot were predictors of how trustworthy the robot was perceived (i.e., *trust core*).

We can hence state that the *perceived likability and competence of a robot influence children's judgement of its trustworthiness in first encounters*.

VI. DISCUSSION

A. The Effect of Emotion Type (RQ1)

We found that *facial expressions of happiness increase the perceived likability and competence of a robot in child-robot first encounters*. When children met the robot for the first time, they liked more its appearance when it expressed happiness instead of anger and consequently rated the robot as more friendly and smarter. We found no significant difference in evaluations of perceived trustworthiness for any of the four items measured. This result suggests that children cannot judge trust directly from robots' facial expressions as they do with humans [9][38], a claim that is supported by HRI studies that argued that humans are better at attributing emotional states to fellow humans instead of robots [37]. Further analyses that we conducted suggest that children might not judge trustworthiness directly from facial expressions, but rather from the impression of likability and competence that the robot elicits.

We consider important to point out that the majority of the studies from Psychology asked participants to rate the trustworthiness of faces on a single item (i.e., "Do you think the face is trustworthy or untrustworthy?") without considering the multifaceted property of trust like we did [32]. This might have affected the way children assessed trust as the definition of trust can differ among individuals, and especially among children, who use multiple sources of information to make an evaluation of trustworthiness.

B. The Effect of Intensity Level (RQ2)

Concerning the intensity level of emotion, we found that *it affects the perception of likability and competence in child-robot first encounters*. In general, low levels of happiness increased robot's perceived likability relative to the different levels of anger. However, an unexpected finding was that children preferred a robot that expressed low happiness over high happiness. One possible explanation for this result is that low happiness could be perceived as more authentic, and hence elicit higher social acceptance. While low levels of happiness were perceived as more likable, high level of happiness contributed significantly to the robot's perceived smartness. This suggests that children can better judge competence when the robot expresses emotions with high intensity. Our findings reveal that for children the intensity level of an emotion is not a strong cue to directly make a trait

inference of trustworthiness. Similar to emotion type, we presume that it is specifically the perception of competence and likability elicited by happiness and anger at each intensity level that mediates the perception of trustworthiness.

C. The Effect of Gender of the Robot (RQ3)

We found that *the gender of the robot alone did not influence the perception of the robot in terms of trustworthiness, likability, and competence*. However, the interaction between gender and emotion intensity greatly affected children's perception of the robot's likability. We would like to highlight an interesting result: the robot expressing high anger was perceived as extremely likable when male-like and extremely dislikable when it was female-like. This finding could point to gender stereotypes related to the expression of anger. Marsh et al. suggested that anger is associated with maturity, strength, dominance, and masculinity [39]. It might hence be that this emotion was perceived as too dominant on a female-like robot and hence disliked.

Overall, our results suggest that 5 seconds are enough for children to judge the likability and competence of a social robot that displays facial expressions at different levels of intensity but they are not sufficient to directly judge trust. At this level of processing, trust is rather inferred from initial impressions of competence and likability. Also, our findings suggest that mostly the perception of competence affects trust judgments in first encounters. This result is in line with Peatzel et. al. as it confirms that competence is judged immediately in first encounters, and is kept constant during repeated interaction [15].

The main limitation of our study resides in the setting where it was carried out. Field studies are extremely valuable as they safeguard ecological validity. However, especially when carried out in fuzzy environments like science fairs, they make it difficult to control for confounders. Another potential limitation of the study is Furhat's blended robot embodiment. We noticed that it had a strong influence already at baseline. Indeed, during the experiment, when we uncovered the robot, we noticed that children were extremely surprised by the robot's appearance. This might have partially hindered their reactions to the first stimulus. Also, a third limitation is participants met the robot in couples, the presence of another child might have partially interfered with participants' ratings.

VII. CONCLUSIONS

In this study, participants were exposed for a few seconds to a female- and male-like version of the Furhat robot exhibiting emotions of happiness and anger at different intensity levels. We found that few seconds were enough for children to judge likability and competence, and that the perception of trust was not directly judged from facial expressions, but rather mediated by children's perceptions of the robot's likability and competence. Our results stand out as they show that the robot's facial expressions might impact children's trait inference already in first encounters. This is extremely important when robots are used as companions

or peers in children's education, as it might influence their long-term acceptance of robots in the classroom.

ACKNOWLEDGMENT

This work was supported by the European Commission Horizon 2020 Research and Innovation Program under Grant Agreement No. 765955.

REFERENCES

- [1] I. Cohen, R. Looije, and M. A. Neerincx, "Child's perception of robot's emotions: effects of platform, context and experience," *International Journal of Social Robotics*, vol. 6, no. 4, pp. 507–518, 2014.
- [2] C. L. van Straten, J. Peter, and R. Kühne, "Child-robot relationship formation: A narrative review of empirical research," *International Journal of Social Robotics*, pp. 1–20, 2019.
- [3] A. Hamacher, N. Bianchi-Berthouze, A. G. Pipe, and K. Eder, "Believing in bert: Using expressive communication to enhance trust and counteract operational error in physical human-robot interaction," in *2016 25th IEEE international symposium on robot and human interactive communication (RO-MAN)*. IEEE, 2016, pp. 493–500.
- [4] P. Alves-Oliveira, P. Sequeira, F. S. Melo, G. Castellano, and A. Paiva, "Empathic robot for group learning: A field study," *ACM Transactions on Human-Robot Interaction (THRI)*, vol. 8, no. 1, p. 3, 2019.
- [5] C. Bartneck, J. Reichenbach, and v. A. Breemen, "In your face, robot! the influence of a character's embodiment on how users perceive its emotional expressions," in *Proceedings of the Design and Emotion*. Ankara Turkey, 2004, pp. 32–51.
- [6] A. Todorov, "Evaluating faces on trustworthiness: An extension of systems for recognition of emotions signaling approach/avoidance behaviors," *Annals of the New York Academy of Sciences*, vol. 1124, no. 1, pp. 208–224, 2008.
- [7] A. Todorov, P. Mende-Siedlecki, and R. Dotsch, "Social judgments from faces," *Current opinion in neurobiology*, vol. 23, no. 3, pp. 373–380, 2013.
- [8] J. Willis and A. Todorov, "First impressions: Making up your mind after a 100-ms exposure to a face," *Psychological science*, vol. 17, no. 7, pp. 592–598, 2006.
- [9] L. Ewing, F. Caulfield, A. Read, and G. Rhodes, "Perceived trustworthiness of faces drives trust behaviour in children," *Developmental Science*, vol. 18, no. 2, pp. 327–334, 2015.
- [10] A. Jones and G. Castellano, "Adaptive robotic tutors that support self-regulated learning: A longer-term investigation with primary school children," *International Journal of Social Robotics*, vol. 10, no. 3, pp. 357–370, 2018.
- [11] P. Vogt, M. De Haas, C. De Jong, P. Baxter, and E. Krahmer, "Child-robot interactions for second language tutoring to preschool children," *Frontiers in human neuroscience*, vol. 11, p. 73, 2017.
- [12] T. Belpaeme, J. Kennedy, A. Ramachandran, B. Scassellati, and F. Tanaka, "Social robots for education: A review," *Science robotics*, vol. 3, no. 21, p. eaat5954, 2018.
- [13] P. L. Witt, L. R. Wheelless, and M. Allen, "A meta-analytical review of the relationship between teacher immediacy and student learning," *Communication Monographs*, vol. 71, no. 2, pp. 184–207, 2004.
- [14] V. Charisi, D. Davison, D. Reidsma, and V. Evers, "Evaluation methods for user-centered child-robot interaction," in *2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 2016, pp. 545–550.
- [15] M. Paetzel, G. Perugia, and G. Castellano, "The persistence of first impressions: The effect of repeated interactions on the perception of a social robot," in *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, 2020, pp. 73–82.
- [16] M. Heerink, B. Kröse, V. Evers, and B. Wielinga, "Assessing acceptance of assistive social agent technology by older adults: the almere model," *International journal of social robotics*, vol. 2, no. 4, pp. 361–375, 2010.
- [17] N. N. Oosterhof and A. Todorov, "The functional basis of face evaluation," *Proceedings of the National Academy of Sciences*, vol. 105, no. 32, pp. 11 087–11 092, 2008.
- [18] M. Holodynski and W. Friedlmeier, *Development of emotions and emotion regulation*. Springer Science & Business Media, 2006, vol. 8.
- [19] F. Ma, F. Xu, and X. Luo, "Children's facial trustworthiness judgments: Agreement and relationship with facial attractiveness," *Frontiers in psychology*, vol. 7, p. 499, 2016.
- [20] D. J. McAllister, "Affect-and cognition-based trust as foundations for interpersonal cooperation in organizations," *Academy of management journal*, vol. 38, no. 1, pp. 24–59, 1995.
- [21] 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*, In press.
- [22] A. R. Landrum, C. M. Mills, and A. M. Johnston, "When do children trust the expert? benevolence information influences children's trust more than expertise," *Developmental Science*, vol. 16, no. 4, pp. 622–638, 2013.
- [23] M. Tielman, M. Neerincx, J.-J. Meyer, and R. Looije, "Adaptive emotional expression in robot-child interaction," in *2014 9th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 2014, pp. 407–414.
- [24] J. M. Kessens, M. A. Neerincx, R. Looije, M. Kroes, and G. Bloothoof, "Facial and vocal emotion expression of a personal computer assistant to engage, educate and motivate children," in *2009 3rd international conference on affective computing and intelligent interaction and workshops*. IEEE, 2009, pp. 1–7.
- [25] G. Skantze, "Predicting and regulating participation equality in human-robot conversations: Effects of age and gender," in *2017 12th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 2017, pp. 196–204.
- [26] A. Tsiami, P. P. Filntisis, N. Efthymiou, P. Koutras, G. Potamianos, and P. Maragos, "Far-field audio-visual scene perception of multi-party human-robot interaction for children and adults," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2018, pp. 6568–6572.
- [27] N. Calvo-Barajas, M. Elgarf, G. Perugia, C. Peters, and G. Castellano, "Can a social robot be persuasive without losing children's trust?" in *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, 2020, pp. 157–159.
- [28] T. S. Wingenbach, C. Ashwin, and M. Brosnan, "Validation of the amsterdam dynamic facial expression set-bath intensity variations (adfes-biv): A set of videos expressing low, intermediate, and high intensity emotions," *PloS one*, vol. 11, no. 1, 2016.
- [29] B. Tay, Y. Jung, and T. Park, "When stereotypes meet robots: the double-edge sword of robot gender and personality in human-robot interaction," *Computers in Human Behavior*, vol. 38, pp. 75–84, 2014.
- [30] M. Siegel, C. Breazeal, and M. I. Norton, "Persuasive robotics: The influence of robot gender on human behavior," in *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2009, pp. 2563–2568.
- [31] J. Bernotat, F. Eyssel, and J. Sachse, "The (fe) male robot: How robot body shape impacts first impressions and trust towards robots," *International Journal of Social Robotics*, pp. 1–13, 2019.
- [32] M. S. Bernath and N. D. Feshbach, "Children's trust: Theory, assessment, development, and research directions," *Applied and Preventive Psychology*, vol. 4, no. 1, pp. 1–19, 1995.
- [33] C. Bartneck, D. Kulić, E. Croft, and S. Zoghbi, "Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots," *International journal of social robotics*, vol. 1, no. 1, pp. 71–81, 2009.
- [34] S. A. Moubayed, G. Skantze, and J. Beskow, "The furhat back-projected humanoid head-lip reading, gaze and multi-party interaction," *International Journal of Humanoid Robotics*, vol. 10, no. 01, p. 1350005, 2013.
- [35] P. Ekman and W. Friesen, "Facial action coding system: A technique for the measurement of facial movement, consulting psychologists press," *Palo Alto*, 1978.
- [36] P. Ekman, "An argument for basic emotions," *Cognition & emotion*, vol. 6, no. 3-4, pp. 169–200, 1992.
- [37] M. Paetzel, G. Castellano, G. Varni, I. Hupont, M. Chetouani, and C. Peters, "The attribution of emotional state-how embodiment features and social traits affect the perception of an artificial agent," in *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 2018, pp. 495–502.
- [38] Y. Tang, P. L. Harris, H. Zou, and Q. Xu, "The impact of emotional expressions on children's trust judgments," *Cognition and Emotion*, vol. 33, no. 2, pp. 318–331, 2019.
- [39] A. A. Marsh, R. B. Adams Jr, and R. E. Kleck, "Why do fear and anger look the way they do? form and social function in facial expressions," *Personality and Social Psychology Bulletin*, vol. 31, no. 1, pp. 73–86, 2005.