2022 31st IEEE International Conference on Robot and
Human Interactive Communication (RO-MAN)
August 29 - September 2, 2022. Naples, Italy.

2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN) | 978-1-7281-8859-1/22/$31.00 ©2022 IEEE | DOI: 10.1109/RO-MAN53752.2022.9900771

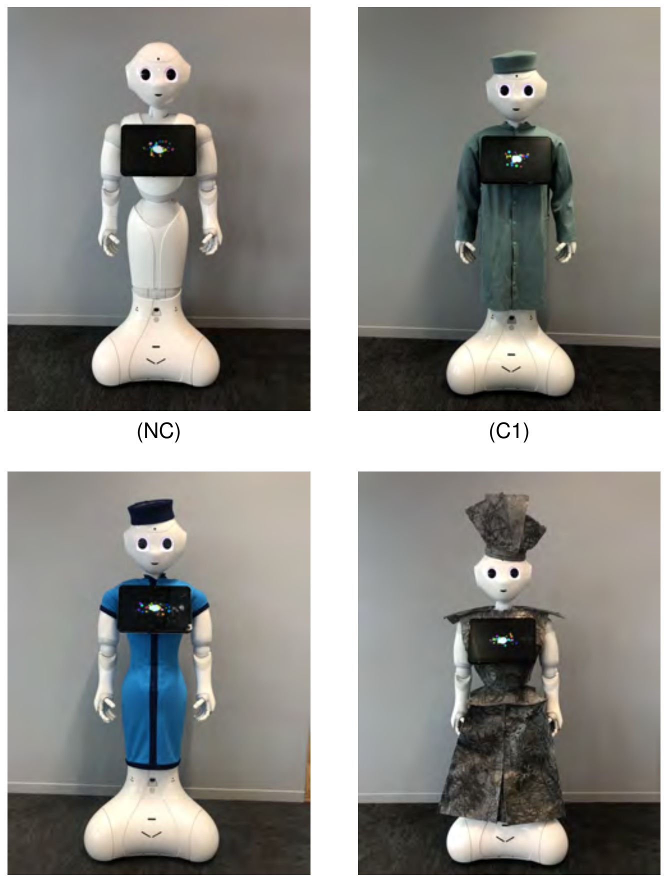
Effects of Robot Clothing on First Impressions, Gender, Human-Likeness, and Suitability of a Robot for Occupations

Koen V. Hindriks[1] and Marijn Hagenaar[1] and Anna Laura Huckelba[1]

***Abstract*— Clothing is often used as a way to communicate about one's identity. It can provide information about characteristics such as gender, age, occupation, and status as well as more subjective attributes such as trustworthiness and friendliness. In this paper we report on exploratory research that we conducted by means of three studies to investigate the effects of clothing on a Pepper robot. Three outfits that varied in style and colour were compared with each other and the robot without clothing. Findings from an online survey suggest that gender perception but not human-likeness can be manipulated by clothing. We also found that first impressions on expertise and likeability may vary with clothing style and may induce stereotypical job associations. In a second study, we interviewed experts which were all familiar with the Pepper platform to obtain a more qualitative perspective on robot clothing. The interviews made clear that specific features of clothing may trigger strong associations with, for example, occupations, and that features such as headdress may make an outfit look more like a uniform. Finally, we conducted a field experiment comparing two clothing conditions for a receptionist robot in a natural setting and found that a robot with uniform appears to be more engaging. Our findings suggest that robot clothing may have an effect on first impressions, gender perception, and engagement in interactions with users in the wild.**

I. INTRODUCTION

For humans, clothing is a form of non-verbal communication [1] which "functions as an effective means of communication during social interaction" [2]. This means that clothing can be used to encode messages for others to interpret and to form impressions about another. Johnson et al. [1] found that people use appearance and dress to form impressions about another's age, ethnicity, gender, social position, occupational role, attitudes and emotions, intentions or motives, and personality traits. Clothing is a means to express oneself [3] and in conveying identity [2]. Roach-Higgins and Eicher [2] note that "dress has a certain priority over verbal discourse in communicating identity since it ordinarily sets the stage for subsequent verbal communication". Dress may also affect behavioural responses [4]. It has been found, for example, that experimenters were obeyed more often when wearing a uniform than when not wearing a uniform. Dress may have an effect on disclosure [4] and on the treatment a customer service employee receives [5]. It is well-known that the appearance of robots plays a role in the interaction between humans and robots. The shape and form of a robot has an effect on the perception of human-likeness [6] and appearance also influences gender perception [7] of a robot. Appearance, however, is a broad term that includes features of the undressed body, such as its shape and colour, as well as expression through gesture and grimace [2]. Our focus is on the effects of robot *clothing*, which we use conform [2] to refer to enclosures that cover the body. A difference between humans and robots is that we can study the effects of wearing clothing versus not wearing any clothing.



(NC) (C1) (C2) (C3)

Fig. 1: Robot clothing conditions: (NC) No clothing, (C1) Green outfit, (C2) Blue outfit, (C3) Grey outfit

We asked a professional designer to design three new outfits for Pepper. The outfits differ in colour but also vary in style. In discussions with the designer aspects related to gender, and role or occupation were taken into account. The resulting outfit designs are shown in Figure 1. The C1 outfit was specifically designed to influence gender inferences. It is not clear how clothing will influence gender perception, but we know from literature that body shape is a factor in

[1]Vrije Universiteit Amsterdam, Social AI group, The Netherlands k.v.hindriks@vu.nl

how people infer gender. More specifically, Bernotat et al. [7] showed that shoulder width and waist-to-hip ratio play a big role in how people perceive the gender of a robot. Because we can use clothing to (de)emphasise certain parts of a robot's body. The width of the clothing in C1 makes the body shapeless, which makes the waste invisible and focuses more on the shoulders. This could make the robot look more male-like. It was also thought that the shape and a more neutral colour such as green would be more appropriate for robot roles in healthcare, whereas the blue outfit C2 was thought to be more appropriate for service roles such as a receptionist [8]. In contrast to the green outfit, C2 emphasises the hips which could make the robot appearance more feminine. The outfit C3 was designed to provide the robot with a more extraordinary and unusual outfit for an "out-of-everyday" social role or activity that would likely draw more attention ("dress to impress"). We thought this outfit might be fitting for a particular restaurant or theatre setting where such an appearance would stand out. For all outfits, the designer also created matching headdress.

There is still little known about how a robot that wears clothing affects the perception of and interaction with that robot. In a recent study, Hurtienne and Arnold [9] report that clothing did not affect the user's experience of the robot at all. As it seems likely, however, that robot clothing could potentially have many of the same effects as clothing has for humans, our research is aimed at further exploring the effects of robot clothing. Our work does not focus on clothing design for robots specifically, but rather on exploring *whether* robot clothing may influence first impressions (e.g., expertise), perception of gender, human-likeness, and the suitability for a particular role or occupation. Our approach to study the effects of robot clothing has been to broadly survey these aspects first using an online questionnaire (Section III). Johnson et al. [1], however, also encourage researchers interested in the role of appearance and dress on impression formation to allow participants to freely express their thoughts about the meanings of dress and appearance. Inspired by their work we also conducted a structured interview study to collect free-response data and more qualitative insights about the influence of robot clothing from experts from various domains (Section IV). Finally, as it is very important to evaluate expected results also with robots in context, in a third study that builds on the findings of the previous two we conducted a field experiment. In this study we investigate whether any behavioural responses of clothing can be identified by means of a use case in a natural setting (Section V).

II. Related work

Studies on human clothing reveal that clothing affects people's first impression and how they perceive others [10]. For example, more masculine attire is mostly associated with traits such as dominance and power whereas more feminine attire is perceived as more friendly and likeable [10]. In order to explore first impressions on expertise, trustworthiness, social dominance, likeability and friendliness, we include similar survey items as have been used before for studies on human clothing [10] in our survey on robot clothing.

It is questionable if the clothes of a robot have to be the same kind of clothes as we wear as humans. Robot clothing should fulfil needs robots have, rather than just being human clothes on a robot [11]. Robot clothes need to meet different requirements than human clothing related to factors such as shape and used materials. But putting clothes on a robot could still cause people to anthropomorphise a robot more. If that would be the case, it may be less desirable to make robots wear clothing. For this reason we also investigate whether clothes would make a robot more human-like.

We know from previous work that appearance is significant for how people perceive the suitability of a robot for a particular application, or occupational field [12], [13]. Robot clothes need to be adapted to the context and task that it performs, protect the robot and signal its function [11]. Clothing may signal the role of the robot and the tasks that it could perform and thereby make the robot more transparent for its users. Bryant et al. [14] asked participants in their study how likely they think that the robot could perform certain tasks that are needed for occupations such as a home health aid, news anchor, receptionist and security guard and evaluated whether certain occupations were seen as more fit for a male or female robot. We also look into whether robots that perform certain roles are seen as more competent if their appearance (including their clothing) would better match respective gender stereotypes, cf. [15].

III. Survey: First Impressions

In an online survey we explore how first impressions, gender perception, human-likeness rating, and suitability for occupations is influenced by robot clothing.

A. Method

1) Participants: 100 participants were recruited from European countries (Austria, Netherlands, Spain, Belgium, Denmark, Finland, France, Italy, Germany, Luxembourg, Slovakia, Slovenia, Sweden and Switzerland) via Amazon Mechanical Turk (Mturk). We selected these countries because within Europe they have the highest robot density [16]. Moreover, because the field experiment was performed in one of these countries, we wanted to avoid effects due to differences in cultural background. HIT (Human Intelligence Task) approval for participants was required to be 95% or higher and the number of HITs needed to exceed 5000. Participants were rewarded with a payment of \$1.50. Of all participants, 24 were asked to rate items for condition NC (the picture labelled as such in Figure 1), 25 for condition C1, 27 for condition C2, and 24 for condition C3. Gender division was Male (60%) and Female (40%). Age ranges were divided as follows: 18-24 years old (18%), 25-34 (40%), 35-44 (30%), 45-54 (8%), 55-64 (3%) and 65+ (1%). 95% of the participants had never seen Pepper before.

2) Design and Materials: To measure first impressions, we used various scales from the literature, inspired by [10]: An expertise scale adapted from [17] consisting of 5 semantic differential items ($\alpha = .94$), a trustworthiness scale adapted from [17] consisting of 5 semantic differential items ($\alpha = .82$), a social dominance scale adapted from [18] consisting of 5 semantic differential items ($\alpha = .83$), nine Likert-scale items from the Reysen likeability scale [19] ($\alpha = .89$), and a friendliness scale taken from [20] consisting of 5 Likert-scale items ($\alpha = .79$). To measure gender, we used a semantic differential item *male-female* (1 = male, 5 = female) and a naming task which asked participants to provide a name for Pepper (0 = male, 0.5 = unisex, 1 = female), see also e.g. [21]. To measure human-likeness, we used a semantic differential item *not human-just like a human* often used in the literature [6], 4 items from the anthropomorphism scale from [22] ($\alpha = .80$), and five yes/no questions from the anthropomorphism scale from [23].[1] If not indicated otherwise, all items were rated on a 7-point scale.

3) Procedure: People were directed from Mturk to Qualtrics in which they completed the survey. Participants were first asked for their consent to participate in the survey. Then participants were asked to complete the items of the various scales used. At the end of the survey they were asked a few general questions about demographics. We also asked participants to indicate the colour of the robot clothing (white if no clothing) in the picture presented to them (corresponding to the condition of each participant) to check whether they took the questionnaire seriously.

B. Results

Table I summarises the results and presents the means and standard deviations for the various scales and items.

First Impressions. A one-way ANOVA revealed that there was a significant difference in mean expertise score $F(3,96) = 6.19$, $p = 0.001$ with a large effect size $\eta^2 = 0.16$), in mean trust score $F(3,96) = 3.58$, $p = 0.02$ with medium effect size $\eta^2 = 0.10$), and in mean likeability $F(3,96) = 3.19$, $p = 0.03$ with a medium effect size $\eta^2 = 0.09$). Tukey's HSD tests for multiple comparisons found that the mean value for expertise was significantly different between NC and C3 ($p < 0.001$, 95% C.I. = $[-2.44, -0.57]$) and C2 and C3 ($p = 0.03$, 95% C.I. = $[-1.88, -0.63]$) and for likeability between NC and C1 ($p = .04$, 95% C.I. = $[0.01, 1.58]$) but did not find any significant differences between conditions for trust although the difference between C2 and C3 is nearly significant at $p = 0.05$ and between C1 and C3 is significant at the $p < 0.1$ level.

Gender. We performed a Welch ANOVA as Levene's test indicated unequal variances for both gender and name. This revealed a significant difference in gender $F(3,96) = 5.92$, $p = .001$ with large effect size $\eta^2 = 0.16$ and in name $F(3,96) = 6.92$, $p < .001$ with large effect size $\eta^2 = 0.18$ (results are equal for a one-way ANOVA). Both gender and name were found to strongly correlate $r(98) = 0.86$, $p < .001$. Games-Howell tests for multiple comparisons found that the mean value for gender was significantly different between conditions C1 and all three other conditions (NC: $p = .02$, 95% C.I. = $[0.22, 2.56]$, C2: $p < .05$, 95% C.I. = $[0.01, 2.46]$, C3: $p < .001$, 95% C.I. = $[0.88, 3.56]$) and for name between conditions C1 and NC ($p = .01$, 95% C.I. = $[0.07, 0.60]$ and C1 and C3 ($p < .001$, 95% C.I. = $[0.24, 0.81]$).

Human-likeness. We found no significant differences on any of the three scales used for human-likeness. All three variables Human-likeness, Anthropomorphism 1 and 2 were found to strongly correlate: Human-likeness and Anthropomorphism 1 $r(98) = 0.69$, $p < .001$, Human-likeness and Anthropomorphism 2 $r(98) = 0.51$, $p < .001$, Anthropomorphism 1 and 2 $r(98) = 0.58$, $p < .001$.

Occupations. One-way ANOVA only showed a significant difference for Personal Assistant $F(3,96) = 3.96$, $p = 0.01$ with a medium effect size $\eta^2 = 0.11$. Tukey's HSD test showed that NC and C3 ($p = 0.03$, 95% C.I. = $[0.07, 1.77]$) and C2 and C3 ($p = 0.02$, 95% C.I. = $[0.14, 1.79]$) are significantly different. It is noticeable that both NC and C2 are nearly always rated very close to or above (total) average on all occupations. We used 3.50 (70% on a $1-5$ range, which is a good indication that the robot on average is perceived as suitable for the job) as a cut-off point to further inspect the ratings for occupations. If on (total) average an occupation is rated on or above 3.50, then this is also true for C2, NC (with the exception of Household tasks with a rating of $3.42 < 3.50$), as well as almost always true for C3 (with the exception of Tour-guide with a rating of $3.42 < 3.50$ and of Personal Assistant with a rating of $2.96 < 3.50$). This is not true for C1 which is only perceived as more than suitable relative to the cut-off point for Receptionist and Tour-guide. Finally, note that all clothing conditions are perceived as more than suitable only for Receptionist and almost so for Tour-guide.

C. Discussion

Overall, there is a very consistent pattern that mean scores for first impressions for NC and C2 are higher than those for C1 and C3. Scores for NC and C2 are very similar except perhaps for expertise but this difference is not significant. It is clear that C1 is viewed as more male and C3 as more female than any of the other clothing conditions. Clothing thus can have a significant effect on gender perception. Some items suggest stereotyping. It is noticeable, for example, that C3, which is considered more female, most of the time scores lower than the other clothing conditions on occupations but is rated higher on Household tasks. A very similar pattern as for first impressions is found for human-likeness with NC and C2 both scoring higher than C1 and C3. The first Anthropomorphism (1) scale shows a similar pattern but the differences are smaller. Interestingly, the pattern for the second Anthropomorphism scale is somewhat different, lightly increasing from NC to C3. The main difference seems to be that participants believe that Pepper in the grey outfit (C3) is able to enjoy something more than in the other outfits, which is in line with the more festive look of this outfit. We can conclude that putting clothes on robots does not

[1] We selected the following questions, all using the format 'Do you think that Pepper...', (1) Can be sad? (2) Remember something? (3) Can think? (4) Understands when you say something? (5) Can enjoy something?

TABLE I: Average Scores on scales for First Impression, Gender, Anthropomorphism, and on items for Occupations

Construct	Total *M*	Total *SD*	No Clothing (NC) *M*	No Clothing (NC) *SD*	Clothing C1 *M*	Clothing C1 *SD*	Clothing C2 *M*	Clothing C2 *SD*	Clothing C3 *M*	Clothing C3 *SD*	ANOVA $F(3,96)$	ANOVA p	ANOVA η^2
Expertise	4.59	1.33	5.30	1.12	4.49	1.38	4.76	1.07	3.79	1.36	6.19	.001	0.16
Trustworthiness	5.05	1.09	5.36	1.12	4.76	0.89	5.41	1.02	4.65	1.17	3.58	.02	0.10
Social Dominance	3.54	1.19	3.66	1.07	3.42	1.42	3.73	1.13	3.34	1.13	0.61	.61	0.02
Likeability	4.58	1.08	4.93	0.91	4.13	1.10	4.84	0.85	4.40	1.30	3.19	.03	0.09
Friendliness	5.26	1.03	5.46	1.04	5.08	1.02	5.41	0.76	5.06	1.26	1.06	.37	0.03
Gender	3.52	2.00	3.71	1.88	2.32	1.03	3.56	2.12	4.54	2.21	5.92	.001	0.16
Name (coded; *[0-1]*)	0.34	0.44	0.40	0.44	0.06	0.17	0.32	0.46	0.58	0.48	6.92	.000	0.18
Human-likeness	2.64	1.59	3.00	1.45	2.44	1.29	2.89	1.63	2.21	1.91	1.36	.26	0.04
Anthropomorphism 1	3.28	1.28	3.29	1.32	3.08	1.22	3.37	1.27	3.35	1.35	0.27	.85	0.01
Anthropomorphism 2 *[0-5]*	2.41	1.17	2.29	1.30	2.36	1.19	2.41	1.05	2.58	1.21	0.27	.85	0.01
Home Health Aide	**3.56**	1.08	**3.54**	1.18	3.36	1.22	**3.81**	0.88	**3.50**	1.02	0.82	.49	0.03
Security Guard	3.00	1.29	3.42	1.06	2.80	1.38	3.00	1.18	2.83	1.49	1.17	.33	0.04
Teacher	3.28	1.26	3.33	1.44	3.24	1.33	3.41	1.22	3.13	1.12	0.23	.88	0.01
News Anchor	3.17	1.22	3.42	1.44	2.96	1.24	3.37	1.01	2.92	1.14	1.18	.32	0.04
Receptionist	**4.06**	0.95	**4.13**	1.12	**4.16**	0.99	**4.19**	0.83	**3.75**	0.85	1.14	.34	0.03
Household tasks	**3.52**	1.21	3.42	1.25	3.20	1.26	**3.56**	1.01	**3.92**	1.28	1.53	.21	0.05
Package deliverer	3.21	1.33	3.08	1.28	2.92	1.41	3.41	1.28	3.42	1.35	0.86	.47	0.03
Therapist	2.28	1.20	2.75	1.29	2.28	1.31	2.07	0.96	2.04	1.16	1.86	.14	0.06
Personal Trainer	2.83	1.24	3.13	1.26	2.76	1.33	2.67	1.07	2.79	1.32	0.64	.59	0.02
Tour-guide	**3.67**	1.11	**3.67**	1.20	**3.60**	1.08	**3.96**	0.85	3.42	1.28	1.08	.36	0.03
Personal Assistant	**3.56**	1.18	**3.88**	1.08	3.44	1.12	**3.93**	0.87	2.96	1.40	3.96	.01	0.11
Restaurant Server	**3.59**	1.16	**3.71**	1.16	3.16	1.18	**3.85**	0.91	**3.63**	1.35	1.72	.17	0.05
Sales Agent	2.63	1.22	2.83	1.27	2.28	1.17	2.78	1.25	2.63	1.17	1.04	.38	0.03

All occupation items were rated on a 5-point Likert-scale while all other scale items were rated on a 7-point Likert scale except where ranges [...] are explicitly indicated. Occupation ratings in bold score relatively high exceeding or being equal to the cut-off point 3.50 *that we used.*

make them appear more human-like. Such an effect might have been a reason to avoid the use of robot clothing, but for the clothing designs that we used in our studies this does not seem to be the case. Finally, C2 scores better on suitability for most occupations except for Security Guard, Therapist, and Personal Trainer, where NC scores better. The Receptionist and Tour-guide occupations are rated very high for all conditions and stand out as the most suitable occupations for the Pepper robot. The fact that the green outfit only scores higher than the cut-off point for 2 out of 13 occupations suggests that outfits can trigger stronger associations with specific roles.

IV. Interviews: Perspectives of Experts

We conducted interviews with experts to obtain more qualitative insights on the effects and interpretation of robot clothing. Our aim here is to improve our understanding of the findings of the survey and also to collect any unanticipated new insights on robot clothing.

A. Method

1) Participants: Several experts (N = 7), 5 Female and 2 Male within the age range of 25-60 years old, were asked to participate in a short online interview by contacting them directly by email. All experts who participated were known to the researchers and/or had been previously involved in deploying a Pepper robot in a real world use case. The participants consisted of two interviewees with healthcare expertise, one with marketing expertise in the cultural sector, one with expertise in the fashion industry, one with educational robotics expertise, one with expertise in hospitality robotics, and one with expertise on robots and theatre.

2) Design: We were interested in the appearance cues that are used to infer information and form impressions of a robot with or without clothing. To extract relevant information from the interviews, we used ATLAS.ti 9 and a subset of categories deemed relevant and applicable for our research from [4] for coding the interviews. The categories used were (1) behaviours, activities, and interests, (2) body form, (3) body motion, (4) body surface, (5) character traits, (6) clothing, (7) demographic information, (8) garment aesthetics, (9) level of attention seeking, and (10) social role.

3) Materials: A PowerPoint presentation was used to structure the interview, to display pictures of the four robot conditions (Fig. 1), and to present some of the Likert-scale items from the survey used in the first study (we included expertise, trust, dominance, friendliness, and human-likeness).

4) Procedure: Interviewees were asked for their consent to record the interview and informed that the interview is about first impressions related to robot clothing. They were told they could respond freely and comment on anything they wanted to share during the interview. They were then briefly asked about their experience with robots. The interview then continued with presenting each of the four conditions in order (NC, C1, C2, C3) and the interviewees were asked about their opinion of each of the four robot clothing conditions. As participants of the survey, for each condition, they were also asked to rate their impressions for expertise, trust, dominance, friendliness, and human-likeness on Likert-scale items. Gender was not rated but included as a topic in the interview questions.

B. Results

Interviews took between 19 and 36 minutes with an average length of 29:39 minutes. We used Amberscript to automatically process the audio recordings of the interviews. The quality of the resulting transcripts was sufficient for our coding purposes, although we needed to correct the transcript at several places.

All three outfits triggered strong associations with at least some of the participants. The first outfit (C1) triggered strong associations with a "liftboy" - the exact word used by 2 interviewees. They explicitly mentioned this type of job for this outfit or with the military for 4 other interviewees because of the particular green colour of this outfit. The second outfit (C2) was strongly associated with stewardesses and/or the Dutch airline company KLM for all 7 interviewees because of the specific type of blue that is also used by the dress used by this airline. The third outfit (C3) was associated with Asia for 4 interviewees, which according to one also would draw more attention. As is clear from this discussion, the colour of an outfit can raise various associations and it was often mentioned (17x). Several suggestions were made for changing it. For example, changing C1 from green to white for use in healthcare, or changing C2 from blue to red to remove the association with the airline. Finally, one attribute of the clothing, the headdress, was mentioned explicitly (12x) by 5 of the 7 interviewees. Especially the headdress of C1 triggered associations with formal uniforms, and the military. The headdress of C3 triggered an association with Asia (1x) and it was argued that by removing it the clothing would become more suitable for use in an application context (1x).

Apart from the more general service categories of greeting (2x), providing (8x) and collecting information (4x), and entertaining (6x), interviewees also mentioned more specific tasks for a social robot such as exercise training, and keeping companionship. A role for social robots in healthcare was also often mentioned (17x), sometimes explicitly associated with outfits C1 (2x) and C2 (1x). Interviewees indicated that social robots would be useful in schools (2x; Teacher), as a hospitality robot (16x; Receptionist), and in shops (3x; Sales Agent). Outfit C1 was associated in particular with policing tasks and the military (Security guard). As discussed above, outfit C2 triggered particularly strong associations with stewardesses. Outfit C3 was associated most with entertainment (5x), museums (1x), and theatre (2x). It was also associated with (Asian) restaurants (4x) and taking orders (Restaurant Server) and as a travel guide (Tour guide).

All interviewees indicated that the Pepper robot had a friendly (5x) or cute (3x) appearance, referring to the eyes (3x), face (2x), and height (1x) of the robot. Pepper also scored highest on friendliness in the survey and likewise no significant differences between conditions were found. Human-likeness was rated only slightly higher for a robot with clothes than without, which is similar to what we found in the survey. Only one interviewee indicated that the clothing made the robot look substantially more human-like. Most interviewees rated C1 as somewhat more human-like than NC, where some (2x) argued this is because the clothing covers more of the robot arms. All interviewees thought C1 to be more male, C2 to be more feminine, and C3 to be even more feminine than NC, which, except for C2, is in line with the survey findings. Some indicated that the "broad shoulders" (2x) of the robot suggested a more male and the "hips" (1x) a more female figure (cf. also [7]). One of the interviewees suggested that removing the headdress of C1 would make it less male.

C. Discussion

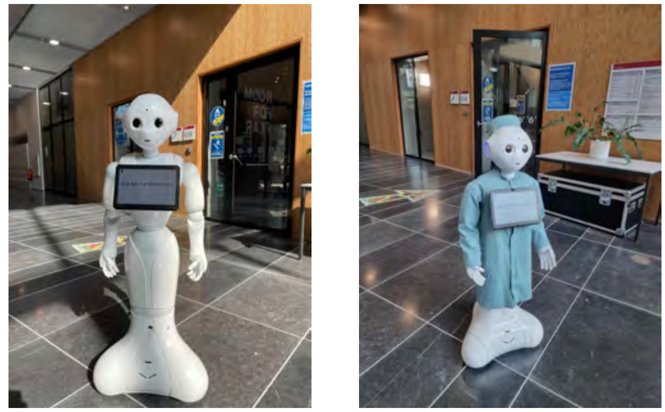
The interviews yielded several new insights. Our outfits did not have particularly strong effects on dominance and friendliness. The non-clothed robot is rated significantly higher on expertise (only in the survey) and also on trust (both in the survey and by the experts). As 95% of our survey participants indicated that they were not familiar with the Pepper robot whereas all our experts were, higher ratings on expertise initially may be adjusted after becoming more familiar with a robot. This is different for trust where the first impression based on appearance seems to remain stable over time but more research is needed to support this finding and establish why. How much a robot's body is covered by clothing may also have an effect on human-likeness.

Many occupations listed in the survey were also mentioned during the interviews except for home health aide, news anchor, household tasks, package deliverer, therapist, and personal assistant. We think that most of these (e.g., news anchor) were simply not explicitly mentioned because the job label is very specific while associated tasks such as providing information or more generic job labels such as healthcare assistant were mentioned during the interviews as an appropriate activity for a social robot. We speculate that household tasks and package deliverer were not mentioned because all interviewees were familiar with the capabilities of the Pepper robot and recognise that these jobs are not within the capabilities of this robot platform. Although we do not think the Pepper platform is particularly suitable as a security guard either, this job was explicitly mentioned for one of the outfits, which highlights the effects that specific design features (colour, headdress) of an outfit can have.

It is clear from our results that clothing can affect inferred information such as gender. By hiding specific robot body shapes by clothing such as C1 does a more female gender identification of NC may be changed into a more male perception. An outfit like C3 that emphasises the hips can 'compensate' for the broader shoulders of a robot to suggest an even more feminine appearance. It appears that certain design features such as colour and headdress can have stronger effects than anticipated and need to be more carefully taken into account when designing robot clothing. Headdress was associated with a uniform, which may suggest a more formal role. Overall, these insights confirmed and framed the findings from Study 1 as well as shed new light on the importance of headwear. Moreover, it helped us select what appear to be the most contrasting outfits for the two conditions in our 3rd and final study.

V. Use Case: Clothing for a Robot Receptionist

In our third study we aimed to explore the effects of clothing on engagement with a robot. We chose to compare the robot without clothing (NC) and with the green outfit (C1) in a field experiment. Our choice to compare these two conditions was motivated by several findings reported above. First, based on the survey findings, conditions NC and C2 are overall very similar and we did not expect to find any differences in a field experiment between these conditions. Second, the NC and C1 conditions were found to differ significantly on likeability as well as on gender (the green outfit is perceived as more male than all other conditions). Finally, we speculated that the associations several experts had for outfit C1 with a liftboy (section IV) would make it particularly well-suited for a receptionist job. Because of these associations, this uniform-like outfit might make it more transparent why the robot would be positioned at a building entrance and raise expectations that it could deliver a service. As both conditions were rated very suitable for performing a receptionist task and this occupation was rated highest amongst all other occupations (see Table I), we choose to implement one of the tasks of this occupation for our field experiment. We designed a simple use case for helping people finding directions in a campus university building. To further inform our design and identify what type of questions passersby would ask the robot we conducted a small pilot study where we controlled the robot remotely (WoZ style) to collect that information. The robot we used in the study operated fully autonomously and was also able to provide directions to some other locations nearby on the campus. On the robot's tablet in very large font the text "Ask Me For Directions" was displayed. The interaction flow we designed enabled the robot to greet passersby that it detected automatically using its cameras, and, once interaction with someone was initiated, a variety of questions could be asked, including, for example, directions for study rooms, the cinema, or the restaurant inside the building. Thereafter, whether or not the robot was able to address a question, it would ask the person it was interacting with to rate its performance ("How did I do?") on a 7-point scale with smiley symbols ranging from a frown to a smiling face.



(NC) (C1)

Fig. 2: No clothing (NC) and green uniform (C1) conditions for receptionist use case in a university building

A. Method

1) Participants: We observed 2440 people that entered the building and were counted as participants. We included students but excluded employees as we assumed that they would already know their way around the building and would only interact with the robot to try it out. Sample size for both conditions was about equal (1338 for NC condition and 1102 for C1 condition). Of the 2440 participants, 1077 were labelled male, 1338 female, and 25 were labelled as 'other'.

2) Design: We performed a field experiment in a university campus building. The independent variable was the clothing that the robot was wearing. The robot was either not wearing any clothing (NC condition) or it was wearing a green uniform with cap (C1 condition), see also Fig. 2. The dependent variables included the ignore rate (percentage of people that entered the building but did not interact with the robot and simply passed by), the dropout rate (people leaving before they rated the robot), and ratings of how well the robot did on a 7-point smiley scale. To check whether our use case design was not a differentiating factor for the two robot conditions we also analysed the robot's ability to recognise locations people were asking for. Ethical approval for the study was obtained through the Ethical Board procedure of the university faculty of the researchers. The robot executed the navigation assistance use case autonomously. In other words, interactions were not controlled by a "wizard." An observer was present only for logging observations and handling technical issues with the WiFi connection we used. We wished to conduct a field study with high ecological validity, which did not provide us with any control over the study population nor did it allow us to use any self-report measures, as then the purpose of the field study would likely have been revealed to other participants (since this was mainly a student population).

3) Materials: We used the Pepper robot designed by SoftBank Robotics, version 1.8 and Google Dialogflow to enable speech recognition for identifying the directions participants were asking for. For logging observations, we used a simple web interface with buttons to create records for type of person (we tried to differentiate students from employees based on estimate of their age) and sex (we tried to differentiate males and females but also had an 'other' button for cases where the observer was not sure).

4) Procedure: A Pepper robot was placed near one of the entrances inside the building over the course of five days for about $2-3$ hours each day. An observer was always present, who was situated at least 10 meters away from the robot which we judged to be sufficient for hiding the fact that the robot was observed. The observer's task involved logging information using the web interface and to resolve any technical difficulties (e.g., with the network) when they arose. Only people that entered the building were counted as participants while people that left the building and interacted with Pepper were excluded to keep the procedure for computing the ignore rate clear and simple.

B. Results

9.7% of people walking into the building initiated an interaction in the NC condition (90.3% ignore rate) whereas 13.9% of people did so in the C1 condition (86.1% ignore rate). A chi-square test of independence showed that there was a significant association between clothing condition and ignore rate ($\chi^2(1, N = 2440) = 9.56$, $p = .02$) with a small to medium effect size ($w = 0.19$). We did not find that clothing was associated with dropout ($\chi^2(1, N = 153) = 0.29$, $p = .59$). Dropout happened about 10% of the time in both conditions. More males than females interacted with the robot in the C1 than the NC condition, but no significant association between sex and clothing condition was found. We noticed that people tried to test the capabilities of the robot. They did so in various ways ranging from where-questions unrelated to the building (e.g., “Where is the Sahara?”), open-ended requests (e.g., “pick a room anywhere”) to attempts to see how the robot would respond to insensible phrases (e.g, “website happy space”) and the one or two insults the robot got (e.g., “you suck”). We found testing behaviour depended on clothing condition, as significantly more testing behaviour was found in the C1 condition (7%) compared to the NC condition (4%; $\chi^2(1, 2440) = 5.59$, $p = .01$), with a small effect size ($w = 0.11$).

We compared the robot’s performance in both conditions using the transcripts from the speech recognition module. In our analysis we excluded testing behaviour to ensure performance was only compared for interactions that were in scope of the use case we designed. In particular, we analysed how often a place someone was asking directions for was or was not recognised. We found no difference related to this performance metric ($\chi^2(1, 153) = 0.35$, $p = .56$). We found above average ratings for the NC ($M = 5.12$, $SD = 2.49$) and C1 ($M = 5.27$, $SD = 2.23$) condition but did not find a significant difference on how the robot’s performance was rated by participants, $t(137) = -0.388, p = .08$. Ratings were significantly higher though when the robot could successfully give an adequate response to a question ($M = 5.85$, $SD = 2.03$) compared to when it could not ($M = 4.20$, $SD = 2.48$), $t(137) = 4.27$, $p =< .001$.

C. Discussion

Our findings suggest that the robot with green outfit (C1) is more engaging than the robot without clothing (NC) as more participants initiated interactions in that case. We expected this based on our findings from the interviews where experts associated this outfit with a liftboy and indicated it looks like a uniform which might more strongly suggest that the robot would be able to deliver a service. We cannot exclude, however, that our results may also have been partly due to a novelty effect as the outfits were new for a robot that may have been familiar to at least some of our participants.

Participants initiated also more testing behaviour when the robot was wearing the green outfit compared to when it did not wear clothing. We speculate that this may also be related to the green outfit being perceived as a uniform, which may have raised expectations. It may be that participants perceive a robot in a uniform as more competent and therefore expected it to be able to answer more complex questions [24]. We do not see any of this reflected in the performance ratings, however. Gender perception may also have been a factor here. In our survey results (Section III) we found that the robot wearing the green outfit is perceived as more male than any of the other clothing conditions. This perception may also have been associated with more competence, an association which has been reported more often in the literature (e.g., [25], [26]).

Finally, we did not find any differences related to the execution of the use case. All measures related to how well the robot performed its task (giving directions) were similar for both conditions, including speech recognition performance (locations recognised), dropout behaviour, and ratings of how well the robot did. We can conclude from this that performance in both conditions was similar and appears unrelated to any of the differences that we found.

VI. Limitations and Future Work

Our studies were exploratory in nature and have several limitations. We discuss some of the more important ones. First of all, we only compared 3 very specific outfits that were green, blue and grey and a condition without clothing and used a single robot platform. All outfits used in our study are perhaps rather formal. Formal clothing is associated with competence (and expertise varied between conditions in our survey study) whereas casual clothing is associated with friendliness (for which we found no significant difference between conditions) [27]. It may be interesting therefore to explore more casual robot clothing with different colours in future work. In our first survey study the sample was small with only 100 participants. In addition, all participants were European citizens, so it is possible that these findings are not applicable in other cultures [28]. Future work should include other regions and cultural backgrounds. The experts we interviewed all had relevant views on some of the aspects related to clothing that we were interested and expertise on both application domain and robotics was present. In future work, more experts could be consulted to include more perspectives on different applications, robotics, and fashion. We chose to conduct a field experiment but such a design allows only very limited control. In particular, we have treated our sample as independent even though it is possible that participants interacted multiple times with the robot and we know for sure that this happened at least once as someone asked the robot why it was not wearing clothes that day. As far as we know, ours is the first field experiment to examine the effect of clothing on a receptionist robot that is autonomously executing a use case in a natural setting. Our use case, however, was small, and future work should explore the effects of clothing for other tasks and contexts. More work is also needed to explore which outfits are more suitable for particular contexts. Another important question that remains for future work is how effects of clothing are different for humans and robots.

VII. Conclusion

We conducted three studies to investigate the influence of clothing for social robots. Three outfits were designed which varied in style in order to compare these with a robot without clothing. Gender and role or occupation were also taken into account when designing the outfits. The findings of the survey suggest that gender perception but not human-likeness can be manipulated by clothing. In addition we found that first impressions on expertise and likeability may vary with dress and that the dress itself may induce stereotypical job associations. In our second study, we interviewed experts to get a more qualitative perspective on robot clothing. The interviews made clear that specific features of clothing such as colour may trigger strong associations with, for example, occupations, and that features such as headdress may make an outfit look more like a uniform. In our final study, a field experiment, we compared a robot without clothing and with a green, uniform-like outfit in a receptionist use case. The findings suggest that robots with uniforms appear to be more appealing. They also provide some evidence that a robot "identity" may be useful in practice. Overall the findings suggest that clothing can have an effect on first impressions, gender perception, and engagement in interactions with users in the wild as well as signify a robot's identity. Clothing shapes our perception of others, ranging from gender, age, status, occupation, and even personality, so it seems logical that these perceptions would apply to robots as well. It is often one of the most prominent characteristics when meeting someone, influencing the formation of first impressions [29]. As robots become more embedded in our daily lives over the coming decades, clothing could provide robots with identities based on their application and purpose, increasing transparency and understanding for social robots in various environments and conditions.

Acknowledgment

The authors would like to thank *Fajjaaz Chandoe, Georgiana Juglan, Nihed Harrak, and Vincent Koeman* for developing the use case and assisting with the experiment.

References

[1] K. K. P. Johnson, N. A. Schofield, and J. Yurchisin, "Appearance and dress as a source of information: A qualitative approach to data collection," *Clothing and Textiles Research Journal*, vol. 20, no. 3, pp. 125–137, 2002.

[2] M. E. Roach-Higgins and J. B. Eicher, "Dress and identity," *Clothing and Textiles Research Journal*, vol. 10, no. 4, pp. 1–8, 1992.

[3] I. W. King *et al.*, "'What to wear?': Clothing as an Example of Expression and Intentionality," *Argument: Biannual Philosophical Journal*, no. 1, pp. 59–78, 2015.

[4] K. K. P. Johnson, J.-J. Yoo, M. Kim, and S. J. Lennon, "Dress and human behavior: A review and critique," *Clothing and Textiles Research Journal*, vol. 26, no. 1, pp. 3–22, 2008.

[5] K. Nelson and J. Bowen, "The effect of employee uniforms on employee satisfaction," *Cornell Hotel and Restaurant Administration Quarterly*, vol. 41, no. 2, pp. 86–95, 2000.

[6] E. Phillips, X. Zhao, D. Ullman, and B. F. Malle, "What is human-like? decomposing robots' human-like appearance using the anthropomorphic robot (abot) database," p. 105–113, Feb 2018.

[7] J. Bernotat, F. Eyssel, and J. Sachse, "The (fe)male robot: how robot body shape impacts first impressions and trust towards robots," *Int. Journal of Social Robotics*, vol. 13, no. 3, pp. 477–489, 2021.

[8] K. C. Malakwen, "The Influence of Workplace Attire on Employees' Productivity in the Service Industry: A Case of Nairobi Serena Hotel," Ph.D. dissertation, United States International University-Africa, 2015.

[9] J. Hurtienne and D. Arnold, "The naked truth?" New York, NY, USA: Association for Computing Machinery, 2020.

[10] J. Lower, "Style Speaks: Clothing Judgments, Gender Stereotypes, and Expectancy Violations of Professional Women," Master's thesis, University of Central Florida, 2018.

[11] N. Friedman, K. Love, A. Bremers, A. Parry, R. LC, B. Amgalan, J. Liu, and W. Ju, "Designing functional clothing for human-robot interaction," in *Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*, 2021, p. 703–705.

[12] F. Hegel, M. Lohse, and B. Wrede, "Effects of visual appearance on the attribution of applications in social robotics," in *18th Int. Symposium on Robot and Human Interactive Communication*, 2009, pp. 64–71.

[13] N. Savela, T. Turja, and A. Oksanen, "Social acceptance of robots in different occupational fields: A systematic literature review," *International Journal of Social Robotics*, vol. 10, no. 4, pp. 493–502, 2018.

[14] D. Bryant, J. Borenstein, and A. Howard, *Why Should We Gender? The Effect of Robot Gendering and Occupational Stereotypes on Human Trust and Perceived Competency*. New York, NY, USA: Association for Computing Machinery, 2020.

[15] B. Tay, Y. Jung, and T. Park, "When stereotypes meet robots: The double-edge sword of robot gender and personality in human–robot interaction," *Computers in Human Behavior*, vol. 38, p. 75–84, 2014.

[16] IFR Press, "Robot density rises globally," 2018. [Online]. Available: https://ifr.org/ifr-press-releases/news/robot-density-rises-globally

[17] R. Ohanian, "Construction and validation of a scale to measure celebrity endorsers' perceived expertise, trustworthiness, and attractiveness," *Journal of Advertising*, vol. 19, no. 3, pp. 39–52, 1990.

[18] A. D. Bryan, G. D. Webster, and A. L. Mahaffey, "The big, the rich, and the powerful: Physical, financial, and social dimensions of dominance in mating and attraction," *Personality and Social Psychology Bulletin*, vol. 37, no. 3, pp. 365–382, 2011, pMID: 21252382.

[19] S. Reysen, "Construction of a new scale: The reysen likability scale," *Social Behavior and Personality: an international journal*, vol. 33, no. 2, pp. 201–208, 2005.

[20] J. F. Dovidio, K. Kawakami, and S. L. Gaertner, "Implicit and explicit prejudice and interracial interaction." *Journal of personality and social psychology*, vol. 82, no. 1, p. 62, 2002.

[21] R. C. Ladwig and E. C. Ferstl, "What's in a name? an online survey on gender stereotyping of humanoid social robots," in *Proceedings of the 4th Conference on Gender & IT*, ser. GenderIT '18. New York, NY, USA: Association for Computing Machinery, 2018, p. 67–69.

[22] C. Bartneck, D. Kulić, E. Croft, and S. Zoghbi, "Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots," *International journal of social robotics*, vol. 1, no. 1, pp. 71–81, 2009.

[23] R. van den Berghe, M. de Haas, O. Oudgenoeg-Paz, E. Krahmer, J. Verhagen, P. Vogt, B. Willemsen, J. de Wit, and P. Leseman, "A toy or a friend? children's anthropomorphic beliefs about robots and how these relate to second-language word learning," *Journal of Computer Assisted Learning*, vol. 37, no. 2, pp. 396–410.

[24] R. Mauro, "The constable's new clothes: Effects of uniforms on perceptions and problems of police officers 1," *Journal of Applied Social Psychology*, vol. 14, no. 1, pp. 42–56, 1984.

[25] D. Oh, E. A. Buck, and A. Todorov, "Revealing hidden gender biases in competence impressions of faces," *Psychological Science*, vol. 30, no. 1, pp. 65–79, 2019.

[26] W. Wood and S. J. Karten, "Sex differences in interaction style as a product of perceived sex differences in competence." *Journal of personality and social psychology*, vol. 50, no. 2, p. 341, 1986.

[27] K. Johnson, S. J. Lennon, and N. Rudd, "Dress, body and self: Research in the social psychology of dress," *Fashion and Textiles*, vol. 1, no. 1, pp. 1–24, 2014.

[28] H. Over, A. Eggleston, and R. Cook, "Ritual and the origins of first impressions," *Philosophical Transactions of the Royal Society B*, vol. 375, no. 1805, p. 20190435, 2020.

[29] N. Howlett, K. Pine, I. Orakçıoğlu, and B. Fletcher, "The influence of clothing on first impressions: Rapid and positive responses to minor changes in male attire," *Journal of Fashion Marketing and Management: An International Journal*, 2013.