Effects of Robot's Facial Expressions on Children's Perception of Trustworthiness in First Encounters

Natalia Calvo[1], Giulia Perugia[1], and Ginevra Castellano[1]

Department of Information Technology, Uppsala University, Sweden
{natalia.calvo, giulia.perugia, ginevra.castellano}@it.uu.se

Abstract. Facial expressions of emotion have an impact on interpersonal trait inference. Facial cues of emotions serve as indicators to know when to approach or avoid a person. In child-robot interaction, the first impression of a robot matters to motivate children to engage in long-term interaction. We constructed a set of facial expressions to be projected onto a Furhat robot to evaluate the perception of trustworthiness, likability, and competence. We investigated how the emotion type and intensity of the facial expression, together with the gender-likeness of the robot, affected the perception of the robot in first encounters. Results from a 2x2x3 mixed design study show that happiness and anger at a different level of intensity influence children's judgment of likability and competence, but do not influence the judgment of trustworthiness. Our results demonstrate the importance of considering dynamic facial expression when introducing robots to children.

Keywords: Trustworthiness · Child-Robot Interaction · Facial Expressions

1 Introduction

Robots are increasingly used as children's companions in schools and homes, leading designers, roboticists and computer scientists to build social robots that can interact with humans in a friendly and at the same time safe way.

Some robots are designed to express emotions: this feature has been associated to higher social acceptance and perceived trustworthiness [6, 17]. Ekman's set of six basic expressions of emotion (anger, disgust, surprise, fear, happiness, sadness) can be used to predict social behaviors and help in the interpersonal trait inference (e.g. dominance and trustworthiness) [7]. In particular, when humans are exposed to a stranger's facial expression, they can judge trustworthiness, competence, likeability, aggressiveness, and attractiveness in a few seconds [21].

While adults can accurately evaluate emotions from robots [3], it is still unclear whether children can infer emotions from a robot's facial expressions. This inference is crucial as emotions are used to convey interpersonal information

and influence the perception of trust in Human-Robot Interaction (HRI) [6,17]. 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 robots [6].

This study aims to investigate how a robot's facial expressions and its gender-likeness could affect children's perception of trust. We explored if children can attribute trustworthiness to a male-like or female-like robot based on the type and intensity of their facial expressions during first encounters. We used Furhat, a back-projected human-like robot to convey facial expressions in a smooth way [15]. We conducted a 2(Gender)x2(Emotion)x3(Intensity) study to evaluate children's perception of trustworthiness, likeability, and competence of a robot's facial expression at different levels of emotion intensity.

2 Related work

Trust is a key element when robots interact with humans in a shared work-space. For instance, it influences the human decision of choosing a robot teammate to complete a task [1, 13] or serves as an indicator of social acceptance in the evaluation of a robot's performance [10,12].

In human-robot interaction (HRI), trust has two subcategories: *cognitive-based trust*, which is related to robot performance factors (e.g. reliability, behavior, failure rates), and *affective-based trust*, which refers to attribute factors (robot personality, robot type, and anthropomorphism) [11]. Although a robot's performance has a stronger effect on trust development, there is evidence that first impressions are crucial to establish trust and rapport: these are influenced by human traits, cognitive and social factors [10,11].

In the educational and assistive field, social robots are used as companions and tutors for children, due to their benefits in increasing cognitive and affective outcomes [4]. The understanding of child-robot interaction (cHRI) includes not only the examination of children's social interaction, acceptance, and emotional involvement with a robot [5], but also the impact of robot's behavior and its physical-related features on the interaction. The quality of the interaction is assessed from first encounters with the robot. Good quality ensures the effectiveness of cHRI and motivates children to interact with the robot. For instance, the way a robot is introduced to children generates a first impression formation [20]. Such a first impression affects children's comfort and rapport with the robot and, consequently, trust development and learning outcomes [4–6].

Happiness and anger are part of the set of six basic emotions proposed by Ekman [7], which could be expressed with specific facial expressions. These emotions convey information about the affective state and interpersonal intent of a subject and are correlated with the judgment of trust and dominance [16,19], for instance, happy faces are perceived as more trustworthy than faces that express anger. Based on the theory that humans make trait inference about others' facial appearance to guide their trust perception and behavior [19], and that such skill is developed in early childhood [9], we posit that robot's facial expression

of happiness and anger might affect children's perception of robots' trust in first encounters.

As mentioned before, humans make trustworthiness judgments after brief exposures to faces and these judgments affect intentions and behaviors towards the other person. Happiness signals that the observed person can be approached, anger that s/he should be avoided [19]. Besides, literature suggests that people can better judge emotions that are high in intensity [3]. However, there is not enough evidence that they can make these judgments based on robot's faces. In [3], users could infer the emotions of the iCat robot and rate their intensity. Their result suggests that users are more prone to detect the exaggerated expression of emotion in a robot. We hence pose the following research questions (RQs):

RQ1: Do children perceive a robot that expresses happiness as more trustworthy, likable, and competent, than a robot that expresses anger?

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

Also, there is evidence that female stimuli are perceived as significantly more likable and trustworthy than male stimuli [18]. These findings reveal the importance of gender in social judgment. We want to understand if gender-likeness affects cHRI in first encounters. Hence, we state the following RQ:

RQ3: Do children perceive a female-like robot as more trustworthy, likable, and competent than a male-like robot?

3 Method

3.1 Design and Measures

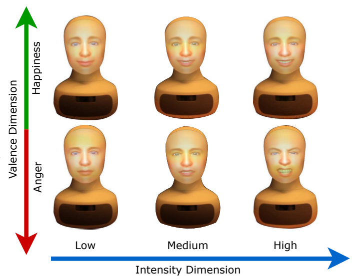
We constructed a study with a 2x2x3 design. We had three independent variables: Robot Gender (Female-like/Male-like), Emotion Type (Happiness/Anger), and Intensity Level (Low/Medium/High); and eight dependent variables— we used 8 items of the Godspeed questionnaire [2]— which refer to three factors: Trustworthiness (Core, Secrets, Truth-teller, Influence), Likability (Appearance, Friendly), and Competence that investigates the perceived intelligence of the robot by using two items (Smartness, Help) [21]. We consider the last two factors as facilitators of trust in cHRI. These judgments may reflect the evaluation of the robot's facial expression that approximates children's perception of a trustworthy robot in first encounters.

3.2 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. 129 children (46 female, 56 male, 27 No Answer) took part in the study, they ranged in age from 9 to 14 years ($M = 11.29, SD = 0.85$).

3.3 Apparatus and Stimuli

We designed a set of dynamic facial expressions to be projected onto the Furhat robot [15]. We used the valence emotional dimension—that signals whether to approach or avoid a person—for happiness (approach) and anger (avoid) as proposed by [16, 19], and the intensity dimension for the strength/weakness of the emotion (Figure 1.a).



(a) Exemplar Male Stimuli

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

(b) Emotion Recognition

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

(c) Intensity of Happiness

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

(d) Intensity of Anger

Fig. 1: (a). Exemplar male stimuli projected onto Furhat. (b) Confusion matrix emotion recognition. (c) Confusion matrix intensity of Happiness. (d) Confusion matrix intensity of Anger. Percentage of correctly classification are shown in bold

Authentic smiles have an onset duration of 0.8s, an offset duration of 2.1s, and an apex duration of 1.8s [14]. To avoid having stimuli of different duration, we kept this onset, apex, and offset length both for happiness and anger. We used an in-built domain-specific-language provided by Furhat in the Kotlin programming language to create the stimuli. We manipulated the action units suggested by Ekman to animate the mouth and eye-browns regions to create dynamic facial expressions for happiness and anger [7]. To give the impression of a change in intensity, we modified the facial expression displayed in the apex by varying its intensity in agreement with the Facial Action Coding System (FACS) [8]. In the study, we used two textures (male/female) in-built in the Furhat platform, expressing two emotions (happiness/anger), and three levels of intensity (low/medium/high). See Figure 1.a.

Validating the set of stimuli. To asses whether the proposed set of stimuli was perceived as expected, we conducted a validation 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 1.b-d: the rows indicate the label assigned by the participant and the columns indicate the actual label of the stimuli. Results revealed that happiness was correctly identified by almost all participants [92%], while anger was more difficult to identify [53%], as participants confused it with disgust [21%], happiness [10%], and surprise [8%] (Figure 1.b). This result is similar to previous studies that showed that anger, disgust, and fear are more difficult to decode [17]. Concerning intensity, we found that high intensity for anger had the highest recognition rate [76%], followed by low for anger and medium for happiness [51%], and high happiness [47%] (Figures 1.c and 1.d). 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.*

3.4 Experimental Setup and Procedure

During data collection at Scifest, 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. 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 expressing facial expressions twice, 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. The allocation of the experimental conditions was randomized.

4 Results

We employed a 2X2X3 MANOVA with Robot Gender (male-like/female-like), Emotion Type (happiness/anger) and Intensity Level (low/medium/high) as independent variables. All ρ-values of the post-hoc analysis are adjusted with the 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. 111 participants completed all the questionnaires, only their responses were considered in our results.

The Effect of Emotion Type. The results of the MANOVA showed a significant main effect of Emotion Type, $F(8, 191) = 2.869, \rho = .005$, on the dependent variables. Post-hoc tests revealed that children perceived the robot as more likeable in terms of appearance, $\rho = .001$, if it expressed happiness ($M = 3.38, SD = 1.01$) instead of anger ($M = 2.89, SD = 1.14$), and, in terms of friendliness, $\rho = .0.023$, 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 more competent in terms of smartness, $\rho = .002$, if it expressed happiness ($M = 4.12, SD = .95$) instead of anger ($M = 3.64, SD = 1.25$). We did not find any significant effect of Emotion Type on the perception of trustworthiness. Therefore, we can conclude that *in our study, happiness increased the likability and competence of the social robot in the first encounter.*

The Effect of Intensity Level. The MANOVA revealed that there was a significant main effect of the Intensity Level, $F(40, 935) = 1.525, \rho = .021$, 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 ($M = 2.76, SD = 1.43$), $\rho = .007$, and instead of low anger ($M = 2.92, SD = .98$), $\rho = .037$. Also the robot was found more likable in terms of friendliness when it expressed low happiness ($M = 3.68, SD = .91$) instead of medium happiness ($M = 2.73, SD = 1.04$), $\rho = .011$, and instead of high anger ($M = 2.67, SD = 1.59$), $\rho = .003$. 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 ($M = 3.48, SD = 1.17), \rho = .038$. We did not find a significant effect of Intensity Level on the perception of trustworthiness. Therefore, we can conclude that *in our study, the intensity level of emotion is a factor that affects the perception of likability and competence in the first encounter.*

The Effect of Gender of the Robot. Results revealed no significant main effect of Gender Robot, $F(8, 183) = 1.275, \rho = .259$, on the dependent variables, nor a significant interaction effect of Robot Gender and Emotion Type, $F(8, 191) = 1.252, \rho = .271$. However, we found a significant interaction effect of Robot Gender and Intensity of Emotion, $F(40, 935) = 1.453, \rho = .036$. The interaction effect revealed that the male-like robot was perceived as more likable when it expressed high anger ($M = 3.92, SD = 1.4$) instead of medium anger ($M = 2.89, SD = .87), \rho = .037$, and when it expressed low happiness ($M = 3.75, SD = .57$) instead of medium anger ($M = 2.89, SD = .87), \rho = .015$, and low anger ($M = 3, SD = 1), \rho = .037$. We also found that the female-like robot was perceived as less likable when the robot expressed high anger ($M = 1.94, SD = .89$) instead of: medium anger ($M = 3.15, SD = 1.14), \rho = .002$, low anger ($M = 2.85, SD = 0.98), \rho = .008$, high happiness ($M = 3.55, SD = 1.03), \rho \leq .001$, medium happiness ($M = 2.88, SD = 1.2), \rho = .010$, and low happiness ($M = 3.57, SD = .81), \rho \leq .001$; also, when the robot expressed low anger, $\rho = .026$, and medium happiness, $\rho = .042$, instead of low happiness. Therefore, we can conclude that *in our study, the gender of the robot influences children's judgment of appearance only when the robot expresses happiness or anger at different intensity levels.*

5 Discussion

Our results suggest that 5s are enough for children to judge likability and competence of a social robot that displays facial expressions at different levels of intensity. We did not find evidence that our set of dynamic facial expressions

directly affected the perception of the trustworthiness of the robot. However, as competence and likability are facilitators of trust, we argue that children are more prone to build *cognitive-based trust* when the robot expresses happiness with high levels of intensity (RQ1). This might be due to the fact that children could make a better trait inference when the robot expressed the high intensity of the emotion (RQ2). We would like to highlight an interesting result: children perceived the female-like robot expressing high anger as less likable than all the other stimuli. This might have a negative effect on *affective-based trust* development in long-term interaction (RQ3).

We believe that the 3D blended robot embodiment itself has a stronger influence on the perception of trustworthiness rather than the emotional state expressed by the robot. As a limitation, we found that most of the children were surprised when we uncovered the robot. This might have partially hindered their reactions to the first stimulus presented.

In conclusion, we argue that the emotion and the intensity of a facial expression affect children's first impression of a robot in terms of competence and likability. Our results stand out the impact of the robot's facial expression when are introduced to children, especially when robots are used as companions or peers in education. Also, the robot's behaviors are the key factor for eliciting trust, because it would allow maintaining children's engagement in long-term interactions. Trustworthiness is not directly influenced by facial expressions. However, as likability and competence are facilitators of trustworthiness, the emotions displayed in first encounters and their intensity might affect the development of trust over time. These findings are important not just for educational scenarios, but also for assistive settings, as trust is a key factor for long-term adoption of social robots [12, 20].

Acknowledgement

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

References

1. Alves-Oliveira, P., Sequeira, P., Melo, F.S., Castellano, G., Paiva, A.: Empathic robot for group learning: A field study. ACM Transactions on Human-Robot Interaction (THRI) **8**(1), 3 (2019)
2. Bartneck, C., Kulić, D., Croft, E., Zoghbi, S.: Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. International journal of social robotics **1**(1), 71–81 (2009)
3. Bartneck, C., Reichenbach, J., Breemen, v.A.: 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. pp. 32–51. Ankara Turkey (2004)
4. Belpaeme, T., Kennedy, J., Ramachandran, A., Scassellati, B., Tanaka, F.: Social robots for education: A review. Science robotics **3**(21), eaat5954 (2018)

5. Charisi, V., Davison, D., Reidsma, D., Evers, V.: Evaluation methods for user-centered child-robot interaction. In: 2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN). pp. 545–550. IEEE (2016)
6. Cohen, I., Looije, R., Neerincx, M.A.: Child's perception of robot's emotions: effects of platform, context and experience. International Journal of Social Robotics **6**(4), 507–518 (2014)
7. Ekman, P.: An argument for basic emotions. Cognition & emotion **6**(3-4), 169–200 (1992)
8. Ekman, R.: What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS). Oxford University Press, USA (1997)
9. Ewing, L., Caulfield, F., Read, A., Rhodes, G.: Perceived trustworthiness of faces drives trust behaviour in children. Developmental Science **18**(2), 327–334 (2015)
10. Gaudiello, I., Zibetti, E., Lefort, S., Chetouani, M., Ivaldi, S.: Trust as indicator of robot functional and social acceptance. an experimental study on user conformation to icub answers. Computers in Human Behavior **61**, 633–655 (2016)
11. Hancock, P.A., Billings, D.R., Schaefer, K.E., Chen, J.Y., De Visser, E.J., Parasuraman, R.: A meta-analysis of factors affecting trust in human-robot interaction. Human factors **53**(5), 517–527 (2011)
12. Heerink, M., Kröse, B., Evers, V., Wielinga, B.: Assessing acceptance of assistive social agent technology by older adults: the almere model. International journal of social robotics **2**(4), 361–375 (2010)
13. Jones, A., Castellano, G.: Adaptive robotic tutors that support self-regulated learning: A longer-term investigation with primary school children. International Journal of Social Robotics **10**(3), 357–370 (2018)
14. Krumhuber, E., Manstead, A.S., Cosker, D., Marshall, D., Rosin, P.L., Kappas, A.: Facial dynamics as indicators of trustworthiness and cooperative behavior. Emotion **7**(4), 730 (2007)
15. Moubayed, S.A., Skantze, G., Beskow, J.: The furhat back-projected humanoid head–lip reading, gaze and multi-party interaction. International Journal of Humanoid Robotics **10**(01), 1350005 (2013)
16. Oosterhof, N.N., Todorov, A.: The functional basis of face evaluation. Proceedings of the National Academy of Sciences **105**(32), 11087–11092 (2008)
17. Paetzel, M., Castellano, G., Varni, G., Hupont, I., Chetouani, M., Peters, C.: 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). pp. 495–502. IEEE (2018)
18. Siegel, M., Breazeal, C., Norton, M.I.: Persuasive robotics: The influence of robot gender on human behavior. In: 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems. pp. 2563–2568. IEEE (2009)
19. Todorov, A.: Evaluating faces on trustworthiness: An extension of systems for recognition of emotions signaling approach/avoidance behaviors. Annals of the New York Academy of Sciences **1124**(1), 208–224 (2008)
20. Vogt, P., De Haas, M., De Jong, C., Baxter, P., Krahmer, E.: Child-robot interactions for second language tutoring to preschool children. Frontiers in human neuroscience **11**, 73 (2017)
21. Willis, J., Todorov, A.: First impressions: Making up your mind after a 100-ms exposure to a face. Psychological science **17**(7), 592–598 (2006)