



The Influence of Robot Designs on Human Compliance and Emotion: A Virtual Reality Study in the Context of Future Public Transport

OLIVIA HERZOG, Technical University Munich, Chair of Ergonomics, Germany, and TUMCREATE Ltd., Design for Autonomous Mobility, Singapore

NIKLAS FORCHHAMMER and **PENNY KONG**, TUMCREATE Ltd., Design for Autonomous Mobility, Singapore

PHILIPP MARUHN, Technical University Munich, Chair of Ergonomics, Germany

HENRIETTE CORNET, TUMCREATE Ltd., Design for Autonomous Mobility, Singapore

FRITZ FRENKLER, Technical University Munich, Chair of Industrial Design, Germany

As robots enter everyday environments, they start performing tasks originally performed by humans. One field of application is the public transport sector. The deployment of autonomous transport systems comes with a lack of human contact persons for help, guidance, and crowd management. This elicits challenges regarding redirecting and managing passengers. Current solutions on platforms can be replaced or enriched with service robots whose task includes crowd management as well as social interaction. This study investigates how the human-likeness of a robot influences the compliance and emotions of public transport users. A Virtual Reality experiment was conducted (N=33) to evaluate two different robot designs in a bus stop boarding scenario. The two robot designs differ in terms of humanoid appearance. In different experimental trials, participants had to perform a given task that was nullified by instructions from one of the two robots. Additionally, the dissonance of the situation was altered so that the environment either justified the robot's interference or not. Compliant behavior, pleasure, and arousal ratings, as well as task processing times were recorded. The experiment included an individual interview and a post-study questionnaire. The results suggest that future deployment of service robots has the potential to redirect passengers. In dissonant situations, clear reasoning must be given to make the robot effective. However, the robot's visual appearance has a more substantial impact on arousal and subjective preferences than on evoked behavior. The study implies that the presence of a service robot can influence peoples' choices and gives hints about the importance of giving a reason. However, objectively, the level of the robot's humanoid appearance did not make a difference.

This work is part of a research program funded by the National Research Foundation (NRF) Singapore under its Campus for Research Excellence and Technological Enterprise (CREATE) program. The work was also supported by a scholarship of Bund der Freunde der TUM (Technische Universität München) e. V.

Authors' addresses: O. Herzog, Technical University Munich, München 80333, Germany, TUMCREATE Ltd., Design for Autonomous Mobility, Singapore; email: olivia.herzog@tum.de; N. Forchhammer, P. Kong, and H. Cornet, TUMCREATE Ltd., Design for Autonomous Mobility, Singapore; emails: niklas.forchhammer@gmail.com, penny.kong@tum-asia.edu.sg, henriette.cornet@gmail.com; P. Maruhn, Technical University Munich, Chair of Ergonomics, München 80333, Germany; email: philipp.maruhn@tum.de; F. Frenkler, Technical University Munich, Chair of Industrial Design, München 80333, Germany; email: fritz.frenkler@tum.de.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2022 Association for Computing Machinery.

2573-9522/2022/02-ART21 \$15.00

<https://doi.org/10.1145/3507472>

CCS Concepts: • **Human-centered computing** → **Empirical studies in interaction design**;

Additional Key Words and Phrases: Human-robot interaction, service robots, affective design, public transport, crowd management, virtual reality

ACM Reference format:

Olivia Herzog, Niklas Forchhammer, Penny Kong, Philipp Maruhn, Henriette Cornet, and Fritz Frenkler. 2022. The Influence of Robot Designs on Human Compliance and Emotion: A Virtual Reality Study in the Context of Future Public Transport. *ACM Trans. Hum.-Robot Interact.* 11, 2, Article 21 (February 2022), 17 pages. <https://doi.org/10.1145/3507472>

1 INTRODUCTION

In the near future, interactive robots will be increasingly used in everyday public spaces, performing service roles traditionally played by humans [16]. Studies found that the general public has intense interest in service robots [8]. Nevertheless, while much progress has been made to develop functional criteria for artificial agents and machines, guidelines to design the sociality of human-machine interactions to foster user acceptance are lacking [52].

One vast field with substantial, persistent user presence, which is poised to benefit greatly from automation (e.g., service robots), is the public transport sector. It will undergo significant disruption with the deployment of autonomous services. **Automated Road Transport Systems (ARTS)** are on the rise with more systems being evaluated in field studies [30]. The operation of autonomous buses is expected to enable faster and safer journeys for commuters [15]. However, the absence of the human driver leaves a gap in the service experience, both functionally and socially, in terms of providing help and guidance [32]. Furthermore, future dynamic, modular autonomous bus systems may come with constantly changing boarding zones and destinations [36]. The pod solution with changing boarding zones requires crowd redirection [12]. Passengers might need to be redirected to dynamically changing departure platforms. These systems require more and more flexible crowd management on the platforms.

Currently, solutions for crowd redirection mainly rely on static information like displays and service personnel on the platforms. These current solutions can be replaced or enriched with service robots in the future. They offer the possibility of human-like social interaction in multiple languages, they are cost-efficient, do not get fatigued and are able to give physical assistance [8]. Furthermore, Shinozawa, Naya, Yamato, and Kogure (2005) argue that they are more attractive, productive, and easy to use than screens [45].

However, for service robots to be deployed on public transport platforms, passengers must be comfortable and compliant with them to ensure efficient crowd management [5]. An important aspect of developing socially interactive robot companions is the design for effective **human-robot interaction (HRI)** as well as for acceptance by users. One approach to increase people's acceptance of robots is to increase the familiarity of a robot by using anthropomorphic designs for its physical shape [16]. The psychological effects of robots' visual appearance, and especially in terms of human-likeness, are very important in HRI research and are elaborated further in the next section. Usability, as defined by ISO 9241-11:2018 [22], incorporates effectiveness, efficiency, and satisfaction as three dimensions that can be measured either objectively or subjectively (perceived) [22]. This article aims to perform research on the design of an assistive robot companion for public transport and its effects on compliance and emotion.

The dimensions effectiveness and efficiency were addressed objectively, namely if the redirection was successful (effectiveness) and how long the task processing took in seconds (efficiency). Additionally, the participants' affective state was measured in terms of arousal and pleasure, where pleasure can be a perceived measure for satisfaction, according to ISO 9241-11:2018 [22].

2 STATE OF THE ART

According to Breckler, Olson, and Wiggins (2005), compliance refers to a change in behavior that was requested by another individual, even though it was possible to refuse [7]. Haring et al. (2019) investigated the application of compliance concepts to robots [19]. Studies suggest a link between anthropomorphism and human-likeness to obedience [1]. Following this, Haring et al. tested the effect of a robot's human-like appearance on compliance with a request by comparing two different robots, a high and a low human-like robot, to a human control condition. The robot's task was to continuously prompt participants to continue practicing a task. Their results showed that compliance with the request was the highest with a human and compliance with both robots was significantly lower. There were no differences between the high and low human-like robots for compliance. However, they showed that robots could be used as persuasive coaches that can help a human teammate to persist in a training task. In summary, they found that robots are currently inferior to humans, but future robots can be designed carefully in an authoritative way to gain and maximize compliance [19]. Zhang et al. investigate robot feature design effects on user perceptions in their studies and recommend measuring compliance-related emotional responses [54].

It is to be expected that human perceptions of robot behavior depend to a large extent on its visual appearance [17]. Human-machine communication appears to rely on an anthropocentric expectancy bias, where people tend to follow a human-human social script while communicating and experiencing a violation when their partners are not or less human-like [13]. How this framework, also known as the **Computers Are Social Actors (CASA)** paradigm [34], applies to instructive robots is unclear. At the same time, [44] compared robots (embodied agents) with virtual assistants (disembodied/pervasive agents). They found that the agent's form did not influence behavioral outcomes. Even nudges from the pervasive agent were sufficient to induce the intended normative behavior due to participants' perceptions of the agent's social presence, or, the "watching-eyes" effect [51].

Salem et al. (2011) recommend equipping a robot with humanoid features such as a head, two arms, and two legs to support an intuitive and meaningful interaction with humans [42]. Emotional responses to **anthropomorphic robots (ABOT)** and phenomena connected to them are very important in HRI research and have to be investigated when designing robots. The psychological process of anthropomorphism is a strategy to increasingly make sense of a non-human's actions and thus to reduce the uncertainty that is often associated with interactions with non-human agents [14]. Supporting this assumption, Luczak et al. (2003) demonstrated that anthropomorphism in technical devices, such as cars or computers is associated with reduced stress in HRIs [29]. At the same time, positive emotions are believed to result from exposure to a human-like stimulus [29], at least as long as the stimulus does not become eerie.

Results by Torta et al. (2014) showed that a humanoid robot was trusted by participants in care of the elderly and that the participants might even engage in an emotional relationship with the robot [48]. The presence of anthropomorphic and interactive features in a service robot promoted positive emotional responses (subjective feelings and physiological responses) in elderly participants, with physical appearance contributing more to perceived humanness than the sound of the robot's voice [55].

3 RESEARCH OBJECTIVES

The overall goal of this research is to investigate the efficacy and user acceptance of an assistive robot for crowd management on public transport platforms in Singapore. Natarajan and Gombolay (2020) found that anthropomorphism elicited by humanoid features is one of the most significant factors in predicting trust and compliance with a robot [35]. Therefore, we hypothesize that

Table 1. Study Hypotheses

H1.1	Passengers are more likely to board a slower bus when a humanoid tells them to do so in comparison to a functional robot.
H1.2	Passengers are more likely to board a slower bus when a robot tells them to do so, when there is clear reasoning.
H2.1	The humanoid robot leads to higher pleasure ratings.
H2.2	The humanoid robot leads to lower arousal ratings.
H3	The humanoid robot leads to higher task processing times than the functional robot when the dissonance is high.

compliance is higher when a humanoid robot gives instructions (H1.1) and that low dissonance leads to a higher probability of compliance with the robot (H1.2). In terms of affective states, we assert that a humanoid design of the service robot will lead to a different experience in terms of pleasure and arousal pattern than a functional design (H2.1 and H2.2). Furthermore, we hypothesize that a humanoid design increases the task processing time in high dissonance conditions (H3).

The empirical investigation is based on how anthropomorphic characteristics in the service robot—meaning the attribution of human qualities to a non-living object—influence the elicitation of positive emotions and compliance during the redirection of commuting passengers. For this, an empirical study was conducted to evaluate two different assistive robot designs for crowd management on a public transport platform in Singapore with the following study hypotheses (Table 1).

4 METHODS

Prototyping multiple robot designs would engage high effort in terms of costs and time. Additionally, the selected use case environment for this study of an autonomous shuttle bus stop is not yet available in a real-life setting. Therefore, the empirical study was conducted in **Virtual Reality (VR)** to simulate the selected use case in an immersive environment. The method of VR in research always has to be handled with caution. Pan and Hamilton (2018) state, that VR can provide good results with high validity, while enabling reproducibility and novel experimental contexts. Although they note the generalizability of VR to the real world has not been tested in detail VR is still a useful tool to provide good experimental control [37]. Also, a study conducted by Duguleana showed that using VR was a feasible way of simulating robotic scenarios but they also reported differences in participant behavior between the real world experiment and the virtual environment, with the participants giving more personal space to robots in the real world [11]. Wijnen et al. (2020) investigated whether studies run in VR yield realistic results and thus provide support to the idea that VR is a valid platform for HRI and user studies in general [53]. To achieve best possible results a highly immersive set up was used.

The VR set up consisted of an HTC Vive Pro **head-mounted display (HMD)**, featuring a dual 3.5" diagonal AMOLED screen with a resolution of 1440 x 1600 pixels per eye and a field of view of 110 degrees. The high resolution of the HTC Vive Pro helps to avoid the screen-door effect, a picture artefact occurring when using displays, especially with text displayed in VR. An HP OMEN workstation with an Intel Core i7 (2.8 GHz) processor, 16 GB DDR4 SDRAM, and an NVIDIA GeForce GTX 1070 graphics card with 8 GB VRAM was used. The HMD had a wired connection with a cable length of 5 m (plus 1 m from Link Box to PC). A stable frame rate above 90 frames per second was ensured to prevent motion sickness.

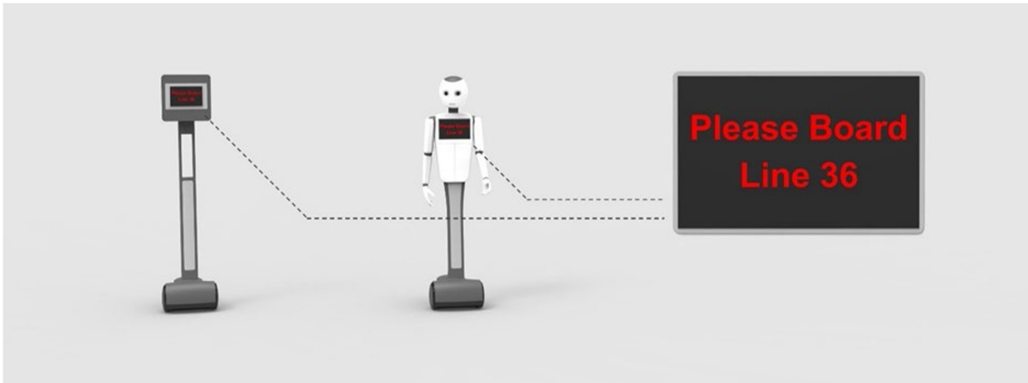


Fig. 1. The two robot designs used, with the screen magnified. The non-humanoid design is shown on the left and the humanoid design is on the right.

A 3D environment of an autonomous bus platform was created in the game engine Unity, version 2019.2. Two different service robot designs, which were animated and incorporated in the VR environment, were used as stimuli for the experiment (Figure 1). The robot designs were based on preliminary studies conducted [24] to determine which features were crucial or non-crucial in human-machine interfaces for autonomous vehicles. This resulted in two distinct concepts: a sociable, humanoid robot that provides physical assistance and an impersonal, screen-based display robot providing information. The concepts were modified for this study so that they can be distinguished from each other in terms of humanoid appearance, but both communicate via a screen and an audio output with the same message. The non-humanoid robot is composed of a simple screen on top of a locomotive mechanism. The design was based on the robot “BEAM” by Suitable Technologies (2018) since it is a frequently used, very basic telepresence platform design which is mobile and incorporates a screen. The humanoid design was also created based on BEAM, but additionally incorporates a head with a simple face, arms, and hands with humanoid gestures in accordance with the literature regarding the creation of humanoid appearances [17, 38, 42, 56]. Both robots moved towards the participants when they approached the bus stop and played the audio message, which was also displayed on their body. The humanoid robot additionally made subtle and neutral arm movements to support its human-likeness without gesturing into one or another direction.

According to the recommendations by Hiroi and Ito (2016) a robot height of 30 cm below the user’s eye height is the most comfortable for verbal interaction [21]. Anthropometry data for Singapore was retrieved from Chuan, Hartono, and Kumar (2010) [9]. A reference value suitable for two genders was chosen: the 50th percentile in eye height for Singaporean males is 163 cm and for females 160 cm. Thus, following Chuan et al.’s recommendation, a value of 131.5 cm (as the mean value of 130 cm and 133 cm) was chosen for the humanoid robot. Both robots should be the same height, so Beam was also scaled to a height of 131.5 cm (original height: 134.4 cm).

Both robots are based on the same basic model to ensure that they have the same visual style to provide comparability. To prevent either of the designs from falling into the Uncanny Valley—meaning eliciting an eerie sensation in humans, typically occurring somewhere between 50% and 100% on the human-likeness axis [33]—both designs were intended to definitely stay below a human-likeness rating of 50%. To further validate the two robots’ visual distinguishability, the ABOT database was used to rate the two designs regarding anthropomorphism [38]. The humanoid design reached a rating of 43.83 out of 100 on the human-likeness scale, while the non-humanoid robot was rated 4.4 out of 100 [47].

Table 2. The Six Study Trials that Every Participant Experienced in a Randomized Order













	LOWDiss		HIGHDiss	
	FAST	SLOW	FAST	SLOW
Baseline				
Functional robot				
Humanoid robot				



Fig. 2. The bus stop environment in VR from the passenger's perspective, with the crowded bus on the left and the empty bus on the right as in the LOWDiss baseline trial.

In total, the participants went through six trials: four experimental and two baseline trials. Half of the experiment trials featured the humanoid robot (H) and the other half the functional robot (F). The baseline trials featured none of the two robots to get participants' normal behavior as a control. Table 2 shows the study trials from the participants' perspective with simplified pictograms of the VR environment. All participants went through the same six trials with different conditions. To counteract study design induced biases, the trial order was randomized for every participant.

In addition to altering the visual appearance of the robot (humanoid vs. functional), induced dissonance served as a second independent variable. The participants' task in the experiment was to board their preferred bus to get home from work on a bus platform, with two autonomous buses from different lines equidistant from the participant's start point. The arm swing locomotion method was used to enable the participants to walk around in the VR environment. This method was chosen due to the small laboratory space and to ensure safety. The arm swing locomotion allows moving in the virtual space by swinging the left and right controller alternately similar to arm movements during actual walking but without physically moving forward. The environment is shown in Figure 2 from the participant's perspective. In the briefing, participants learned that



Fig. 3. The Affective Sliders, implemented into VR based on Betella and Verschure (2016) [3], from the participant's perspective.

they could take both buses to get home. LINE 105 was introduced as the line with a shorter overall travel time and one that the participants therefore usually take to get home even though both buses would get them to their destination. However, in the four experimental trials, a robot appeared on the platform and instructed the participants to take the other, slower bus (LINE 36), hence interfering with their original intention (dissonance conditions). In half of the experiment trials the situation induced low dissonance. The faster bus was crowded with people creating a situation where the robot's advice to take the slower bus made sense and could be connected to an obvious reason (LOWDiss). For the trials with a crowded bus, animated human passengers were visibly placed inside the vehicle. The other half of the trials showed two empty buses and therefore induced high dissonance since the participants could not identify an apparent reason to follow the robot's advice (HIGHDiss). They could only assume that the robot had more information than them.

The participants' behavioral response (i.e., which bus they boarded), pleasure and arousal ratings, as well as the task processing time, were recorded for each trial. Pleasure and arousal ratings were collected after each trial using the Affective Sliders developed by Betella and Verschure (2016) [3] implemented into the VR environment, which participants could manipulate by pointing and dragging with the VR controller (Figure 3). This way the emotional ratings could be measured metrically. The Affective Sliders are based on the Circumplex Model of Affect by Russell (2003) which holds that emotions can be described as linear combinations of the two underlying, independent neuro-physiological systems arousal and pleasure [41]. The experiment also included a semi-structured, in-depth individual interview and a paper-and-pencil demographics and preferences questionnaire where participants were asked about their previous experience with robots and their opinion toward them.

The interview started with the following general questions:

- Did you notice that one bus was sometimes crowded?
- Did you notice there were two different kinds of robots?

Participants who did not notice a robot at all or answered they did not notice a robot but a sign on a stick were excluded from the data analysis. For the second part of the interview, the experimenter

showed an image of each scene the participants had experienced during the experiment. Using these images as a memory support, the experimenter went through all of the scenarios with the interviewed participant. The following questions were asked for each scenario:

- *You took line [36 or 105, in respective to the participant's specific action in the scenario]. Do you remember why?*
- *Only for trials with a crowded bus: There was a crowd in line 105. To what extent did that influence your decision?*
- *Only for robot trials: There was a robot. To what extent did that influence your decision?*

5 RESULTS

Thirty-six participants were recruited and 33 of them (age: $M = 28.91$ years, $SD = 5.86$, 39% female, 61% male, all working and living in Singapore) completed the VR experiment. Three participants were excluded due to motion sickness or based on conflicted answers in their interview (e.g., declaring that they did not notice two different robots at all). Based on the survey answers, participants were identified as having negligible experience with robots (55% had no experience with robots at all). However, 90% of the participants indicated having a fairly positive or very positive view of robots. The results showed a strong general preference for the humanoid robot design (76%) over the non-humanoid design (15%). Only 9% said they preferred to meet neither of the robots on a real-world bus platform.

A Cluster Analysis of the participants' responses to the open question on why they preferred either one of the robots revealed that the humanoid was mostly appreciated for the categories *personality and approachability* (8 times mentioned), *lifelikeness* (6), *entertainment* (4), *pleasantness* (3), and *aesthetics* (3). The non-humanoid was preferred for matching with its functionality (3) and not being distracting (2). The three participants who preferred neither of the two robots declared that they saw no need for a robot at all.

However, for the decision on which bus to take, hence following the robot's advice or not, no significant difference between the robot designs could be found. The participants tended to board the slower bus more often in the LOWDiss conditions (59%) than in the HIGHDiss conditions (45%) including the baseline trials. Figure 4 shows the frequency of participants who followed the robot's instruction and boarded the slower bus in the experimental conditions.

In alignment with our expectations, in the baseline condition with two empty buses, all participants boarded the faster bus. In the second baseline trial, where the faster bus was crowded with people, 39% of participants decided to take the slower bus instead. When a robot, regardless of its appearance, instructed them to take the slower yet empty bus, the rate increased to 68%. In trials where both lines were empty, 45% still followed the instruction. Consequently, the robot had an impact on the participants' decision-making (39% vs. 45%). However, the general compliance with the robot's instruction was very even between the two different robots: In trials with the humanoid 58% of participants followed its instruction. In trials with the non-humanoid robot, it was 56%.

However, the dissonance of the condition had an impact on the participants' decision making. When only taking into account the trials with the crowded fast bus, hence trials where there was a visible reason and justification to take the empty bus (LOWDiss), participants chose the slower bus significantly more often when a robot told them to, in comparison to the baseline condition without a robot. A χ^2 -test of independence ($\chi^2(2) = 7.58$, $p = 0.023$) revealed that the robot indeed had a significant effect on the participants' decision to take the slower bus, beyond the fact that people just preferred an empty bus over a crowded one.

Emotional ratings (pleasure and arousal) were analyzed using linear mixed-effect models including the subject ID and the trials' order number as random factors, using the lmerTest package

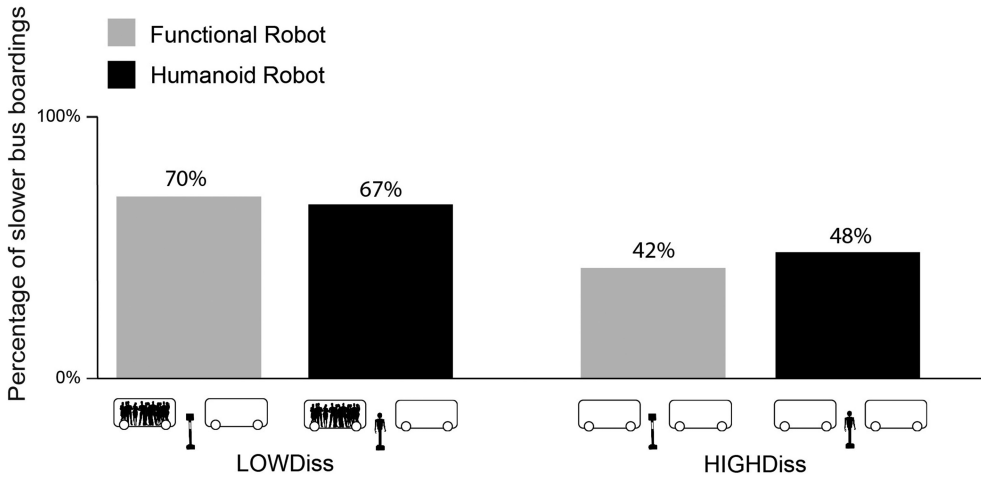


Fig. 4. Percentage of participants compliant with the robot's instruction in the experimental trials, i.e., boarding the slower bus on the right.

Table 3. Hypotheses with Results

Variable	Hypothesis
Compliance	H1.1 Passengers are more likely to board a slower bus when a humanoid tells them to do so in comparison to a functional robot. Non-significant ($\chi^2 (1) = 0.03, p = 0.86$)
	H1.2 Passengers are more likely to board a slower bus when a robot tells them to do so, when the situation is low dissonant. Significant ($\chi^2 (2) = 7.58, p = 0.02^*$)
Core Affect	H2.1 The humanoid robot leads to higher pleasure ratings. Non-significant ($\chi^2 (6) = 4.09, p = .13$)
	H2.2 The humanoid robot leads to lower arousal ratings. Rejected ($\chi^2 (6) = 1.53, p = .47$)
Processing Time	H3 The humanoid robot leads to higher task processing times than the functional robot when the dissonance is high. Significant ($\chi^2 (5) = 5.34, p = .02^*$)

* significant on an alpha level of $p \leq .05$.

[25] in R [39] and the following equation:

$$\text{Core Affect} \sim \text{Robot} + \text{Dissonance} + \text{Robot: Dissonance} + (1|\text{ID}) + (1|\text{TrialOrderNumber})$$

The model revealed no significant effects (Table 3). The average pleasure rating was slightly higher for the humanoid robot ($M = 0.63, SD = 0.23$) than for the non-humanoid robot ($M = 0.59, SD = 0.23$). The highest average pleasure rating, however, became evident in the trials without any robot redirecting ($M = 0.64, SD = 0.23$). With the humanoid robot, the arousal was higher on average ($M = 0.63, SD = 0.23$) than with the non-humanoid ($M = 0.59, SD = 0.23$), or without any robot ($M = 0.61, SD = 0.23$). In LOWDiss trials the average pleasure rating ($M = 0.6, SD = 0.23$) did not differ from HIGHDiss trials ($M = 0.62, SD = 0.22, t(65) = -0.51, p = 0.61$). Arousal ratings were

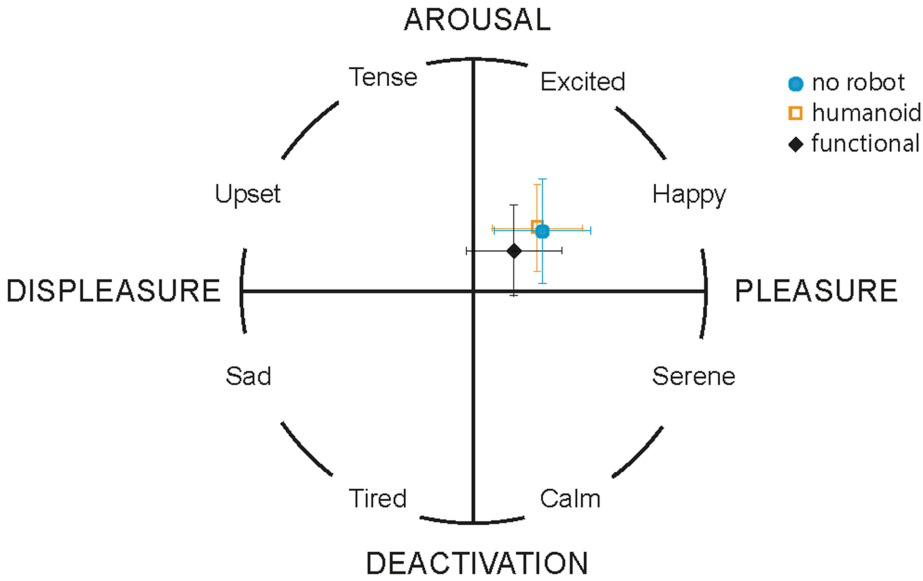


Fig. 5. The Affective Sliders ratings for the robot conditions visualized in the Circumplex Model of Affect based on Russell (2003). The symbols stand for the mean value, the whiskers represent the standard deviation.

higher in HIGHDis trials ($M = 0.63, SD = 0.22$) than in LOWDis ($M = 0.59, SD = 0.22$), but only in a borderline significant way ($t(65) = -1.926, p = 0.058$).

Figure 5 shows the Circumplex Model of Affect, based on Russell (2003) [41], with the Affective Slider results of the three different robot conditions: no robot, functional robot, and humanoid robot. The functional robot trials were, both in pleasure and arousal, rated slightly lower, while the humanoid robot trials and the no robot trials are almost congruent to each other.

The task processing time was measured in seconds from the start of the trial until the moment where the participant’s virtual position collided with one of the buses, indicating that they had boarded one of the buses and ending the trial. The effect of the two independent variables, Robot and Dissonance, on the task processing time TPT was analyzed using a linear mixed-effect model including the subject ID and the trials’ order number as random factors, using the lmerTest package [25] in R [39] and the following equation:

$$TPT \sim Robot + Dissonance + Robot: Dissonance + (1|ID) + (1|TrialOrderNumber)$$

The model revealed a significant main effect of the type of robot on the task processing time on a significance level of $\alpha = .05$, where including type of robot as a predictor significantly improved overall model fit, $\chi^2(5) = 5.34, p = .02$. Task processing time was significantly higher in trials with the humanoid robot ($M = 26,62s, SD = 8.73s$) in comparison to the functional robot ($M = 24,19s, SD = 7.51s$), $b = 2.31, t(94.4) = 2.33, p = .02$. An overview of the results regarding the hypotheses is shown in Table 3.

The qualitative data from the semi-structured post-experimental interview was cluster-analyzed to identify patterns and find significant moments and quotes within the open answers to be matched with the objective results. First off, without being presented with the terms “humanoid” or “functional” before, all participants were able to distinguish between the two robot designs in terms of their human-likeness. Fifteen participants mentioned that it would have made a difference for them if the robot had additionally given them a valid reason for its advice, so that they would have trusted and followed it more. For every trial, the participants were confronted with

their behavior and asked why they acted as they did. As expected, when boarding the faster bus, the most frequently mentioned reason was the shorter travel time (63 times mentioned). When boarding the slower bus, the most frequent explanation was following the robot's instruction (54), followed by avoiding the crowd (23).

6 DISCUSSION

For compliance with the robots' instruction, we found no significant effect for the type of robot (Hypothesis 1.1). However, participants showed significantly more compliant behavior in the trials with low dissonance (Hypothesis 1.2), which aligns with literature implying that less dissonance is connected to less adverse cognition [20]. In summary, the hypotheses regarding compliance can be answered positively for the dissonance of the condition, but not for the visual appearance of the robot. The results are in line with results by Haring et al. (2019), who also found no significant differences between high and low human-like robots for compliance [19]. Moreover, the work of Agrawal and Williams (2017) showed that perceived anthropomorphism did not seem to have a strong relation with obedience, although direct comparison of the mean values suggest that the people following the instructions perceived the robot to be more human-like [1]. However, in the study at hand, the robots can still be regarded as effective, with the result of Hypothesis 1.2 in mind. Both robots inherently had a significant effect, going beyond the fact that people just preferred an empty bus over a crowded one. When transparent reasoning is given, a service robot, therefore, can offer effective behavior modification. Giving a justification and explanation for a robot's behavior was also found a factor to mitigate reactance and lead to higher compliance with a robot's cue in the works of [2] and [6]. Additionally, following Agrawal and Williams's suggestions (2017), emphasis should be put on developing robot behavior that ensures safety, so that people develop a sense of trust in the robot. Giving justified advice, therefore, keeps dissonance low, which may ensure this safety. It is widely known that human decision making can be irrational [7]. Humans show a wide range of decision-making bias. Several of these biases may have been at play in this experiment. For example, people often show reactant behavior in situations where they feel their freedom of choice is threatened, and as a result, they try to regain autonomy and control by doing precisely the opposite [10], which may partly explain why some participants refused to follow the robot's instruction. Here a connection can be drawn to the outcomes of the open interview, where one participant stated that he went against the robot's instruction on purpose. This also aligns with results by Roubroeks implying that under certain circumstances persuasive technology can trigger opposite effects, especially when people have incongruent goal intentions [40].

The hypotheses regarding the affective responses cannot be affirmed with the data at hand. No differences could be found for pleasure. In terms of higher arousal with the humanoid robot, an effect was observed, but in the reverse direction as expected from the literature. This indicates that the humanoid design was not calming and stress-reducing, but rather more exciting than a functional one. This aligns with the results of the robot preference, which implied that the humanoid appearance has a considerable impact on likeability and preference, and with the results of the open interviews. However, since the humanoid robot used slightly more modalities, like arm movements and gaze, the comparison might not have been entirely fair. Not implementing these modalities despite integrating arms and eyes, on the other hand, might have been perceived as unnatural and irritating. Also, the results of the Affective Sliders in the Circumplex Model of Affect showed that the robot designs were located in the upper right quarter of the diagram, indicating an overall positive emotional response. The highest average pleasure rating, which was evident without any robot redirecting, might be due to people not being comfortable with seeing a robot on the platform. More likely, though, it might be closely related to the fact that people were not redirected in the trials without a robot and, therefore, were less irritated in general. Notably, the open answers

in the interview and the preference questionnaire again allow for conclusions about participants' attitudes and motives in this context. Positive emotional responses to automation can influence trust, liking, and reliance on a system [43]. Since the purpose of the robot is to redirect passengers in a public transport system, its role in the entire mobility experience should be considered. While visual appearance did not influence behavior, a robot could, however, serve to make ubiquitous interactions such as crowd redirection more pleasant in the long-term. Besides, following the robot's instruction resulted in higher pleasure ratings. This aligns with results by Biele, Rieskamp, Krugel, and Heekeren (2011) indicating that following advice can be intrinsically rewarding [4].

The hypothesis regarding task efficiency can be affirmed. The humanoid design led to higher task processing times and consequently, worse task efficiency than the functional design (H3.1). In other words, the humanoid appearance seemed to "slow down" passengers, especially in the high dissonance condition. It especially affected the decision time, i.e., the time before the participant started walking. This might be due to a higher interest to interact with or look at it, or due to higher confusion because it is giving the "wrong" advice. This aligns with results by Goodyear et al. (2016) showing that bad advice by a human decreased performance more than advice by a computer agent [18]. Another approach is that the users had higher expectations towards the humanoid design and, therefore, waited for it to speak to them. This also aligns with results by Joosse, Lohse, Perez, and Evers (2013) that role expectations indeed influence people's behavior leading to more or less compliance in particular task settings [23]. However, none of the participants tried to talk to either of the robots or ask them questions. Another point to discuss is that people may not have been influenced solely by the robot, but by other unidentified factors like the crowd in the bus or because they wanted to have a seat. For example, the COVID-19 pandemic might have also influenced the participants' behavior since the task was to choose between a crowded and an empty bus, which can be an issue of keeping distance to other people because of infection risks. However, none of the participants mentioned this as a reason for any scenario during the final interviews. More factors in this context need to be identified and included in future research. As an example, personality traits could play a role, since obedience and compliance are correlated to personality, as was, for example, investigated by Larmour, Bergström, Gillen, and Forth (2015) in the context of interrogative compliance [26]. Bishop et al. (2019), in addition, found relations of robot acceptance with age and mood in HRI [5]. This could serve as implications for future studies. Generally, it can be discussed if a service robot in this context adds value, hence if the existence made a difference at all. According to the results regarding Hypothesis 1.2, it did. However, it was not investigated if a stationary sign may have had a similar effect. This leads to another limitation of the study design: There was no baseline condition with a non-robot intervention. In the context of dynamic autonomous transport, the interactive robot may, however, offer advantages beyond that. Robots might be useful in serving as interactive and physical, maybe even social, guides. An application scope could be an approachable information source not only with AVs in public bus transport, but also, for example, in airports, bus terminals, and train stations. The context of the surrounding (ambient noise, crowdedness) has to be taken into account and examined to determine whether the deployment of a service robot is useful.

7 LIMITATIONS

Regarding objectivity, a discussion point for the study at hand is whether social desirability influenced the experiment. Few people mentioned in the interview that they wanted to respect the creator's choice, indicating the presence of social desirability towards the experimenter. On the other hand, the participants were immersed and isolated in the VR and could not see if the investigator could observe what they were doing inside the virtual environment, which could mitigate this effect. To get data on the reliability of the results, the study should be replicated, and

the results compared. Ideally, it should also be translated into a real-world experiment to see if the VR test delivered reliable data since at this point in time it stays unclear how the results can be interpreted in real-world settings. This is one of the biggest limitations of the study at hand. However, even though an external validity may not be given in VR studies, relative validity when comparing different conditions inside VR can be present. When it comes to validity, the method of the Affective Sliders needs to be discussed. The method led to some ceiling effects, meaning that three participants always responded to the sliders with the highest possible rating of 1. Even though excluding them from the analyses did not influence the outcome, it still indicates that the sliders must be used with caution. The original sliders were validated for the purpose of measuring core affect. For this study, they were translated into a virtual application, modifying the mean of interaction from digital to physical by physically moving the controller. Results by Laurans, Desmet, and Hekkert (2009) imply that the modality of interaction with the slider might have an impact on the self-report ratings [27].

When doing experiments in VR, frequent problems are a lack of immersion, resulting in behavioral artefacts, i.e., participants do not show authentic behavior. In relation to this, it could be that the people crowding the bus were not perceived as persons because they were virtual agents. These assumptions can be drawn from interview answers, where participants indicated that they were not bothered by the virtual characters because they did not perceive them as real. Another point to discuss is the subject sample. Due to restrictions and suspension of subject recruiting outside the building because of the COVID-19 pandemic, the sample was small and not as heterogeneous as desired. Besides, the pandemic might have also influenced the participants' behavior since the task was to choose between a crowded and an empty bus, which can be an issue of keeping distance to other people because of infection risks. However, none of the participants mentioned this as a reason or thought they had for none of the scenarios during the final interviews. In addition, the data collection took place in the very early beginnings of the pandemic when the distance awareness was still comparatively low. The participants were all recruited from a technical research entity, impacting the overall technology affinity and experience. To statistically identify small significant effects based on a power analysis, a larger subject sample could have been helpful. The small sample size was a result of the recruiting situation due to the COVID-19 outbreak. Even though the participants were all residents in Singapore for at least six months, a larger sample would also have been more representative of the population. The service robot on the platform appeared to give an official instruction to passengers from a public transport authority. Considering Singaporean culture and habits, people in Singapore may comply with official instructions in general, especially since Singapore is known for rigorous and high fines for misdemeanors ([28, 31, 49]). Participants mentioned in the interview that their decision is in connection with the state and rules of Singapore. One even mentioned that he would be scared not to follow official instructions because of potential penalization. Therefore, a replication of the experiment with a sample from a different cultural background could give further insights and potentially reveal differences. Another major point to discuss is the study design and the order of study trials. It may have influenced the behavior of participants dramatically in the form of practice effects, sequence effects, and novelty effects. In particular, the very first encounter with a robot of any kind may have induced a novelty effect [46]. Even though learning effects were dealt with by counterbalancing, carry-over effects remain present. Although a within-subject design reduces the error variance dramatically, they are prone to sequence effects. Nevertheless, for the research design a within-subject design was chosen since emotional ratings were measured. They are very subjective and variable between individuals, since there are genetically-based individual differences in average levels of core affect [41]. However, a study design with a between-subject group for the variable dissonance may have mitigated these effects. Additionally, since there were no distractor trials included in the design,

participants may have been able to predict trials after the first trials. The individual study trials were quite short, so it is unclear whether this provided enough time to alter individual core affect, measured by pleasure and arousal.

Furthermore, the humanoid robot was developed especially for this study and is therefore not thoroughly validated. Even though the ABOT database rating [47] was used to validate human-likeness, this is a singular, objective method. Moreover, the ratings are based on dichotomous questions and can therefore, sometimes be debatable. The perceived anthropomorphism is important and has to be assessed subjectively. In the same vein, the **Uncanny Valley Hypothesis (UVH)** can be discussed. Since there is no absolute threshold for the UVH, it cannot be clearly said that the effect has been avoided, since one participant described the humanoid robot as creepy during the interview. Besides, the story framework might not have affected participants' decisions because there were no real consequences when they took the slower bus, for example, in terms of waiting time. A real-time delay could have been implemented as a kind of "time penalty", but it was not feasible for the study in VR as there was no way to keep participants occupied during the waiting time. It would not have created a realistic situation because in real life, they would normally have a phone or any other preoccupation during the waiting time. One of the most significant limitations of the study could be that it examined a first encounter with the robot. Effects like trust and compliance might develop over time and multiple encounters. However, in this context, Agrawal and Williams (2017) showed that a majority of participants were cooperative and willing to take instructions from a robot, even if they had not encountered the robot before [1]. This offers an implication for future studies, to find out whether multiple encounters and, especially, consistency effects change the outcomes.

8 CONCLUSION

We conducted a study using VR to evaluate two different robot designs in terms of human-likeness in a public transport scenario. Participants had to perform a given task, which was nullified by instructions from one of the two robots. Compliant behavior, emotional ratings, as well as task processing times were recorded.

The results suggest that future deployment of service robots has the potential to be effective in redirecting passengers on a public transport platform, but not necessarily to improve people's emotional states and therefore their overall travel experience. The visual appearance of the robot plays an important role regarding preferences, but especially in dissonant situations, it seems like clear reasoning must be given to make the robot more effective and users more compliant. However, appearance has a larger impact on emotional response to the situation and preference than on compliant behavior.

In summary, the presence of a service robot can influence people's choices when there is a recognizable reason to do so. Objectively, a robot's appearance did not make a difference. Subjectively, a humanoid robot seemed to be more exciting and preferred over a functional version. The implication of this study is that when it comes to the visual appearance of service robots, the differences lie rather in qualitative than in quantitative aspects.

Further studies must be conducted to explore potential effects, reasons and get a deeper understanding, such as by including more parameters like feedback or personality traits. Further, group effects might be interesting, i.e., what happens when people around the participant comply/ignore the robot instruction. Multiple encounters and development of trust over time must be examined, especially if consistency effects change the outcomes. Practical implications for the design of future robot companions can be drawn. From the results of this study, a humanoid appearance cannot be recommended in terms of efficiency and effectiveness. However, regarding service experience, a humanoid appearance may be useful to facilitate interaction subjectively. This is aligned

with findings by Tung (2016) indicating that especially for children or seniors, a moderate level of anthropomorphic appearance combined with appropriate social cues can enhance preferences for and acceptance of robots [50]. This, of course, depends on the goal of the design process: whether it is the improvement of subjective user experience or objective goal attainment. For the latter, it is subject to discussion if society even wants and needs robots to be sociable. Many scientists have already mentioned concerns about ethical issues as humanoid service robots might lead to social isolation and loss of privacy [44]. These are factors that have to be carefully considered when designing future service robots for public space.

REFERENCES

- [1] Siddharth Agrawal and Mary-Anne Williams. 2017. Robot authority and human obedience. In *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*. ACM, New York, 57–58. DOI : <https://doi.org/10.1145/3029798.3038387>
- [2] Franziska Babel, Johannes M. Kraus, and Martin Baumann. 2020. Development and testing of psychological conflict resolution strategies for assertive robots to resolve human-robot goal conflict. *Frontiers in Robotics and AI* 7, 591448 (2020), 216. DOI : <https://doi.org/10.3389/frobt.2020.591448>
- [3] Alberto Betella and Paul F. M. J. Verschure. 2016. The affective slider: A digital self-assessment scale for the measurement of human emotions. *PLoS one* 11, 2 (2016), e0148037. DOI : <https://doi.org/10.1371/journal.pone.0148037>
- [4] Guido Biele, Jörg Rieskamp, Lea K. Krugel, and Hauke R. Heekeren. 2011. The neural basis of following advice. *PLoS biology* 9, 6 (2011), e1001089. DOI : <https://doi.org/10.1371/journal.pbio.1001089>
- [5] Laura Bishop, Anouk van Maris, Sanja Dogramadzi, and Nancy Zook. 2019. Social robots: The influence of human and robot characteristics on acceptance. *Paladyn, Journal of Behavioral Robotics* 10, 1 (2019), 346–358. DOI : <https://doi.org/10.1515/pjbr-2019-0028>
- [6] Annika Boos, Michaela Sax, and Jakob Reinhardt. 2020. Investigating perceived task urgency as justification for dominant robot behaviour. In *Proceedings of the 22nd International Conference on HCI International 2020 - posters*. Constantine Stephanidis and Margherita Antona (Eds.), Communications in Computer and Information Science, 1224–1226. Springer, 117–124. DOI : https://doi.org/10.1007/978-3-030-50726-8_15
- [7] Steven J. Breckler, James M. Olson, and Elizabeth C. Wiggins. 2006. *Social Psychology Alive*. Thomson Wadsworth, Belmont, CA.
- [8] Yingfeng Chen, Feng Wu, Wei Shuai, and Xiaoping Chen. 2017. Robots serve humans in public places— KeJia robot as a shopping assistant. *International Journal of Advanced Robotic Systems* 14, 3 (2017), 172988141770356. DOI : <https://doi.org/10.1177/1729881417703569>
- [9] Tan K. Chuan, Markus Hartono, and Naresh Kumar. 2010. Anthropometry of the singaporean and indonesian populations. *International Journal of Industrial Ergonomics* 40, 6 (2010), 757–766. DOI : <https://doi.org/10.1016/j.ergon.2010.05.001>
- [10] Matthew T. Crawford, Allen R. McConnell, Amy C. Lewis, and Steven J. Sherman. 2002. Reactance, compliance, and anticipated regret. *Journal of Experimental Social Psychology* 38, 1 (2002), 56–63. DOI : <https://doi.org/10.1006/jesp.2001.1481>
- [11] Mihai Duguleana, Florin G. Barbuceanu, and Gheorghe Mogan. 2011. Evaluating human-robot interaction during a manipulation experiment conducted in immersive virtual reality. In *Proceedings of the Virtual and Mixed Reality - New Trends*. Randall Shumaker, (Ed.), Lecture Notes in Computer Science. Springer Berlin, 164–173. DOI : https://doi.org/10.1007/978-3-642-22021-0_19
- [12] Grace Eden, Benjamin Nanchen, Randolph Ramseyer, and Florian Evéquoz. 2017. On the road with an autonomous passenger shuttle. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems - CHI EA'17*. ACM Press, New York, New York, 1569–1576. DOI : <https://doi.org/10.1145/3027063.3053126>
- [13] Autumn Edwards, Chad Edwards, David Westerman, and Patric R. Spence. 2019. Initial expectations, interactions, and beyond with social robots. *Computers in Human Behavior* 90, 308–314. DOI : <https://doi.org/10.1016/j.chb.2018.08.042>
- [14] Nicholas Epley, Adam Waytz, and John T. Cacioppo. 2007. On seeing human: A three-factor theory of anthropomorphism. *Psychological review* 114, 4 (2007), 864–886. DOI : <https://doi.org/10.1037/0033-295X.114.4.864>
- [15] Daniel J. Fagnant and Kara Kockelman. 2015. Preparing a nation for autonomous vehicles: Opportunities, barriers and policy recommendations. *Transportation Research Part A: Policy and Practice* 77, 167–181. DOI : <https://doi.org/10.1016/j.tra.2015.04.003>
- [16] Julia Fink. 2012. Anthropomorphism and human likeness in the design of robots and human-robot interaction. In *Proceedings of the 4th International Conference on Social Robotics*. Shuzhi S. Ge, Oussama Khatib, John-John Cabibihan, Reid Simmons and Mary-Anne Williams (Eds.), Lecture Notes in Computer Science, Lecture Notes in Artificial Intelligence, 7621. Springer, Berlin, 199–208. DOI : <https://doi.org/10.1007/978-3-642-34103-820>

- [17] Terrence Fong, Illah Nourbakhsh, and Kerstin Dautenhahn. 2003. A survey of socially interactive robots. *Robotics and Autonomous Systems* 42, 3–4 (2003), 143–166. DOI : [https://doi.org/10.1016/S0921-8890\(02\)00372-X](https://doi.org/10.1016/S0921-8890(02)00372-X)
- [18] Kimberly Goodyear, Raja Parasuraman, Sergey Chernyak, Poornima Madhavan, Gopikrishna Deshpande, and Frank Krueger. 2016. Advice taking from humans and machines: An fMRI and effective connectivity study. *Frontiers in Human Neuroscience* 10, (2016), 542. DOI : <https://doi.org/10.3389/fnhum.2016.00542>
- [19] Kerstin S. Haring, Ariana Mosley, Sarah Pruznick, Julie Fleming, Kelly Satterfield, Ewart J. de Visser, Chad C. Tossell, and Gregory Funke. 2019. Robot authority in human-machine teams: Effects of human-like appearance on compliance. In *Proceedings of the 11th International Conference on Virtual, Augmented and Mixed Reality. Applications and Case Studies*. Jessie Y. Chen and Gino Fragomeni (Eds.), *Information Systems and Applications, incl. Internet/Web, and HCI*, 11575. Springer International Publishing; Imprint: Springer, Cham, 63–78. DOI : <https://doi.org/10.1007/978-3-030-21565-15>
- [20] Eddie Harmon-Jones and Judson Mills. 2019. An introduction to cognitive dissonance theory and an overview of current perspectives on the theory. In *Proceedings of the Cognitive Dissonance. Reexamining a Pivotal Theory in Psychology*. Eddie Harmon-Jones (Ed.). American Psychological Association, 3–24. DOI : <https://doi.org/10.1037/0001135-001>
- [21] Yutaka Hiroi and Akinori Ito. 2016. Influence of the height of a robot on comfortableness of verbal interaction. *IAENG International Journal of Computer Science* 43, 4 (2016), 447–455.
- [22] International Organization for Standardization. 2018. ISO 9241-11: 2018, Ergonomics of human-system interaction-Part 11: Usability: Definitions and concepts. ISO standards catalogue.
- [23] Michiel Jooisse, Manja Lohse, Jorge G. Perez, and Vanessa Evers. 2013. What you do is who you are: The role of task context in perceived social robot personality. In *Proceedings of the IEEE International Conference on Robotics and Automation*. IEEE, Piscataway, NJ, 2134–2139. DOI : <https://doi.org/10.1109/ICRA.2013.6630863>
- [24] Penny Kong, Henriette Cornet, and Fritz Frenkler. 2020. Closing the human-machine disconnect: Design requirements for two extreme companions for future autonomous mobility. In *Proceedings of the 16th International Design Conference - DESIGN 2020*.
- [25] Alexandra Kuznetsova, Per B. Brockhoff, and Rune H. B. Christensen. 2017. lmerTest Package: Tests in linear mixed effects models. *Journal of Statistical Software* 82, 1 (2017), 1–26. DOI : <https://doi.org/10.18637/jss.v082.i13>
- [26] Simon R. Larmour, Henriette Bergström, Christopher T. A. Gillen, and Adelle E. Forth. 2015. Behind the confession: Relating false confession, interrogative compliance, personality traits, and psychopathy. *Journal of Police Criminal Psychology* 30, 2 (2015), 94–102. DOI : <https://doi.org/10.1007/s11896-014-9144-3>
- [27] Gael Laurans, Pieter M. A. Desmet, and Paul Hekkert. 2009. The emotion slider: A self-report device for the continuous measurement of emotion. In *Proceedings of the 3rd International Conference on Affective Computing and Intelligent Interaction and workshops, 2009*. IEEE, Piscataway, NJ, 1–6. DOI : <https://doi.org/10.1109/ACII.2009.5349539>
- [28] David Leffman and Richard Lim. 2018. *The Rough Guide to Malaysia, Singapore & Brunei* (9th edition). Rough guides. Rough Guides Ltd, [London].
- [29] Holger Luczak, Matthias Roetting, and Ludger Schmidt. 2003. Let’s talk: Anthropomorphization as means to cope with stress of interacting with technical devices. *Ergonomics* 46, 13–14, 1361–1374. DOI : <https://doi.org/10.1080/00140130310001610883>
- [30] Ruth Madigan, Tyron Louw, Marc Dziennus, Tatiana Graindorge, Erik Ortega, Matthieu Graindorge, and Natasha Merat. 2016. Acceptance of automated road transport systems: An adaptation of the UTAUT Model. *Transportation Research Procedia* 14, (2016), 2217–2226. DOI : <https://doi.org/10.1016/j.trpro.2016.05.237>
- [31] Tim McDonald. 2018, October 29. *The Economics of Change. The Cost of Keeping Singapore Squeaky Clean*. Retrieved 11 Aug, 2021 from <https://www.bbc.com/worklife/article/20181025-the-cost-of-keeping-singapore-squeaky-clean>.
- [32] Alexandra Millonig and Peter Fröhlich. 2018. Where autonomous buses might and might not bridge the gaps in the 4 a’s of public transport passenger needs. In *Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. ACM, 291–297. DOI : <https://doi.org/10.1145/3239060.3239079>
- [33] Masahiro Mori, Karl MacDorman, and Norri Kageki. 2012. The uncanny valley [from the field]. *The Uncanny Valley [From the Field]* 19, 2 (2012), 98–100. DOI : <https://doi.org/10.1109/MRA.2012.2192811>
- [34] Clifford Nass, Jonathan Steuer, and Ellen R. Tauber. 1994. Computers are social actors. In *Proceedings of the Conference Companion on Human Factors in Computing Systems*. ACM Press, New York, New York, 204. DOI : <https://doi.org/10.1145/259963.260288>
- [35] Manisha Natarajan and Matthew Gombolay. 2020. Effects of anthropomorphism and accountability on trust in human robot interaction. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*. ACM, New York, NY, 33–42. DOI : <https://doi.org/10.1145/3319502.3374839>
- [36] Teron Nguyen, Meng Xie, Xiaodong Liu, Nimal Arunachalam, Andreas Rau, Bernhard Lechner, Fritz Busch, and Y. Wong. 2019. Platooning of autonomous public transport vehicles: The influence of ride comfort on travel delay. *Sustainability* 11, 19 (2019), 5237. DOI : <https://doi.org/10.3390/su11195237>

- [37] Xueni Pan and Antonia F. D. C. Hamilton. 2018. Why and how to use virtual reality to study human social interaction: The challenges of exploring a new research landscape. *British Journal of Psychology* 109, 3 (2018), 395–417. DOI : <https://doi.org/10.1111/bjop.12290>
- [38] Elizabeth Phillips, Xuan Zhao, Daniel Ullman, and Bertram F. Malle. 2018. What is human-like? Decomposing robots' human-like appearance using the anthropomorphic roBOT (ABOT) database. In *HRI'18 Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*. 105–113 DOI : <https://doi.org/10.1145/3171221.3171268>
- [39] R. Core Team. 2017. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- [40] M. A. J. Roubroeks, J. R. C. Ham, and C. J. H. Midden. 2010. The dominant robot: Threatening robots cause psychological reactance, especially when they have incongruent goals. In *Proceedings of the Persuasive Technology*. David Hutchison, Takeo Kanade, Josef Kittler, Jon M. Kleinberg, Friedemann Mattern, John C. Mitchell, Moni Naor, Oscar Nierstrasz, C. Pandu Rangan, Bernhard Steffen, Madhu Sudan, Demetri Terzopoulos, Doug Tygar, Moshe Y. Vardi, Gerhard Weikum, Thomas Ploug, Per Hasle and Harri Oinas-Kukkonen. (Eds.), Lecture Notes in Computer Science. Springer Berlin, 174–184. DOI : https://doi.org/10.1007/978-3-642-13226-1_18
- [41] James Russell. 2003. Core affect and the psychological construction of emotion. *Psychological Review* 110, 1 (2003), 145–172. DOI : <https://doi.org/10.1037/0033-295X.110.1.145>
- [42] Maha Salem, Friederike Eyssel, Katharina Rohlfing, Stefan Kopp, and Frank Joublin. 2011. Effects of gesture on the perception of psychological anthropomorphism: A case study with a humanoid robot. In *Proceedings of the 3rd International Conference on Social Robotics*. Lecture Notes in Computer Science, Lecture Notes in Artificial Intelligence, 7072. Springer, Berlin, 31–41.
- [43] Kristin E. Schaefer, Jessie Y. C. Chen, James L. Szalma, and P. A. Hancock. 2016. A meta-analysis of factors influencing the development of trust in automation: Implications for understanding autonomy in future systems. *Human Factors* 58, 3 (2016), 377–400. DOI : <https://doi.org/10.1177/00187208166634228>
- [44] Alessandra Sciutti, Martina Mara, Vincenzo Tagliasco, and Giulio Sandini. 2018. Humanizing human-robot interaction: On the importance of mutual understanding. *IEEE Technology and Society Magazine* 37, 1 (2018), 22–29. DOI : <https://doi.org/10.1109/MTS.2018.2795095>
- [45] Kazuhiko Shinozawa, Futoshi Naya, Junji Yamato, and Kiyoshi Kogure. 2005. Differences in effect of robot and screen agent recommendations on human decision-making. *International Journal of Human-Computer Studies* 62, 2 (2005), 267–279. DOI : <https://doi.org/10.1016/j.ijhcs.2004.11.003>
- [46] Catharina V. Smedegaard. 2019. Reframing the role of novelty within social HRI: from noise to information. In *Proceedings of the 14th ACM/IEEE International Conference on Human-Robot Interaction*. IEEE, Piscataway, NJ, 411–420. DOI : <https://doi.org/10.1109/HRI.2019.8673219>
- [47] 2019. *The ABOT Database* (2019). Retrieved January 20, 2020 from <https://abotdatabase.info/>.
- [48] Elena Torta, Franz Werner, David O. Johnson, James F. Juola, Raymond H. Cuijpers, Marco Bazzani, Johannes Oberzaucher, John Lemberger, Hadas Lewy, and Joseph Bregman. 2014. Evaluation of a small socially-assistive humanoid robot in intelligent homes for the care of the elderly. *Journal of Intelligent & Robotic Systems* 76, 1 (2014), 57–71. DOI : <https://doi.org/10.1007/s10846-013-0019-0>
- [49] Carl A. Trocki. 2006. *Singapore. Wealth, Power and the Culture of Control*. Routledge.
- [50] Fang-Wu Tung. 2016. Child perception of humanoid robot appearance and behavior. *International Journal of Human-Computer Interaction* 32, 6 (2016), 493–502. DOI : <https://doi.org/10.1080/10447318.2016.1172808>
- [51] Lis Tussyadiah and Graham Miller. 2019. Nudged by a robot: Responses to agency and feedback. *Annals of Tourism Research* 78, (2019), 102752. DOI : <https://doi.org/10.1016/j.annals.2019.102752>
- [52] Eva Wiese, Tyler Shaw, Daniel Lofaro, and Carryl Baldwin. 2017. Designing artificial agents as social companions. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 61, 1 (2017), 1604–1608. DOI : <https://doi.org/10.1177/1541931213601764>
- [53] Luc Wijnen, Paul Bremner, Severin Lemaignan, and Manuel Giuliani. 2020. Performing human-robot interaction user studies in virtual reality *. In *Proceedings of the 2020 29th IEEE International Conference on Robot and Human Interactive Communication*. IEEE, Piscataway, NJ, 794. DOI : <https://doi.org/10.1109/RO-MAN47096.2020.9223521>
- [54] Tao Zhang, David B. Kaber, Biwen Zhu, Manida Swangnetr, Prithima Mosaly, and Lashanda Hodge. 2010. Service robot feature design effects on user perceptions and emotional responses. *Intelligent Service Robotics* 3, 2 (2010), 73–88. DOI : <https://doi.org/10.1007/s11370-010-0060-9>
- [55] Jakub Zlotowski, Diane Proudfoot, Kumar Yogeewaran, and Christoph Bartneck. 2015. Anthropomorphism: Opportunities and challenges in human–robot interaction. *International Journal of Social Robotics* 7, 3 (2015), 347–360. DOI : <https://doi.org/10.1007/s12369-014-0267-6>

Received January 2021; revised August 2021; accepted November 2021