A Little Chit-Chat Goes a Long Way

Design and Evaluation of Task-and Person-Oriented Styles for Social Robots

Saad, Elie; Broekens, Joost; Neerincx, Mark A.

DOI
10.1109/RO-MAN60168.2024.10731475

Publication date
2024

Document Version
Final published version

Published in
33rd IEEE International Conference on Robot and Human Interactive Communication, ROMAN 2024

Citation (APA)
Saad, E., Broekens, J., & Neerincx, M. A. (2024). A Little Chit-Chat Goes a Long Way: Design and Evaluation of Task-and Person-Oriented Styles for Social Robots. In *33rd IEEE International Conference on Robot and Human Interactive Communication, ROMAN 2024* (pp. 1712-1719). (IEEE International Workshop on Robot and Human Communication, RO-MAN). IEEE. https://doi.org/10.1109/RO-MAN60168.2024.10731475

2024 33rd IEEE International Conference on
Robot and Human Interactive Communication (RO-MAN)
August 26-30, 2024. Pasadena, California, USA.

2024 33rd IEEE International Conference on Robot and Human Interactive Communication (ROMAN) | 979-8-3503-7502-2/24/$31.00 ©2024 IEEE | DOI: 10.1109/RO-MAN60168.2024.10731475

A Little Chit-Chat Goes a Long Way: Design and Evaluation of Task- and Person-Oriented Styles for Social Robots

Elie Saad[1,2], Joost Broekens[3] and Mark A. Neerincx[1,4]

Abstract— **Whereas the reception task is a promising application domain for social robots, knowledge is lacking about how to design the appropriate re-usable communication styles for a reception robot. This paper presents the use and evaluation of an iterative interaction-design (*ID*) method with which *task-* and *person-oriented* multi-modal communication styles have been designed for such a robot. First, we report on an evaluation study of the ID-method with Industrial Design students ($N = 13$) who designed these two communication styles for a Pepper robot. This provided a set of distinct designs of the two styles, for which the differences in design parameters were in line with social science theory. The *task-oriented* style showed a more formal, shorter and less chatty communication. Second, we present findings from a Mechanical Turk study conducted to evaluate the perception of these style designs. Participants ($N = 301$) were presented with videos showing the robot acting as a receptionist and were asked to rate their perception of the robot, the service experience and the orientation of the designs. Overall, the interaction with the robot was appreciated well. The robot with a *person-oriented* style was perceived to be more animate and likeable. Analysis showed that chit-chat was the main contributor to the perceived difference between the *person-oriented* and *task-oriented* styles. This is an important finding as it gives interaction designers a validated best-practice approach to make interaction style more or less personal.**

Index Terms— **Human-Robot Interaction; Multi-Modal Communication; Interaction Design; Communication Styles; Expressive Behaviors; Reception Robot.**

I. Introduction

Assisting receptionists to deal with visitors at public spaces, such as meeting centers and hospitals, is a promising task for social robots. This helps free (human) staff from basic and repetitive tasks (e.g., welcoming visitors, confirming appointments and providing time & location indications) to focus on more complex requests (e.g., asking for further explanations and finding and communicating the answers). However, providing such a service by social robots requires that they are able to appropriately communicate with people, i.e., use the appropriate communication styles to initiate, maintain and end an interaction (e.g., [1], [2]). Most often, the interaction with visitors at a reception consists of a practical (goal-driven) exchange of information which follows a *task-oriented* communication style. However, the interaction may also include a social (context-driven) exchange of information, following a *person-oriented* style (e.g., [3]).

Our application domain, the reception in healthcare organisations, is selected for being able to provide a "rich" test environment to study such communication styles in a human-robot interaction (HRI) context. There are relatively high potential benefits for robot deployment (i.e., "fit for application"), and the communication appeals to social, cognitive and affective skills (i.e., "fit for research"). The communication has to meet high demands on effectiveness (high costs of errors), efficiency (scarcity of personnel and time) and subjective experience (the setting and reason of the visit can evoke emotional responses). In addition to the administrative and visitor support tasks (e.g., [4]), the Health-Care Receptionists (HCRs) have to deal with the rather continuous attendance of visitors having a diversity of states which affects (1) the way to approach the visitor (e.g., [3]); and (2) the choice of the communicative behavior (e.g., [5]). The behavior of healthcare staff, i.e., their caring approach, is also found to have an impact on the visitors' acceptability of the provided service and influence their continuity of care (e.g., [6], [4]). Therefore, it is important to carefully examine and design the visitor-technology communication in healthcare.

Designing the communication entails the specification of different variables, such as the speech acts (e.g., text), the voice (e.g., speed), the posture (e.g., openness) and the gazing (e.g., diversion). There are examples of how a humanoid robot can express moods or learning styles with such variables (e.g., [7], [8]). However, as far as our knowledge extends, re-usable proven designs of comprehensive *task-* and *person-oriented* communications for a reception robot are not available yet. This paper focuses on the creation of these designs to acquire further insight in both the process and outcomes of designing such communications.

Having a multi-dimensional design space, a sound HRI design requires a process to iteratively refine the design variables (e.g., speech acts and gesture openness) extracted from a systematic theory-driven exploration. For this, we need to answer four research questions: (1) How to support the proposed design refinement process, (2) does such support bring about the desired theory-based different designs, (3) how does the refinement process evolve over different designers, and (4) are the differences in the theory-based designs perceivable?

To answer these research questions, we first report on a design study which we conducted with Industrial Design students [9]. Participants ($N = 13$) were asked to design

[1] Faculty of Electrical Engineering, Mathematics and Computer Science, Department of Intelligent Systems, Delft University of Technology, Delft, The Netherlands. E.Saad@tudelft.nl
[2] School of Engineering, Holy Spirit University of Kaslik, Jounieh, Lebanon.
[3] The Leiden Institute of Advanced Computer Science, Leiden University, Leiden, The Netherlands. Joost.Broekens@gmail.com
[4] The Netherlands Organization for Applied Scientific Research (TNO), Soesterberg, The Netherlands. Mark.Neerincx@tno.nl

TABLE I

PARAMETERS USED FOR DESIGNING *task*- AND *person-oriented* COMMUNICATION STYLES

Parameter		**Variation**
Gesture Openness		*Close* or *Open*
Gaze Diversion		*Fixated* or *Diverted*
Proxemic Closeness (m)		from *Close* ($0.5m$) to *Far* ($3m$)
Voice	**Pitch**	from *Low* (-50% than default)
	*default is 100%	to *High* ($+50\%$ than default)
	Speed	from *Slow* (50% slower than default)
	*default is 100%	to *Fast* (150% faster than default)
	Prosody	*Weak* or *Strong*
Speech Acts		*Informal* or *Formal*
Chit-Chat		*No-Chat* or *Chat*
Eye Color		*White* / *Red* / *Green* / *Blue*

TABLE II

VALUES OF THE PARAMETERS USED IN THE LAST ITERATION CYCLE OF THE ITERATIVE DESIGN

	Communication Style	
	Task-Oriented	Person-Oriented
Gesture Openness	Close	Open
Gaze Diversion	Fixated	Fixated
Proxemics (m)	0.75	0.75
Pitch	90 (-10% than default)	100 (default pitch)
Speed	93 (7% slower than default)	92 (8% slower)
Prosody	Weak	Weak
Speech Acts	Formal	Informal
Chit-Chat	No	Yes
Eye Color	Blue / Green	White / Blue
# of Blocks	7	9

communication styles for a social robot by following a two-step process: create an individual design of *person*- and *task-oriented* communications for a reception robot; and contribute to an iterative design by refining the design produced by the previous participant. Second, we conducted a follow-up Mechanical Turk (MTurk) study ($N = 301$) for evaluating the perception of the designed communication styles.

This paper builds on the conference paper [9] that reported about the first study with 13 Industrial Design students, mentioned above. Section II gives an overview of the first design study. Section III provides the second follow-up perception study. Sections IV and V contain the discussion and general conclusions of the paper.

II. FIRST STUDY: DESIGNING THE STYLES

In [9], we conducted a study with Industrial Design students ($N = 13$) who designed *person*- and *task-oriented* styles for social robots. The designers created interactions between a receptionist robot and a visitor entering a clinic to check-in for a scheduled appointment. This Section provides a brief summary of the experimental methods and findings of our previous work.

A. Modeling the Communication Styles

To model the *task*- and *person-oriented* communication styles, we created a list of interaction-design blocks and derived a set of design parameters (Table I). These parameters, paired with chit-chat, were used to express the two communication styles, i.e., *person*- and *task-oriented* by designing multi-modal communicative robot behaviors. The blocks were used by the designers (i.e., participants) for creating the interaction flow while adjusting the parameter-settings to express a given style. The content of the chit-chat was the same, the only difference was using the block or not.

B. Design Method

In this study, designers (i.e., participants) were asked to create two versions of a social robot (i.e., a hospital reception robot) by manipulating the intended communication style (i.e., independent variable) following the *person*- and *task-oriented* style. Each participant designed for the next iteration by following an iterative interaction-design (*ID*) method. The styles were designed and deployed directly on the *Pepper*[1] robot.

The *ID*-method consists of a two-step process. In the first – individual – step, the designer adjusts the parameter-settings to create an individual design of the *task* and *person-oriented* styles without any influence from other designs. In the second – iterative – design step, the designer is asked to fine-tune the parameter-settings of the current design iteration received from previous designers to produce the next iteration of the *task* and *person-oriented* styles. In the end, each participant thus delivers 4 different designs: two *person-oriented* designs and two *task-oriented* designs, for both the individual and iterative steps.

C. Findings

The findings showed that designing multi-modal communicative behaviors for social robots following the *ID*-method and tool was effective and provided promising results. The iterative step helped designers to reflect on their own individual designs when adjusting the parameter-settings of the iterative design towards a stable end point. This was based on the overall convergence of the designs and the harmonization of the individual designs into the iterative design. Further, there was a difference in the process and outcome of designing a *task*- or a *person-oriented* interaction: the former showed a more formal, shorter and less chatty communication. The values of the parameters used to design the styles in the last iteration cycle of the iterative design are shown in Table II.

The interaction time with the *Pepper* robot was significantly longer for the *person-oriented* style in both the individual and iterative designs. Designers significantly used chit-chat to express the *person-oriented* style, which corroborated this finding. It seems that spending more time interacting with visitors is expected from a robot with a *person-oriented* style, which is consistent with [10].

The findings from parameter-settings also show that the preferred proxemic for both styles was within the personal zone (i.e., less than $1.5m$). Further, the voice design was

[1] SoftBank Robotics, `https://www.softbankrobotics.com`

Fig. 1. The *Pepper* robot acting as a healthcare receptionist assistant. The 3D rendering of the *Pepper's* background image is courtesy of Artistic Visions, LLC, `https://artisticvisions.com`).

consistent for expressing the styles of the *Pepper* robot, which is in line with [11] where the authors signal the importance of matching the robot's appearance with its voice.

III. Second Study: Evaluating the Styles

We conducted an Amazon MTurk[2] perception study to evaluate the development method and the designs expressing the *person-* and *task-oriented* communication styles (Sect. II). This approach allowed us to recruit a high number of MTurk workers (i.e., participants) and gather enough data points for analyzing the results. In this section, we discuss the experimental methods and findings from this study.

A. Hypotheses

With respect to the perception of *task-oriented* and *person-oriented* communication styles, we formulated the following hypothesis:

*a) **H1**: Task-* and *person-oriented* communication styles for a social robot are perceived differently in accordance with social science theory. Perception differences are measured in terms of the perception of a) the robot, b) the service experience and c) the design orientations.

With respect to the design method, we formulated the following hypothesis:

*b) **H2**:* The perception of the iterative design converges. Perception convergence is defined as an increased specificity of the task and person orientation of the designs for each design iteration.

B. Participants

We employed Amazon MTurk webservices to recruit anonymous workers based on their geographical location (i.e., United States, Canada, Grand Britain and the Netherlands), the number of submitted hits (> 500) and the hits acceptance rate ($> 98\%$). All qualified MTurk workers, i.e., participants ($N = 301$), were included in the study and consisted of 187 (62.13%) males, 108 (35.88%) females and 6 (1.99%) who did not specify. Each participant received a compensation of 2 USD for their time ($M = 13.79$ minutes, $SD = 7.50$).

[2]Amazon Mechanical Turk, `https://www.mturk.com`

C. Materials

In the following we discuss the materials used for designing the MTurk human intelligence tasks (HITs) to be completed by the participants.

1) Recorded Videos: To validate the designs from [9], we recorded 52 videos[3], each corresponding to an iteration (out of 13) of a design source (i.e., individual or iterative) and style (i.e., *task-* or *person-oriented*). The videos showed an interaction between the *Pepper* robot (acting as an HCR assistant) and a visitor (i.e., patient) wanting to check-in for an appointment. For this, we used our interaction design tool to deploy the designs on the *Pepper* robot which would autonomously interact with a (human) user. The video backgrounds were edited to show the robot positioned at the reception of a healthcare establishment (Fig. 1).

2) Survey: We created a survey using Qualtrics[4] and posted it on Amazon MTurk webservices. The survey was designed to gather demographic info from the participants (i.e., age, gender, education level, employment status and previous interaction with humanoids) and record their perception after watching each video. First, the participants' perception of the person and task orientations of each video was recorded, via two questions, on a 7-point Likert scale (higher=better). Second, the perception of the robot was recorded, on a 5-point Likert scale, using the Godspeed questionnaire by Bartneck et al. [12] (the questions measuring the perceived safety and the robot's anthropomorphism were excluded for not being relevant to this study). Third, the perception of the service experience was recorded, on a 7-point Likert scale, using the questionnaire by Froehle et al. [13]. An additional question was included about the content of the dialogue to verify that the participants were paying attention when watching each video and disqualify those who were not, i.e., exclude them from the study without being compensated for their time.

D. Measurements

To verify our hypotheses, we collected measures as follows.

a) Perception of the robot: We recorded the participants' perception of the robot's animacy, likeability and perceived intelligence using a questionnaire (Sect. III-C.2). These measures are used to compare the perception of the designs and verify $H1.a$.

b) Perception of the service experience: To measure the participants' perception of the robot as a service provider, we used the 7 metrics defined in [13] (Sect. III-C.2). Three of the metrics (i.e., courtesy, professionalism and attentiveness) are characteristics of the relationship-building communication (i.e., the first measure). Another three are task-related (i.e., the second measure), namely knowledgeableness, preparedness and thoroughness. The last metric is to measure the participants' satisfaction of the service experience (i.e., the third measure). These measures will be used to verify $H1.b$.

[3]The videos will be available online when this publication is accepted.
[4]Qualtrics Survey Software, `https://www.qualtrics.com`

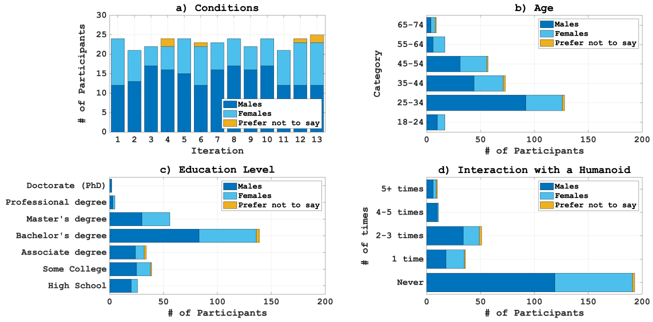

Fig. 2. Distribution of the participants ($N = 301$): a) across the conditions, i.e., iterations; b) their age category; c) their education level; and d) their previous interaction with a humanoid.

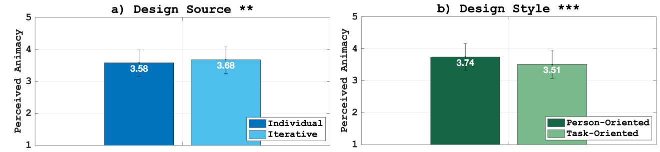

Fig. 3. Significant effects on the perception of the robot's animacy (mean scores with error bars showing 95% CI). ** = significant at .01; *** = significant at .001, explanation in text.

c) Perception of the task and person orientations of the designs: We recorded the participants' perception of the orientation of the videos, i.e., toward the task and toward the person, via a questionnaire (Sect. III-C.2). This measure will be used to verify $H1.c$.

d) Perceived specificity: We computed the perceived specificity (ΔS) which is the sum of the differences between the perceived person and task orientations of the designs at each iteration using (1). A positive ΔS at iteration i indicates that the designs were perceived differently depending on the expressed style (i.e., *person-* and *task-oriented*). The ΔS will be used to evaluate the perceived orientations of the designs over time (i.e., the iterations) and verify $H2$.

$$\begin{aligned} \Delta S_i = &(P\_Person(POD_i) - P\_Person(TOD_i)) \\ &+ (P\_Task(TOD_i) - P\_Task(POD_i)) \end{aligned} \quad (1)$$

where ΔS is the perceived specificity; P_Person and P_Task are the perceived person and task orientations; POD and TOD are the designs expressing the *person-oriented* and *task-oriented* styles; and i is the iteration number.

E. Procedure

To avoid the complexity of evaluating all the videos (i.e., a total of 52) by each participant (i.e., MTurk worker), we decided that each worker would evaluate four videos: two videos expressing a *person-oriented* style and two expressing a *task-oriented* style. The videos presented to a worker were selected from the same designer. This was useful for comparing the personal and iterative designs and, at the same time, evaluating the design process over time (i.e., iterations).

When starting the survey (Sect. III-C.2), participants were provided with a description of the study followed by their informed consent. Then, they were presented with 4 videos from the same designer, displayed in random order and without knowing which style was being used. After watching each video, participants were asked to fill-in a questionnaire to record their perception of the robot, the service experience and the design orientations. Finally, some general demographic information was collected.

F. Results

Participants ($N = 301$) were almost evenly distributed between the conditions (Fig. 2.a, $Min = 21$ and $Max = 25$) and were aged 18 years old and over (Fig. 2.b). Of the participants, 248 (82.39%) were employed full time and 202 (67.10%) had an education at the Bachelor level or above (Fig. 2.c). Further, 193 (64.12%) mentioned that they had never interacted with a humanoid robot before (Fig. 2.d).

To analyze the results, we conducted a two-way repeated measures ANOVA in which design source (i.e., iterative and individual) and design style (i.e., *person-* and *task-oriented*) were the within-subjects factors, i.e., independent variables. We applied the statistical test on the 8 measures (i.e., dependent variables) discussed in Section III-D, using design iteration as the between-subjects factor.

The multivariate tests (Table III) showed that all the within-subjects effects were statistically significant. This is an indication that the main factors (i.e., design source and style) had an effect on the observed measures, i.e., the participants' perception. However, the between-subjects factor (i.e., iteration) was not significant (Wilks' $\Lambda = .71$, $F(96, 1903) = 1.02$, $p > .05$). A further analysis showed no significant differences between the iterations for the measures where the univariate tests had a significant effect of design source $\times$ iteration, design style $\times$ iteration and/or design source $\times$ style $\times$ iteration (post hoc pairwise comparison using the Bonferroni correction, $p > .05$). Therefore, in the following, we analyze the results using only the within-subjects factors.

1) Perception of the Robot: Participants rated their impression of the robot using a 5-point Likert scale. A reliability analysis showed a good internal consistency of the scale items for animacy (mean Cronbach's $\alpha = .89$), likeability (mean Cronbach's $\alpha = .91$) and perceived intelligence (mean Cronbach's $\alpha = .89$).

a) Perceived Animacy: The main effects of design source and style were statistically significant on the perception of the robot's animacy (Fig. 3.a-b), $F(1, 288) = 10.12$, $p < .01$; and $F(1, 288) = 47.66$, $p < .001$, respectively. The perceived animacy mean score was significantly higher for the iterative than the individual design ($Mean = 3.68$, $SD = .89$; and $Mean = 3.58$, $SD = .89$, respectively) and for the *person-* than the *task-oriented* style ($Mean = 3.74$, $SD = .86$; and $Mean = 3.51$, $SD = .91$, respectively).

b) Perceived Likeability: The main effects of design source and style were significant on the perception of the robot's likeability (Fig. 4.a-b), $F(1, 288) = 8.64$, $p < .01$;

TABLE III

MULTIVARIATE TESTS FROM THE TWO-WAY REPEATED MEASURES ANOVA

		Wilks' Λ	F	df	p
Between-Subjects Effect	Iteration	0.71	1.02	(96,1903)	0.437
Within-Subjects Effects	Design Source	**0.94**	**2.30**	(8,281)	*
	Design Style	**0.69**	**15.60**	(8,281)	***
	Design Source × Style	**0.92**	**3.05**	(8,281)	**
	Design Source × Iteration	**0.65**	**1.31**	(96,1903)	*
	Design Style × Iteration	**0.62**	**1.44**	(96,1903)	**
	Design Source × Style × Iteration	**0.58**	**1.70**	(96,1903)	***

*Note: significant values are in bold text, * $p < .05$; ** $p < .01$; *** $p < .001$*

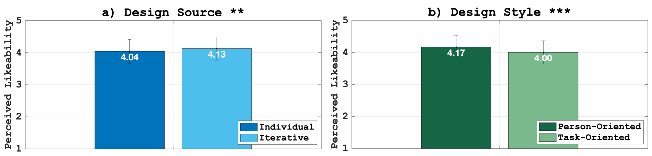

Fig. 4. Significant effects on the perception of the robot's Likeability (mean scores with error bars showing 95% CI). ** = significant at .01; *** = significant at .001, explanation in text.

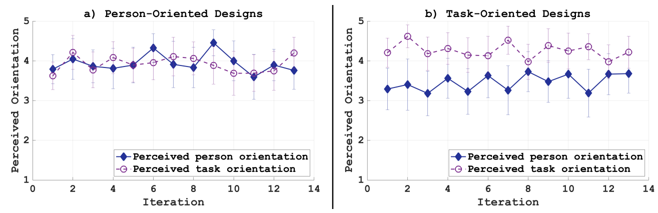

Fig. 5. Perceived person and task orientations of the designs expressing the (a) *person-oriented* and (b) *task-oriented* styles.

and $F(1, 288) = 23.79$, $p < .001$, respectively. The perceived likeability mean score was significantly higher for the iterative than the individual design ($Mean = 4.13$, $SD = .74$; and $Mean = 4.04$, $SD = .77$, respectively) and for the *person-* than the *task-oriented* style ($Mean = 4.17$, $SD = .76$; and $Mean = 4.01$, $SD = .75$, respectively).

c) Perceived Intelligence: The main effects of design source and style were not significant on the perception of the robot's perceived intelligence ($F(1, 288) = 3.72$, $p > .05$; and $F(1, 288) = .09$, ns, respectively).

The aforementioned findings from the perception of the robot's animacy and likeability showed that the *task-* and *person-oriented* styles were perceived differently ($H1.a$). Perception differences were also present between the iterative and the individual designs. The robot was perceived as more animate and likeable in the *person-oriented* style and in the iterative design.

2) Perception of the Service Experience: The following metrics were used to measure the participants' perception of the service experience and were rated on a 7-point Likert scale.

a) Relationship-Building: A reliability analysis carried out on the relationship-building scale items (i.e., courtesy, professionalism and attentiveness) showed that the questions reached a good internal consistency (mean Cronbach's $\alpha = .83$). The main effects of design source and style were not significant ($F(1, 288) = 3.84$, $p > .05$; and $F(1, 288) = .41$, ns, respectively). There was a significant interaction effect of design source × style ($F(1, 288) = 6.65$, $p < .01$). The mean relationship-building score was (slightly) higher for the *person-* than the *task-oriented* style in the individual ($Mean = 5.80$, $SD = .97$; and $Mean = 5.74$, $SD = .95$, respectively) and not in the iterative design ($Mean = 5.79$, $SD = 1.00$; and $Mean = 5.89$, $SD = .87$, respectively).

b) Task-Related: There was a good reliability for the task related scale items, i.e., knowledgeableness, preparedness and thoroughness (mean Cronbach's $\alpha = .88$). There was a significant main effect of design source and not style on the perception of the task-related characteristics of the robot ($F(1, 288) = 8.97$, $p < .01$; and $F(1, 288) = .16$, ns, respectively). The task-related mean score was higher for the iterative ($Mean = 5.91$, $SD = .96$) than the individual design ($Mean = 5.80$, $SD = 1.0$).

c) Service Satisfaction: There was a significant main effect of design source and not style on the participants' satisfaction of the service experience ($F(1, 288) = 7.31$, $p < .01$; and $F(1, 288) = 0.12$, ns, respectively). The mean satisfaction score was higher for the iterative ($Mean = 5.69$, $SD = 1.29$) than the individual design ($Mean = 5.55$, $SD = 1.23$). The interaction effect of design source × style was significant ($F(1, 288) = 5.23$, $p < .05$). The mean satisfaction score was higher for the *person-* than the *task-oriented* style in the individual design ($Mean = 5.61$, $SD = 1.29$; and $Mean = 5.49$, $SD = 1.19$, respectively) and not in the iterative ($Mean = 5.65$, $SD = 1.43$; and $Mean = 5.73$, $SD = 1.13$, respectively).

The aforementioned findings show that the manipulation of the styles had no clear effect on people's perception of the service experience, which does not support $H1.b$. There was some differences between the perception of the iterative and the individual designs, with the iterative design having higher service satisfaction score.

3) Perceived Person and Task Orientations: Participants rated (on a 5-point scale) how much they perceived each design as oriented toward the person and toward the task. The mean perceived orientations of the designs expressing the *person-* and *task-oriented* styles are illustrated in Fig. 5.a-b. The correlation between the perceived orientations was significant for both styles (Pearson's correlation $r = .130$, $p < .05$, for the *person-oriented*; and $r = -.133$, $p < .05$, for the *task-oriented*) with a small magnitude ($.1 < |r| <$

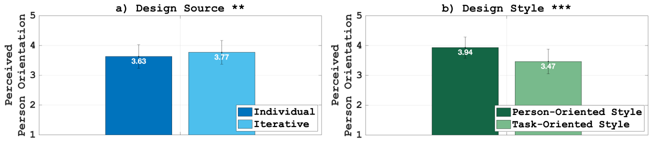

Fig. 6. Significant effects on the perceived person orientation of the designs (mean scores with error bars showing 95% CI). ** = significant at .01; *** = significant at .001, explanation in text.

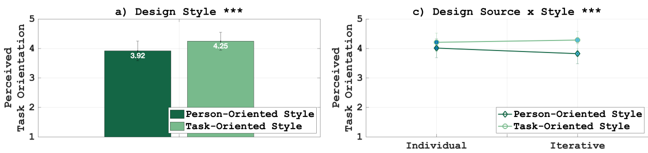

Fig. 7. Significant effects on the perceived task orientation of the designs (mean scores with error bars showing 95% CI). *** = significant at .001, explanation in text.

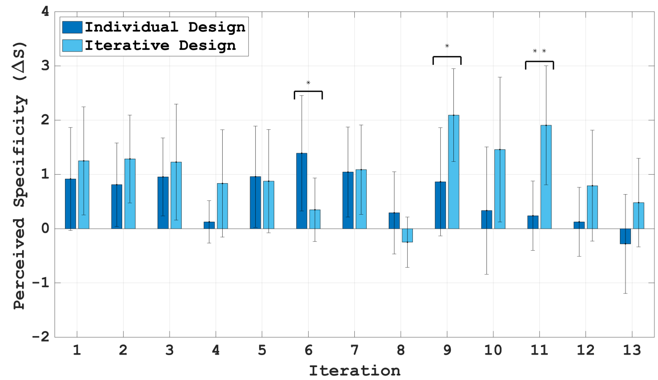

Fig. 8. Perceived specificity (ΔS), which is the sum of the differences of the perceived person and task orientations for the individual and iterative designs expressing the two styles. A positive ΔS at iteration i indicates that the perception of the designs is going in the direction of the intended expressed styles. (* = significant at .05; ** = significant at .01, explanation in text).

.3). The direction of the correlation was positive for the *person-oriented* designs, i.e., a higher orientation toward the person was associated with a higher attention to the task. However, the direction was negative for the *task-oriented*, i.e., a higher orientation toward the task was associated with a lower orientation toward the person.

The univariate statistical tests showed that the main effect of design source was significant on the perception of the designs' person orientation and not the task orientation ($F(1, 288) = 9.93$, $p < .01$; and $F(1, 288) = 2.38$, $p > .05$, respectively). The perceived person orientation (Fig. 6.a) was (slightly) higher for the iterative than the individual design ($Mean = 3.77$, $SD = 1.08$; and $Mean = 3.63$, $SD = 1.08$, respectively).

The main effect of design style was significant on the perception of both person and task orientations of the designs ($F(1, 288) = 84.16$, $p < .001$; and $F(1, 288) = 57.50$, $p < .001$, respectively). The perceived person orientation (Fig. 6.b) was higher for the *person-oriented* than the *task-oriented* style ($Mean = 3.94$, $SD = .97$; and $Mean = 3.47$, $SD = 1.14$, respectively). In contrast, the perceived task orientation (Fig. 7.a) was higher for the *task-* than the *person-oriented* style ($Mean = 4.25$, $SD = .83$; and $Mean = 3.92$, $SD = .92$, respectively).

The interaction effect of design source $\times$ style was significant on the perceived task orientation and not the person orientation of the designs ($F(1, 288) = 12.09$, $p < .001$; and $F(1, 288) = 3.76$, $p > .05$, respectively). The perceived task orientation was higher for the *task-* than the *person-oriented* style in the iterative and individual design (Fig. 7.b). Further, for the designs expressing the *person-oriented* style, the perceived task orientation was lower for the iterative than the individual design ($Mean = 3.82$, $SD = .95$; and $Mean = 4.01$, $SD = .89$, respectively).

The aforementioned findings from the perceived task and person orientations support $H1.c$, i.e., the variations in the designs expressing the *task-* and *person-oriented* styles are perceivable. The *task-oriented* style was perceived as more oriented toward the task and less toward the person than the *person-oriented* style.

TABLE IV
MULTIPLE LINEAR STEPWISE REGRESSION COEFFICIENTS

Predictor Diff. in	Unstand. C. B	Unstand. C. Std. Error	Stand. C. Beta	t	Sig.
(Constant)	1.217	0.105		11.560	***
Chit-Chat	1.175	0.093	0.568	12.683	***
Voice Pitch	-0.018	0.006	-0.123	-2.760	**
Voice Speed	-0.017	0.007	-0.108	-2.422	*
Gaze	-0.212	0.093	-0.103	-2.289	*

*Note: * = significant at .05; ** = sig. at .01; *** = sig. at .001*

4) Perceived Specificity: There was a significant difference in the perceived specificity (ΔS, Sect. III-D.0.c) between the iterations for both the individual and iterative designs (one-way ANOVA, $F(12, 288) = 1.83$, $p < .05$; and $F(12, 288) = 2.51$, $p < .01$, respectively). A post hoc pairwise comparison with Bonferroni correction showed a significant difference between the individual designs 6 and 13 ($p < .05$), and between the iterative designs at iteration 8 and iterations 9 and 11 ($p < .01$).

Further, as shown in Fig. 8, the ΔS was significantly higher for the iterative than the individual design at iterations 9 and 11 ($t(21) = -2.37$, $p < .05$; and $t(20) = -3.55$, $p < .01$, respectively); and higher for the individual than the iterative design at iteration 6 ($t(22) = 2.38$, $p < .05$). The average specificity (i.e., ΔS) was significantly higher than 0 for both the individual ($Mean = 0.60$, $SD = 0.48$) and iterative ($Mean = 1.03$, $SD = 0.63$) designs ($t(12) = 4.45$, $p < .001$; and $t(12) = 5.89$, $p < .001$, respectively).

To further analyze the perceived specificity, we compared the parameter-settings of the designs. It was observed that the designs without a chit-chat block for expressing any of the two styles (e.g., the individual designs 4, 8, 11, 12, 13, and the iterative design 8) had the lowest ΔS (Fig. 8), i.e., the lowest distinctions of the perceived task and person orientations between the styles. A low ΔS was also found for the individual design 10 which was the only design having a chit-chat block in both styles. To verify these observations and know how design features impact the perceived specificity, a multiple linear stepwise regression analysis was conducted. We used the presence of a difference in the design parameters that were adopted for expressing the styles (Sect. I), except for the proxemics setting for not being visible in the videos as independent variables. A significant regression was found ($F(1, 24) = 33.39$, $p < .001$), with an R^2 of .564 and chit-chat being the only significant predictor ($t(26) = 5.78$, $p < .001$). The predicted ΔS was found equal to $.083 + .999(ChitChat)$, with $ChitChat$ coded as $1 = Chat$ and $0 = NoChat$. The predicted ΔS increased .999 when chit-chat was used to differentiate the styles. This finding implies that the presence of chat induces a higher perceived specificity of the styles, i.e., the perceived person and/or task orientations of the styles are more distinguishable.

To verify our finding that chit-chat is important to distinguish the communication styles, we performed a second deep analysis. We conducted a multiple linear stepwise regression analysis to predict, based on the presence of a difference in the design parameters used to express the styles (with the exception of proxemics), the magnitude of the t-test results of the style-by-style comparisons. We used the t-test results between the ΔS of all the designs, i.e., paired t-tests between the ΔS at the same iteration and from the same design source (individual or iterative), and non-paired t-tests between the ΔS at different iterations. A significant equation was found ($F(4, 320) = 45.50$, $p < .001$), with an R^2 of .355. The difference in chit-chat was the highest significant predictor ($t(325) = 12.68$, $p < .001$) for predicting the style differences in t-test, as summarized in Table IV and expressed in (2):

$$Y' = 1.217 + 1.175 * x1 - 0.018 * x2 \\ - 0.017 * x3 - 0.212 * x4 \quad (2)$$

with Y' being the predicted t-test value; $x1$ the difference in chit-chat between the styles (coded 0 or 1); $x2$ the difference in voice pitch (Hz); $x3$ the difference in voice speed (words-per-minute); and $x4$ the difference in gaze diversion (coded 0 or 1). The predicted t-test increased 1.175 when the designs differ based on the presence of chit-chat (i.e., difference in chit-chat is 1). In these cases, the designs also differ in the perceived style difference, i.e., if the *person-* and *task-oriented* styles differ in their designs due to the presence of chit-chat, then they also differ more in terms of perceived style. The differences in voice pitch, voice speed and gaze diversion had a lower and inverse effect on predicting the perception of the design differences, i.e., the more different the parameter-settings are, the less they are perceived. This is an indication that the differences in these parameters do not help in differentiating the styles.

The aforementioned findings indicate that the design differences between the styles were perceived by the participants ($H1$), with chit-chat being the most significant factor. However, the perceived specificity of the iterative design did not increase over time and the design converged to a setup that was not perceived as expected. This is an indication that the perception of the iterative design did not converge, which does not support $H2$.

G. Findings

In this study we validated the iterative development method and the multi-modal communication designs for social robots. The perception of a design expressing a *task-* or a *person-oriented* interaction proved to differ for the two communication styles.

The findings imply that the robot was perceived as more animate and likeable in the designs expressing a *person-oriented* style. The perception differences between the styles were further corroborated with the perceived task and person orientations of the designs. The *task-oriented* designs were perceived as more oriented toward the task and less oriented toward the person than the *person-oriented* designs. Further, the use of chit-chat had a significant effect on differentiating the two styles, i.e., the presence of chit-chat induced a higher specificity of the styles. The perceived person and task orientations were more aligned with the intended styles when chit-chat was used for expressing the *person-oriented* style and not the *task-oriented* style (i.e., no chit-chat). The amount of chit-chat also increased the interaction time with the robot.

The iterative designs showed higher perception scores than the individual designs, especially for the perceived animacy and likeability of the robot, satisfaction of the service experience and person orientation of the designs. These findings are an indication that the fine-tuning that the designers did in the iterative step affected the participants' perception. However, the findings show that the iterative design converged to a setup that was not perceived as expected, i.e., the perception of the iterative design did not increase over time.

IV. Discussion

A. Interaction Styles

There was a proven difference in the process and outcome of both the design and perception of the interactions created for the *task-* and *person-oriented* styles. This is in line with social science theory and the proven existence of design differences between the styles (e.g., [9]) and that the differences are perceivable. The refinement (i.e., design) process converged over time and individual designs were harmonized into the iterative design. This is an indication that the provided support (i.e., *ID*-method and tool) helped achieve the desired theory-based different designs.

The design of a *task-oriented* style showed a more formal, shorter and less chatty communication with more close gestures than the *person-oriented*. The recommended design parameters for both styles are summarized in Table V. It seems that spending more time interacting with visitors is expected from a robot with a *person-oriented* style, which is consistent with [10] (i.e., a robot that provides more personalized feedback leads to a longer engagement time with people).

TABLE V

RECOMMENDED DESIGN PARAMETERS FOR EXPRESSING *task*- AND *person-oriented* COMMUNICATION STYLES

	Communication Style	
	Task-Oriented	Person-Oriented
Gesture Openness	Close	Open
Proxemics (m)	< 1	< 1
Speech Acts	Formal and Direct	Informal and Personalized
Chit-Chat	No	Yes

Further, as predicted, a *task-oriented* style was perceived as more oriented toward the task and less toward the person (i.e., compared to a *person-oriented* style), and the robot with a *person-oriented* style was perceived to be more animate and likeable. These findings are consistent with [10] where a robot was perceived as more agreeable when it provided personal feedback than when it adopted a task based approach. An important design parameter making this difference is the amount of chit-chat during the conversation. When the person-oriented style included chit-chat and the task-oriented style did not, this led to a higher perceived specificity (i.e., distinction) between the communication styles. This finding gives interaction designers a validated best practice approach to make interaction style more or less personal.

B. Limitations

The results and findings in this study are based on participants' evaluation of video-based interactions between a user (i.e., visitor) and a receptionist robot, *Pepper*. Participants did not interact with the robot and we were not able to analyze their reactions and behaviors toward the robot. Thus, it is unclear to what extent the robot's embodiment (i.e., physical presence) would affect people's perception of the robot's interaction styles: Will people differentiate the robot's style in real-world interactions similar to video-based interactions? To address this limitation and consolidate the findings, this study is to be complemented with a real-world evaluation of the robot's interactions styles.

Another limitation was the presence of chit-chat in the videos expressing a *person-oriented* style. This caused the interactions in the videos to last longer and may have made the difference between the styles more obvious. This limitation was addressed in the survey by randomizing the order of presenting the videos to the participants, without mentioning the style. Further, having a high number of participants allowed us to gather enough data points to evaluate the designs and compare the differences between the styles.

V. CONCLUSION

This paper presented the use and evaluation of a development method (i.e., the *ID*-method) which consists of a two-step process: an individual design is created in the first step then, subsequently, an iteration is added to the evolving joint design in the second step. The *ID-method* and tool provided the support needed for designers to create and refine *task*- and *person-oriented* styles for social robots.

In conclusion, the design method proved to be useful by providing the desired theory-based multi-modal communicative behaviors. Further improvement of the design tool is needed to support the creation of more complex interaction designs. Future work will extract reusable design patterns, cf. [14]. The perception of the designs differed for the *task* and *person-oriented* styles. The effort spent during the design phase on fine-tuning the design parameters for expressing the styles led to optimizing the parameter-settings for each style. For future designs, the optimal settings can be reused while putting more effort on improving the robot's chit-chat, i.e., being the design parameter which has the most significant impact on perceiving the style differences. Future work includes investigating and comparing the effect of the style designs in video-based versus real-world human-robot interactions.

REFERENCES

[1] Y. Kato, T. Kanda, and H. Ishiguro, "May i help you? - design of human-like polite approaching behavior-," in *ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, Mar 2015, pp. 35–42.

[2] K. Mizumaru, S. Satake, T. Kanda, and T. Ono, "Stop doing it! approaching strategy for a robot to admonish pedestrians," in *ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, Mar 2019, pp. 449–457.

[3] H. Hewitt, L. McCloughan, and B. McKinstry, "Front desk talk: discourse analysis of receptionist–patient interaction," *British Journal of General Practice*, vol. 59, no. 565, pp. e260–e266, 2009.

[4] P. M. Neuwelt, R. A. Kearns, and I. R. Cairns, "The care work of general practice receptionists," *Journal of primary health care*, vol. 8, pp. 122–129, June 2016.

[5] J. Ward and R. Mcmurray, "The unspoken work of general practitioner receptionists: a re-examination of emotion management in primary care," *Social science & medicine*, vol. 72, pp. 1583–7, March 2011.

[6] M. Alazri, P. Heywood, and B. Leese, "How do receptionists view continuity of care and access in general practice?" *European Journal of General Practice*, vol. 13, no. 2, pp. 75–82, 2007.

[7] J. Xu, J. Broekens, K. Hindriks, and M. A. Neerincx, "Mood expression through parameterized functional behavior of robots," in *2013 IEEE RO-MAN*, Aug 2013, pp. 533–540.

[8] R. Peters, J. Broekens, and M. A. Neerincx, "Robots educate in style: The effect of context and non-verbal behaviour on children's perceptions of warmth and competence," in *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, Aug 2017, pp. 449–455.

[9] E. Saad, J. Broekens, and M. A. Neerincx, "An iterative interaction-design method for multi-modal robot communication," in *IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 2020, pp. 690–697.

[10] S. Zafari, I. Schwaninger, M. Hirschmanner, C. Schmidbauer, A. Weiss, and S. T. Koeszegi, ""you are doing so great!" – the effect of a robot's interaction style on self-efficacy in hri," in *2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, Oct 2019, pp. 1–7.

[11] C. McGinn and I. Torre, "Can you tell the robot by the voice? an exploratory study on the role of voice in the perception of robots," in *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, March 2019, pp. 211–221.

[12] 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, Jan 2009.

[13] C. Froehle, "Service personnel, technology, and their interaction in influencing customer satisfaction," *Decision Sciences - DECISION SCI*, vol. 37, February 2006.

[14] A. Sauppé and B. Mutlu, "Design patterns for exploring and prototyping human-robot interactions," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. New York, NY, USA: Association for Computing Machinery, 2014, p. 1439–1448.