Robots Educate in Style

The Effect of Context and Non-verbal Behaviour on Children's Perceptions of Warmth and Competence

Peters, Rifca; Broekens, Joost; Neerincx, Mark A.

DOI
10.1109/ROMAN.2017.8172341

Publication date
2017

Document Version
Final published version

Published in
26th IEEE International Symposium on Robot and Human Interactive Communication, IEEE RO-MAN 2017

2017 26th IEEE International Symposium on
Robot and Human Interactive Communication (RO-MAN)
Lisbon, Portugal, Aug 28 - Sept 1, 2017.

Robots Educate in Style: The Effect of Context and Non-verbal Behaviour on Children's Perceptions of Warmth and Competence

Rifca Peters[1], Joost Broekens[1] and Mark A. Neerincx[1,2]

***Abstract*— Social robots are entering the private and public domain where they engage in social interactions with non-technical users. This requires robots to be socially interactive and intelligent, including the ability to display appropriate social behaviour. Progress has been made in emotion modelling. However, research into behaviour style is less thorough; no comprehensive, validated model exists of non-verbal behaviours to express style in human-robot interactions. Based on a literature survey, we created a model of non-verbal behaviour to express high/low warmth and competence—two dimensions that contribute to teaching style. In a perception study, we evaluated this model applied to a NAO robot giving a lecture at primary schools and a diabetes camp in the Netherlands. For this, we developed, based on expert ratings, an instrument measuring perceived warmth, competence, dominance and affiliation. We show that even subtle manipulations of robot behaviour influence children's perceptions of the robot's level of warmth and competence.**

I. Introduction

Robots are emerging as collaboration partners in training and education (e.g., [2], [32], [17]), and the importance of social intelligence has been stressed [13]. Educational robots may take the role of tutor, tool or peer; learning from, about or with robots [24]. Different roles have been linked to learning activities. For example, for basic learning tasks, a peer robot was preferred above a tutor robot [27], but for language learning a tutor was preferred [30]. The quality of human teacher-student interactions is partly determined by the teacher's ability to adapt their style to the student and activity [15]. Thus, roles can be performed with various styles, and, like human educators, effective educational robots should be able to express appropriate styles.

Roles and style are being used in experimental human-robot interaction (HRI) (e.g., [23], [33]). However, these studies seldomly report a validated behaviour model. For example, in [23] a *motivator-robot* was intended to signal *empathy* and *trustworthiness* by listener behaviour (gaze, nod and 'listening expression'), but validation of this model has not been reported. Attempting to bridge the social intelligence gap between humans and machines, social signal processing (SSP) research focused on modelling of and synthesis of agent behaviour based on analysis of human behaviour (e.g., [8], [26]). Up to now, research focused on understanding human behaviour, and exploring effects of robot behaviour. However, the question of *how* to stylize and express this has received less attention. Beneficial personalization of robot behaviour requires understanding of the effect of interaction style, and validation of style perception. In this paper, we report a perception study evaluating non-verbal behaviours to express *warmth* and *competence* for an educational robot.

* This work is funded by the Horizon2020 PAL-project (grant 643783)

[1] Interactive Intelligence Group, Delft University of Technology, Mekelweg 4, 2628 CD Delft, The Netherlands r.m.peters@tudelft.nl
[2] TNO, Postbus 23, 3769 ZG Soesterberg, The Netherlands mark.neerincx@tno.nl

II. Background

The importance of robot behaviour style is grounded in educational psychology. Human educators adapt both content and style of communication to the learner and task at hand. This ability is believed to be crucial for effective and motivating educational interactions. Style refers to (opposed to more stable personality traits and role) the current observable behaviour—the way a role is performed. This can be trained and used strategically.

A. Teaching Style

Teaching styles are behaviour patterns that affect information presentation and interaction. Style is considered a technique addressing differences in learning- and cognitive styles [11]. Grasha [15] constructed five teaching styles: *Expert* (transmitting information), *Formal authority* (providing feedback and establishing boundaries), *Personal model* (showing an example), *Facilitator* (encouraging critical thinking), and *Delegator* (available in the background during project work). Teaching style is selected based on student capabilities (e.g., knowledge, responsibility and motivation), teacher's need for control, and teacher's willingness to build and maintain relationships. Certain clusters of styles appear most frequent in classrooms and were linked to specific situations and strategies. For example, teacher-centred styles (i.e., expert and formal authority) were linked to teacher control, and less capable students, while student-centred styles (i.e., facilitator and delegator) were linked to teachers willing to loosen control, and more capable students.

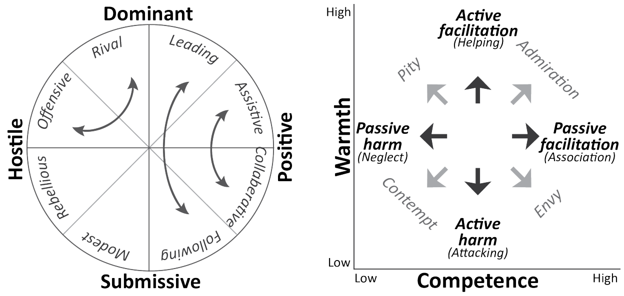
B. Interpersonal Circumplex

This model—also know as Leary's Rose—defines interaction stance by two axes: *dominance* and *affiliation* [20], [31]. The horizontal affective-axis depicts willingness to cooperate. The vertical dominance-axis depicts the degree of power. Commonly, the circumplex is partitioned into eight octants (Fig. 1a). Further, Leary's theory states two interaction rules: dominance is complementary and affiliation symmetric, meaning that, an opposed stance evokes opposed stance and dominance evokes submissiveness. A cooperative style has been linked to maintaining contact, a competitive style to aversion. For example, head nods [14], gaze, and open posture [4] are associated with high affiliation. A dominant style is commonly stereotyped as loud and obtrusive, while submissive style is believed to show discrete, unnoted behaviour. For example, loudness, vocal control, and gaze are associated with perceived dominance [10], [6]. Moreover, dominant people are believed to lean forward, use more gestures, have open and up-right posture, and orient towards others [7]. However, a meta-analysis of the relation between dominance and non-verbal behaviours argued that previous work has been inconclusive or based on limited data, and concluded that the relation exists to different degrees and even directions, depending on the person and situation [16]. Nonetheless, the model is widely used in social skill training.

C. *Stereotype Content Model*

The stereotype content model (SCM) defines *warmth* and *competence* as two fundamental dimensions of social perception [12]. Warmth (kindness, empathy, friendliness and trustworthiness) evaluates valance of intent, competence (intelligence, power, efficacy and skill) assesses the ability to act on these intentions. Perceived warmth and competence generate emotions of admiration, envy, pity, and disgust towards someone, and predict active/passive and facilitative/harmful behaviour patterns (Fig. 1b). Perceived warmth is believed to be evoked by sincere smiles, head tilt, nodding, leaning forward, and open gestures [6], [9]. Coldness is expressed by closed hands, cutting motion, chin down, and the body pivoting away [9]. Upright posture and open gestures, are predictors of perceived power, and associated with competence [6], [9]. Fiddling was suggested to signal low control and confidence, therefore resulting in low-competence judgement [9]. Additionally, warmth judgements are believed to be made before, and influence, competence evaluations—persons evaluated as warm, are likely to be judged more competent [6].

In summary, style can be defined as a behaviour pattern signalling our attitude towards a person or situation, affecting how others evaluate us and subsequently respond. Four social constructs important to the notion of style are discussed: *warmth*, *competence*, *dominance*, and *affiliation*.



(a) Interpersonal Circumplex (b) Stereotype Content Model

Fig. 1: Models of social interaction evaluation and prediction.

III. Related Work

In HRI, *interaction style* has been defined as a combination of behaviours that evoke a perceivable robot role (e.g., [28]). For example, in collaborative play, a NAO robot took the role of *peer* (collaborative behaviour) or *tutor* (scaffolding support) [33]. Recent studies investigated the effect of robot behaviour on user experience or outcome (e.g., [28], [2], [21], [18], [19], [33]). However, validation of behaviour models is scares. At best, evaluation was done with parents [18] or teachers [33]. Thus, it remains unclear if children perceived the robot as intended. Consequently, we cannot compare and build upon results.

Studies validating behaviour models have focused on emotion expression (e.g., [22], [32]) or personality (e.g., [25]). Although, recently interest has developed in modelling social perceptions such as competence and dominance in virtual humans (e.g., [1], [5]). Presence of gestures positively influenced perceived warmth and competence [3]. A more elaborate model[1] of virtual agent behaviours expressing warmth and competence appeared successful [26].

IV. Robot Style Model

Based on the non-verbal behaviours associated with warmth and competence described above, we created a model of non-verbal behaviours for a NAO robot expressing four different style configurations: *high-warmth and high-competence* (HwHc); *high-warmth and low-competence* (HwLc); *low-warmth and high-competence* (LwHc); and *low-warmth and low-competence* (LwLc). The focus on warmth and competence was chosen because of the extensive, applicable work done in [26]—expression of dominance and affiliation are subject to subsequent studies.

First, we selected cues applicable to our set-up. For example, the ‘sync’ cue was rejected because of limitations with respect to timing of behaviours proposed by the software (Section V-A), and the NAO robot is incapable of facial expressions. The resulting model is presented in Table I. Next, we applied our model to the robot by annotating available behaviours, creating a library of animations fitting the style configurations. Lastly, we added fitting behaviours to the text sequence (script). For example, when the robot says ‘Pay attention’, for HwHc the open *StateLeft* behaviour and *head-up* are selected, and for LwLc the closed *CapisceLeft* behaviour and *head-down* (Fig. 2).

V. Study 1: Perceived Warmth and Competence

A. *Method*

Evaluating the effect of non-verbal behaviour on children’s perceptions of an educational robot’s style, we conducted a 2×2 (*warmth* $\times$ *competence*) between-subject perception study at primary schools. For this, a PowerPoint presentation with text and animation script for the four robot styles was created using the RoboTutor framework (`https://github.com/RoboTutor`). The 10-minute lecture on

[1]To conserve space, we refer the interested reader to the authors original work for details on the model.

(a) High Warmth, High Competence (b) Low Warmth, Low Competence

Fig. 2: Stills of behaviour accompanying a statement in two of the four robot style configurations.

TABLE I

OVERVIEW OF THE BEHAVIOUR CUES FOR THE FOUR ROBOT STYLE CONFIGURATIONS (WARMTH × COMPETENCE). ⋆ = COMPETENCE, • = WARMTH – = WARMTH × COMPETENCE.

*CONDITIONS USED IN SECOND EXPERIMENT

High Warmth, High Competence* *Paralinguistic* • low pitch • low volume *Body Posture* ⋆ stable • directed at audience • head tilt – gaze fixed at audience *Hand Gestures* • open • semantic & syntactic ⋆ frequent (every sentence) – mid-peripheral – centre-centre	**Low Warmth, High Competence** *Paralinguistic* • high pitch • high volume *Body Posture* ⋆ stable • pivot away • chin down – gaze fixed at audience *Hand Gestures* • closed • semantic only ⋆ frequent (every sentence) – low-centre – mid-centre-centre
High Warmth, Low Competence *Paralinguistic* • low pitch • low volume *Body Posture* ⋆ wobbling • directed at audience • head tilt – gaze fixed at audience *Hand Gestures* • open • semantic & syntactic ⋆ low frequency – low/mid-peripheral	**Low Warmth, Low Competence*** *Paralinguistic* • high pitch • high volume *Body Posture* ⋆ non-stable • pivot away • chin down – gaze diversion *Hand Gestures* • closed • semantic only ⋆ low frequency – low/high-centre/peripheral – fiddling

robotics, given by a NAO robot (`www.alderbaran.com`), included three multiple choice questions, which could be answered using the Turningpoint polling system (`https://www.turningtechnologies.com`). The robot speech was accompanied by non-verbal behaviours, which varied between groups (*HwHc, HwLc, LwHc, LwLc*). Participant's perceptions of the robot's level of *warmth*, *competence*, *dominance*, and *affiliation* were collected after the lecture by rating 20 adjectives on a three-point Likert scale. Face validity was ensured by selecting adjectives at child level, and in correspondence with a developmental psychologist. Construct validity was ensured by expert ratings.

Measurement: Perceived *competence*, *warmth*, *dominance*, and *affiliation* were measured by an adjective-based instrument developed for the present study. Children rated 20 adjectives on a three-point Likert scale. Perception score for each dependent variable were calculated by multiplying word-rating with a loading based on expert ratings.

The adjectives (translated from Dutch: *bossy, nagging, clumsy, friend, popular, playful, follower, loner, angry, honest, fights, knowledgeable, boring, nice, listener, confident, educational, helpless, helpful, dumb*) were chosen from words commonly used to describe various positions in Leary's rose or the SCM, and likely to be present in young children's vocabulary. Children rated whether each word would describe the robot (yes, sometimes/maybe, no) by placing a sticker in the corresponding column of a response leaflet. Stickers were used because physical activity was suggested to reduce the primacy effect (selecting extreme high or low for all items), thus increasing reliability [29]. Ratings were coded on a numeric interval scale [2–0], where a higher value is associated with better fitting description. Children did not place words they did not understand, these missing values were replaced by the population mean for that adjective.

The loadings for adjectives on the dependent variables were calculated from expert ratings. Eleven experts in human-computer interaction rated each adjective on four bipolar scales [-2–2]. We assessed the reliability of ratings, excluding words with a standard deviation of one or above. Further, words with a median of 0 were excluded for that construct because the association was weak. From the remaining words (10 for competence, 13 for warmth, 5 for dominance, and 11 for affiliation) we took the mean as loading value. This provides a table of 20 adjectives and their loading, if any, on each dependent variable (Table II).

Participants: A total of 101 children, from two primary schools in the Netherlands (S1, n = 40; S2, n = 61), participated in our study. Children at S1 were 10-13 years of age ($M = 11.43, SD = 0.64$), and enrolled in 5th (n = 9) or 6th (n = 31) grade. Children at S2 were aged 5-8 ($M = 6.52, SD = 0.65$), all enrolled in first grade. Gender was evenly distributed (S1 male = 20, female = 20; S2 male = 30, female = 28); three children did not report their gender. All participants were naive to the research aim and had little to no previous experience with a NAO robot. Children within each class were assigned to one of four robot style configurations. Children from the same school, in the same condition were merged into one group. This way, at both schools, all groups contained a minimum of 10 children, controlled for age, class and gender.

Procedure: Before the experiment, the researcher was briefly introduced to the children in the classroom, and children were assigned a group. Afterwards, when all children had participated, children were given the opportunity to ask questions about the robot and experiment, and a demonstration was given. The following steps were repeated for each group:

- children entered, were seated and given a 'clicker';
- researcher introduced the robot and instructed the children to remain seated after the lecture;
- researcher started the selected script;
- robot gave a lecture displaying stylized behaviours;
- researcher explained the questionnaire, accompanied by a brief example rating the popular Disney figure Simba;
- children spread across the room;
- researcher handed the children stickers with 20 adjectives, followed by a response leaflet and pencil; and
- children provided their individual ratings.

TABLE II
LOADINGS FOR 20 ADJECTIVES (TRANSLATED FROM DUTCH) ON THE DEPENDENT VARIABLES. THE VALUES PRESENT THE MEAN EXPERT RATINGS FOR THE ADJECTIVES ON FOUR BIPOLAR SCALES, COMPLIANT WITH OUR CRITERIA ($SD \geq 1$ AND $Mdn \neq 0$).

Adjective	Competence	Warmth	Dominance	Affiliation
Bossy	–	-1.18	2.00	-0.55
Nagging	-0.73	-0.73	–	-1.09
Clumsy	-1.73	–	–	–
Friend	–	1.91	–	1.36
Popular	0.82	0.91	0.55	–
Playful	–	1.36	–	1.82
Follower	–	–	–	–
Loner	–	–	–	–
Angry	–	-1.18	1.09	-1.09
Honest	0.82	1.00	–	1.46
Fight	–	-1.36	1.36	-1.64
Knowledgeable	1.64	–	–	–
Boring	–	-0.64	–	–
Nice	–	1.82	–	1.09
Listener	0.91	1.09	–	1.46
Confident	–	-0.55	1.00	–
Educational	1.64	–	–	0.91
Helpless	-1.45	–	–	–
Helpful	0.64	1.27	–	1.64
Dumb	-1.91	–	–	–

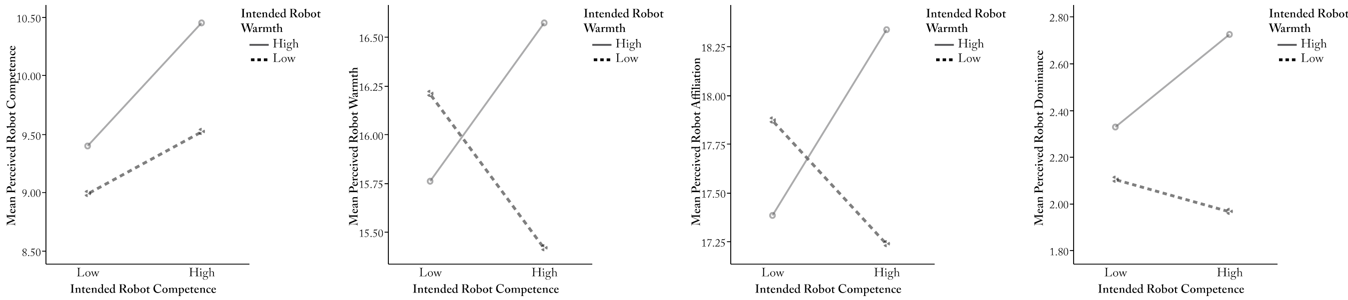
B. Results

We explored differences in children's perceptions of robots displaying high/low warmth and competence-related behaviours. Although, K-S tests indicated non-normal distributions, we decided to perform a MANOVA because we are interested in the interaction effect and the sample size is fair.

Using Pillai's trace, there was a near significant interaction effect of intended warmth $\times$ competence on how children perceived the robot, $V = 0.08, F(4, 95) = 2.09, p = 0.088$. Separate univariate ANOVAs on the outcome variables revealed a significant interaction effect for intended warmth $\times$ competence on perceived affiliation, $F(1, 97) = 4.42, p = 0.038$; and warmth, $F(1, 97) = 4.09, p = 0.046$. Further, near significant main effects were found for intended competence on perceived competence, $F(1, 97) = 3.30, p = 0.072$; and intended warmth on perceived dominance, $F(1, 97) = 3.81, p = 0.054$.

In other words, children perceived a robot displaying high-competence behaviours (stable posture, frequent gestures) as more competent ($M = 10.00$) than low-competence (unstable posture, low frequency of gestures) ($M = 9.20$), regardless of the intended level of warmth (Fig. 3a). And a robot displaying high-warmth behaviours was perceived slightly warmer ($M = 16.18$) than low-warmth robots ($M = 15.83$). However, competence influenced the effect of warmth manipulations on how children perceived the robot; robots displaying high-warmth behaviours were perceived as more warm than low-warmth robots, but only in the high-competence condition. High-competence robots displaying high-warmth behaviours were perceived warmer ($M = 16.58$) than low-warmth robots ($M = 15.42$), low-competence robots displaying high-warmth behaviours were perceived considerably less warm ($M = 15.76$) than low-warmth robots ($M = 16.21$) (Fig. 3b). See also Table III for mean perception scores.

To investigate the bias of individual words on perception scores, we explored differences in word-ratings. We performed a non-parametric Mann-Whitney tests, once for each construct, because normality could not be assumed. Comparison of high and low competence samples showed a significant difference in ratings for 'Follower' ($U = 913.00, z = -2.49, p = 0.013, r = -0.25$) and 'Helpless' ($U = 798.50, z = -2.69, p = 0.007, r = -0.28$); near significant differences were present for 'Helpful' ($U = 933.00, z = -1.83, p = 0.068, r = -0.19$). No difference in word-ratings were found for 'Dumb', 'Nice', 'Knowledgeable', and 'Nagging'. Comparison of high and low warmth samples showed near significant differences for 'Popular' ($U = 1290.00, z = 1.83, p = 0.067, r = 0.19$), 'Loner' ($U = 991.50, z = -1.76, p = 0.078, r = -0.18$), 'Angry' ($U = 1323.00, z = 1.72, p = 0.086, r = 0.17$), 'Confident' ($U = 1305.00, z = 1.65, p = 0.099, r = 0.17$), and 'Helpful' ($U = 1307.00, z = 1.83, p = 0.068, r = 0.19$). Since all children reported exactly the same rating for 'Dumb', there was once again no difference between samples.



(a) Mean perceived competence (range -11.64 to 12.91).

(b) Mean perceived warmth (range -11.27 to 18.72).

(c) Mean perceived affiliation (range -8.73 to 19.46).

(d) Mean perceived dominance (range 0.00 to 12.00).

Fig. 3: Study 1 results showing mean perception scores. Competence condition displayed on horizontal axis, warmth by separate lines.

C. Discussion

The results indicate that children did perceive robots differently based on their non-verbal behaviours—the set-up, PowerPoint, textscript and questions were similar for all groups. However, the differences are small. This may be due to subtlety of the manipulation or measurement.

Variations in non-verbal behaviour were subtle, attempting to avoid creating a caricature and keep behaviour believable. However, observers of all four conditions were not able to note any difference, indicating that we may have been too careful selecting and adapting animations. This can explain the small differences in perception scores between robot styles. Future studies exploring the effect of enlarged differences in behaviour between styles is needed. Overall perception of warmth and competence was on the high end of the range, indicating that children have a positive stance towards the robot.

The measurement instrument, which we created based on literature and in correspondence with experts, was partly validated using expert ratings. Children confirmed to understand all words, with one exception (Popular) for the youngest children. However, children do not necessarily have the same interpretation of words. Analysis of individual word-ratings revealed that children provided significantly different ratings for words not included in calculation of the outcome measures; ratings of 'Loner' differed between high/low intended warmth, but this word did not contribute to calculation of perceived warmth. Similar, was the case for 'Follower' on competence. In fact, 'Loner' and 'Follower' did not load to any of the dependent variables. Further, 'Dumb' and 'Knowledgeable' were used to calculate perceived competence with notable loadings of -1.91 and 1.64 respectively. However, no differences in word-ratings were present. This may indicate a ceiling effect; children never think of the robot as being dumb. Or it may result from a priming effect; the robot stated to know about robotics and teach the children about this subject. Discarding 'Knowledgeable' from calculation of competence scores yields a significant difference in perceived competence between high/low competence samples, $F(1, 97) = 4.41, p = 0.038$. This indicates that validating and improving the reliability of our measurement instrument are worthwhile.

Our findings on competence are in line with the suggestions from the literature: stable posture and frequent gestures evoke higher perceptions of competence. Our findings on warmth seemed in line with the literature: overall, low pitch and volume, body directed at audience, head tilt, open (semantic and syntactic) gestures evoked higher perceptions of warmth. However, perceived warmth declined in the HwLc condition compared to LwLc. This indicates that low-competence behaviours (unstable posture, infrequent gestures) reversed the effect of warmth behaviours. Alternatively, although opposed to the literature, the fiddling and gaze diversion only present in the LwLc condition could account for the effect. Additional head and hand movements might make the robot more lively and therefore perceived warmer. Our findings partially contradict the theory that warmth is evaluated before competence and therefore characters perceived as warm are likely to be found competent. Although competence scores were slightly higher for high-warmth robots than low-warmth samples, the difference was not significant, $F(1, 97) = 2.35, p = 0.129$. The difference in perceived competence between high/low competence samples was larger in the high-warmth (1.04) condition than low-warmth (0.53), and the HwLc robot was perceived marginally less competent than the LwHc robot. This indicates that warmth expression enhances the competence effect rather than biases towards perceived high competence.

VI. Study 2: Context Dependency

A. Method

To explore the effect of context (i.e., location, content, and usergroup) on children's perceptions of the robot, we conducted a 2 $\times$ 2 (*robot style* $\times$ *context*) follow-up study at a camp for children with type 1 diabetes mellitus (T1DM). Only two groups could be formed due to practical limitations, limiting us to two robot style configurations (HwHc and LwLc). The procedure and materials were similar to Study 1, except the lecture was about using MyPAL—an app developed to support coping with and learning about T1DM.

Participants: A total of 72 children participated, 52 from the first study at schools (HwHc = 26, LwLc = 26), and 20 from the camp. We recruited 21 children with T1DM, of which 6 had experienced interacting with a NAO robot before during earlier studies. One participant was excluded from further analysis because the questionnaire was not understood and completed. This left 20 participants from camp (male = 12, female = 8), aged 8-11 ($M = 9.20, SD = 1.10$), divided in two groups (HwHc = 9, LwLc = 11).

B. Results

We explored differences in children's perceptions of the robot between the two contexts (school and camp) and between robot styles (HwHc and LwLc). We performed MANOVA analysis to explore main and interaction effects between robot styles and contexts of use.

Using Pillai's trace, there was a significant main effect for context on perception, $V = 0.16, F(4, 65) = 3.18, p = 0.019$. Separate univariate ANOVAs on the outcome variables revealed significant effects of context on perceived competence, $F(1, 68) = 6.61, p = 0.012$, and affiliation, $F(1, 68) = 6.99, p = 0.010$, and near a significant effect on warmth, $F(1, 68) = 3.52, p = 0.065$. No significant multivariate effect for robot style on perception was found. However, univariate analysis revealed a near significant effect for robot style on competence, $F(1, 68) = 3.03, p = 0.086$. No interaction effects were found for context $\times$ robot style, $V = 0.07, F(4, 65) = 1.14, p = 0.345$.

This indicates that the context (location, users, and content) of the activity (presentation by a NAO robot) influenced perceived warmth, affiliation and competence. Overall, children with T1DM at camp perceived the robot as less warm ($M = 15.28$), affiliated ($M = 16.61$) and competent ($M = 8.14$) than children at school (warmth $M = 16.39$, affiliation $M = 18.11$, competence $M = 9.72$) (Fig 4a, 4b, 4c). Further, we were able to replicate the effect of robot style on perception scores with 20 new participants. Robots displaying high-competence and high-warmth behaviours were perceived more competent ($M = 9.94$) than low-competence and low-warmth robots ($M = 8.65$), independent of the context (Fig. 4a).

C. Discussion

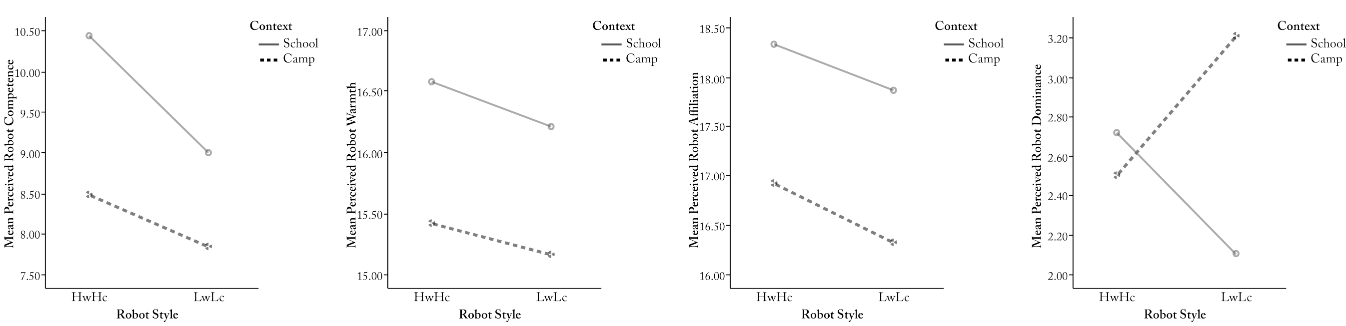
Children at camp perceived the robot less warm, affiliated and competent than children at school, independent of the robot style. Thus, changing the users, content and location of an activity can influence how children perceive an educational robot, meaning that results obtained in one context do not necessarily translate to another context.

Lower perceived competence may indicate that the setting of Study 1 primed the children to think of the robot as competent because it would teach them. Although we expected the children at camp to see the robot as a friendly helper, and therefore perceive it as more warm and affiliated, this was not the case. Six of the children at camp had previously experienced interacting with the NAO robot. Possibly, when novelty wears of, children are more rigorous in evaluating the robot. Or children may compare the robot's behaviour to previous experiences and find the robot less warm in the current activity compared to for example playing a game together. Alternatively, the framing as 'friendly helper' set expectations which the robot could not live up to. Further research is needed to provide solid conclusions on the influence of context on perception scores.

The lack of difference in perceptions between robot styles at camp may result from the small samples or too subtle manipulations. Although this was not the main purpose of the second study, it would be interesting to repeat the study and explore the effect of robot style using a larger sample size and more exaggerated behaviour manipulations.

VII. CONCLUSION

We evaluated an educational robot displaying non-verbal behaviours to express high/low warmth and competence with children at primary schools and camp, and showed that even subtle manipulations in robot behaviour influence children's perceptions of the robot's level of warmth and competence. The competence dimension in our model was successfully, but warmth manipulations had the intended effect only for high-competence robots. Moreover, context influenced children's perceptions; at school the robot was perceived warmer than at camp. Although this was a fist attempt, and further research is needed replicating the study with enlarged behaviour manipulations and other social constructs, to our knowledge this is the first evidence of style in robot teaching—and a first step towards adaptive style.



(a) Mean perceived competence (range -11.64 to 12.91).

(b) Mean perceived warmth (range -11.27 to 18.72).

(c) Mean perceived affiliation (range -8.73 to 19.46).

(d) Mean perceived dominance (range 0.00 to 12.00).

Fig. 4: Study 2 results showing mean perception scores. Robot condition displayed on horizontal axis, contexts as separate lines.

TABLE III

MEAN WORD RATINGS AND PERCEPTION SCORES FOR BOTH CONTEXT, GROUPED BY ROBOT STYLE CONFIGURATION.

	Style	Bossy	Nagging	Clumsy	Friend	Popular	Playful	Follower	Loner	Angry	Honest	Fights	Knowledgeable	Boring	Nice	Listener	Confident	Educational	Helpless	Helpful	Dumb	Dominance	Affiliation	Warmth	Competence
School	HwHc	0.19	0.04	0.19	1.96	1.58	1.92	1.15	0.19	0.04	1.81	0.04	1.92	0.00	2.00	1.88	1.40	2.96	0.85	1.88	0.00	2.72	18.34	16.58	10.44
	HwLc	0.08	0.00	0.27	1.71	1.42	1.79	1.40	0.43	0.08	1.74	0.14	1.87	0.04	2.00	1.78	1.14	1.83	1.14	1.95	0.00	2.33	17.39	15.76	9.40
	LwHc	0.13	0.04	0.21	1.79	1.13	1.83	1.00	0.50	0.00	1.92	0.04	1.75	0.04	1.96	1.67	1.04	1.87	0.71	1.58	0.00	1.97	17.24	15.42	9.52
	LwLc	0.12	0.12	0.38	1.92	1.38	1.85	1.42	0.44	0.00	1.77	0.04	1.79	0.19	1.96	1.85	1.04	1.85	1.17	1.88	0.00	2.11	17.87	16.21	8.99
Camp	HwHc	0.33	0.00	0.56	1.89	1.44	1.67	0.67	0.67	0.11	2.00	0.11	1.89	0.22	2.00	1.44	0.78	1.89	1.22	1.78	0.11	2.51	16.94	15.42	8.49
	LwHc	0.18	0.36	0.55	1.91	1.56	1.91	0.55	0.36	0.36	1.64	0.18	1.73	0.09	2.00	1.73	1.36	1.55	1.00	1.82	0.09	3.22	16.34	15.17	7.86

ACKNOWLEDGMENT

Selection and Participation of Children: We recruited 101 children (aged 5-13, M/F) from primary schools in the Netherlands. An invitation to participate was offered to teachers who requested a robot demonstration. 21 children diagnosed with type 1 diabetes mellitus (aged 8-11, 8 female, 13 male) were recruited by the Dutch Diabetes Association DVN. The research was approved by the Human Research Ethics Committee of Delft University of Technology.

REFERENCES

[1] A. Arya, L. N. Jefferies, J. T. Enns, and S. DiPaola. Facial actions as visual cues for personality. *Computer Animation and Virtual Worlds*, 17(3-4):371–382, 2006.

[2] T. Belpaeme, P. E. Baxter, R. Read, R. Wood, H. Cuayáhuitl, B. Kiefer, S. Racioppa, I. Kruijff-Korbayová, G. Athanasopoulos, V. Enescu, R. Looije, M. Neerincx, Y. Demiris, R. Ros-Espinoza, A. Beck, L. Cañamero, A. Hiolle, M. Lewis, I. Baroni, M. Nalin, P. Cosi, G. Paci, F. Tesser, G. Sommavilla, and R. Humbert. Multimodal Child-Robot Interaction: Building Social Bonds. *JHRI*, 1(2):33–53, 2012.

[3] K. Bergmann, F. Eyssel, and S. Kopp. A second chance to make a first impression? how appearance and nonverbal behavior affect perceived warmth and competence of virtual agents over time. In *IVA*, volume 7502 LNAI, pages 126–138. Springer, 2012.

[4] D. S. Berry and J. S. Hansen. Personality, Nonverbal Behaviour, and Interaction Quality in Female Dyads. *PSPB*, 26(3):278–292, 2000.

[5] A. Cafaro, N. Glas, and C. Pelachaud. The Effects of Interrupting Behavior on Interpersonal Attitude and Engagement in Dyadic Interactions. In *AAMAS*, pages 911–920. ACM, 2016.

[6] L. L. Carli, S. LaFleur, and C. Loeber. Nonverbal behavior, gender, and influence. *Personality and Social Psychology*, 68(6):1030–1041, 1995.

[7] D. R. Carney, J. A. Hall, and L. S. LeBeau. Beliefs about the nonverbal expression of social power. *Nonverbal Behavior*, 29(2):105–123, 2005.

[8] M. Chollet, M. Ochs, and C. Pelachaud. A multimodal corpus for the study of non-verbal behavior expressing interpersonal stances. In *IVA Workshop Multimodal Corpora: Beyond Audio and Video*, 2013.

[9] A. J. Cuddy, M. Kohut, and J. Neffinger. Connect, Then Lead. *Harvard Business Review*, 91(7):54–61, 2013.

[10] N. E. Dunbar and J. K. Burgoon. Perceptions of power and interactional dominance in interpersonal relationships. *Social and Personal Relationships*, 22(2):207–233, 2005.

[11] R. M. Felder and L. K. Silverman. Learning and Teaching Styles in Engeneering Education. *Engr. Education*, 78(7):674–681, 1988.

[12] S. T. Fiske, A. J. C. Cuddy, and P. Glick. Universal dimensions of social cognition: warmth and competence. *Trends in Cognitive Sciences*, 11(2):77–83, 2007.

[13] T. Fong, I. Nourbakhsh, and K. Dautenhahn. A survey of socially interactive robots. *Robotics and Autonomous Systems*, 42(3):143–166, 2003.

[14] R. Gifford. The Role of Nonverbal Communication in Interpersonal Relations. In *Handbook of interpersonal psychology: Theory, research, assessment, and therapeutic interventions*, chapter 11, pages 171–190. Wiley, NY, 2011.

[15] A. F. Grasha. A Matter of Style: The Teacher as Expert, Formal Authority, Personal Model, Facilitator, and Delegator. *College Teaching*, 42(4):142–149, 1994.

[16] J. a. Hall, E. J. Coats, and L. S. LeBeau. Nonverbal behavior and the vertical dimension of social relations: a meta-analysis. *Psychological bulletin*, 131(6):898–924, 2005.

[17] L. Hall, C. Hume, S. Tazzyman, A. Deshmukh, S. Janarthanam, and H. Hastie. Map Reading with an Empathic Robot Tutor. In *HRI*, pages 567–567. IEEE, 2016.

[18] W. Johal, S. Pesty, and G. Calvary. Towards companion robots behaving with style. In *Ro-MAN*, volume 23, pages 1063–1068, Endinbrugh, 2014. IEEE.

[19] J. Kennedy, P. Baxter, and T. Belpaeme. The Robot Who Tried Too Hard: Social Behaviour of a Robot Tutor Can Negatively Affect Child Learning. In *HRI*, pages 67–74, Portland, 2015. ACM.

[20] T. Leary. Interpersonal diagnosis of personality. *American Journal of Physical Medicine & Rehabilitation*, 37(6):331, 1958.

[21] I. Leite, G. Castellano, A. Pereira, C. Martinho, and A. Paiva. Empathic Robots for Long-term Interaction: Evaluating Social Presence, Engagement and Perceived Support in Children. *J of Soc Robotics*, 6(3):329–341, 2014.

[22] Y. H. Lin, C. Y. Liu, H. W. Lee, S. L. Huang, and T. Y. Li. Evaluating emotive character animations created with procedural animation. In *IVA*, volume 5773 LNAI, pages 308–315. Springer, 2009.

[23] R. Looije, M. a. Neerincx, and V. D. Lange. Childrens responses and opinion on three bots that motivate , educate and play. *Physical Agents*, 2(2):13–20, 2008.

[24] O. Mubin, C. J. Stevens, S. Shahid, A. A. Mahmud, and J.-j. Dong. A review of the applicability of robots in education. *Technology for Education and Learing*, 1, 2013.

[25] M. Neff, Y. Wang, R. Abbott, and M. Walker. Evaluating the effect of gesture and language on personality perception in conversational agents. In *IVA*, volume 6356 LNAI, pages 222–235. Springer, 2010.

[26] T.-h. D. Nguyen, E. Carstensdottir, N. Ngo, M. S. El-nasr, M. Gray, D. Isaacowitz, and D. Desteno. Modeling Warmth and Competence in Virtual Characters. In *IVA*, volume 9238, pages 167–180. Springer, 2015.

[27] S. Y. Okita, V. Ng-Thow-Hing, and R. Sarvadevabhatla. Learning together: ASIMO developing an interactive learning partnership with children. In *RO-MAN*, pages 1125–1130. IEEE, 2009.

[28] S. Y. Okita, V. Ng-Thow-Hing, and R. K. Sarvadevabhatla. Multimodal approach to affective human-robot interaction design with children. *TiiS*, 1(1):1–29, 2011.

[29] R. Ros, M. Nalin, R. Wood, P. Baxter, R. Looije, Y. Demiris, T. Belpaeme, A. Giusti, and C. Pozzi. Child-robot interaction in the wild: advice to the aspiring experimenter. In *ICMI*, pages 335–342. ACM, 2011.

[30] M. Saerbeck, T. Schut, C. Bartneck, and M. D. Janse. Expressive robots in education: varying the degree of social supportive behavior of a robotic tutor. In *CHI*, pages 1613–1622. ACM, 2010.

[31] J. S. Wiggins. An informal history of the interpersonal circumplex tradition. *Personality Assessment*, 66(2):217–233, 1996.

[32] J. Xu, J. Broekens, K. Hindriks, and M. Neerincx. Effects of Bodily Mood Expression of a Robotic Teacher on Students. In *IROS*, pages 2614–2620. IEEE, 2014.

[33] C. Zaga, M. Lohse, K. P. Truong, and V. Evers. The Effect of a Robot's Social Character on Children's Task Engagement: Peer Versus Tutor. In *ICSR*, volume 9388 LNAI, pages 704–713, Paris, 2015. Springer.