What is Appropriate? On the Assessment of Human-Robot Proxemics for Casual Encounters in Closed Environments

Nicolas E. Neef[1] · Sarah Zabel[1] · Mathis Lauckner[2] · Siegmar Otto[1]

Abstract

Increasingly autonomous robots become more and more prevalent in daily life and their proximity to humans may affect human well-being and comfort. Consequently, researchers have begun to study the effect of robotic presence on humans and to establish distance rules. However, studies on human-robot proxemics rely on various concepts (e.g. safety, comfort, perceived safety and expectation conformity) to measure the appropriateness of distances which can affect the outcomes. The impact of using diverging operationalization has not been studied explicitly, thus the first aim of our research was to fill this gap. In two experiments (combined $N = 80$), placing participants in indirect hallway human-robot interactions, we found that the way appropriateness is operationalized has a significant impact on the results for lateral passing and frontal approaches. The second goal was to gain new insights into the influence of robot appearance on appropriate proximity. Using an ad-hoc created appropriateness scale we reveal that for robots displaying human faces on screens, closer distances are perceived to be appropriate. Our study provides valuable insights into the relationship between measurement methods, robot appearance, and appropriateness, and offers practical recommendations for future research and development in the field of social robotics.

1 Introduction

Humans will encounter or even live together with increasingly autonomous robots in the years to come and the majority of human-robot encounters will take place without direct interaction (e.g., when robots carry out tasks independently, such as transportation or vacuuming). But we don't need to look far into the future, because already today, autonomous robots are deployed as museum tour guides [1], work closely with humans as collaborative industrial robots, or transport goods in public spaces [2]. While the number of robots surges, it is imperative for humans to adapt to their presence [3]. In order to support this adaptation, robots should obey fundamental human rules of spatial behavior to integrate smoothly into societies still consisting of people that have no or limited prior exposure to robots. In environments where human-robot interactions occur in close physical proximity, such as in a hallway, adhering to these behavior rules is particularly important. Therefore, in order to facilitate seamless integration of robots into human environments, it is crucial to understand human reactions to physical encounters with robots. This knowledge can then inform the programming of robotic motion behavior [4]. The challenge lies in determining appropriate practical distances for indirect interactions, because the best possible distance – usually the greatest, e.g., [5] – may not always be feasible in crowded spaces or narrow corridors. Hence, the development of motion algorithms for autonomous robots must consider appropriate distances that maximize human comfort and minimize restrictions on the robot's mobility. Numerous studies have been in search for the best practical distances, e.g., [5–18], and they arrived at varying estimates of appropriate distances with potentially practical consequences for the implementation in robot motion planning

✉ Nicolas E. Neef
nicolas.neef@uni-hohenheim.de

[1] Institute of Education, Work and Society, Dept. of Sustainable Development and Change, University of Hohenheim, Wollgrasweg 49, 70599 Stuttgart, Germany

[2] San Francisco, CA, USA

algorithms. These varying estimates stem from the use of (slightly) different adequate operationalizations of the construct appropriateness. For example, if perceived safety, e.g., [19], is used as a concept, different distances could be interpreted as appropriate than if spatial comfort, e.g., [20], is measured. In a practical scenario, a robot can only maintain a single physically "appropriate" distance. As a result, developers must either select one particular operationalization of appropriateness, such as perceived safety, and its corresponding measurement, or they may opt for a combined measure of appropriateness for implementation purposes. To determine the feasibility and conditions under which an aggregate measure of appropriateness should be developed and utilized, it is necessary to compare the outcomes of various definitions and their associated measurements, with the goal of identifying a "one-size-fits-all" solution. Thus, the purpose of this study is to specifically investigate and discuss the impact of using various concepts, their measures, and their practical implementation (i.e., operationalization) to evaluate human-robot interaction. Specifically, we will focus on the two most frequent studied forms of indirect interaction, that is, lateral passing of a robot and frontal approach [11]. Further, we aim to determine if the measures employed in previous studies can ad-hoc form an aggregate measure of the appropriateness of spatial distance between robots and humans successfully. Finally, by applying this ad-hoc appropriateness scale, we will extend on existing research regarding the effect of robot appearance on distance preferences. For this purpose we use two different robots - one of them in two variants.

2 Theory

2.1 Concepts of Appropriateness in Human-Robot Proxemics

Human spatial behavior has been extensively studied in anthropology and psychology [10, 21, 22]. In his early works Hall [22] conceptualized the space (proxemics) around a person and divided it into four subspaces, the *intimate zone* ranging from 0 to 0.45 m, the *personal zone* ranging from 0.46 to 1.20 m, the *social zone* ranging from 1.21 to 3.50 m and the *public zone* from above 3.51 m. Proxemics represents a central social convention that encompasses the relative positioning and orientation of interacting [23]. In addition, they determine the quality of the interaction as well as the well-being of the interaction partners. Spatial behavior, deviating from these social conventions, for example inappropriately invading someone's *personal zone*, can cause people stress, anxiety, or discomfort [24–26]. Previous research has demonstrated the existence of similarities in human-human proxemics and human-robot proxemics [7]. It has been proposed that underlying psychological motives, such as threat or gaze behavior, that guide human-human proxemics also apply to human-robot proxemics [12, 27]. This is likely due to the fact that robots actively create a social space [28]. Several studies have examined the relationship of spatial proximity in indirect human-robot interactions, and it has been found consistently that humans tend to prefer greater distances in these interactions and have more positive impressions of robots when they are encountered at greater distances [5–18]. Recently, even an extensive model of personal space in relation to human-robot interactions has been introduced [11].

2.1.1 Problem Statement

It is evident, that the primary objective of most human-robot proxemics studies is to evaluate the appropriateness of human-robot distances through a particular operationalization of the construct appropriateness. However, there is currently a lack of discussion on how appropriateness should be operationalized – even though, practically, a robot can only chose a certain distance at one point in time or motion. Thus, depending on the specific operationalization of appropriateness and hence varying measures, different outcomes are to be expected. For instance, if the evaluation of appropriateness is based on **spatial comfort** (e.g., "How comfortable were you with the passing of the robot?"), **perceived safety** (e.g., "I feel that this situation is dangerous for the passerby."; 19), **expectation conformity** (e.g., "The robots motion behavior was exactly how I expected it"), **or general motion acceptance** (e.g., "The autonomous assistant's motion behavior was good."), the subjective evaluation of appropriate distances will likely differ. In other terms, most studies aim to determine the most appropriate distance, but rely on different operationalizations of appropriateness as their dependent measure. As a result, these different approaches are likely to lead to divergent assessments of appropriate distances due to participants' potential to understand and/or weigh concepts, such as safety or discomfort, in different ways. Moreover, this divergence effect could be amplified by the fact that most of the studies cited in this article use only single-item measures. While this approach can be effective in measuring well-defined and specific constructs [29–31] single-item measures are prone to measurement error and therefore lack reliability [32–36]. Nevertheless it does seem likely that most readily applied measures each capture some of the nuances of the overall appropriateness in human-robot proxemics. Therefore, we will base our study on the most commonly used operationalizations of appropriateness – presented in the following.

2.1.2 Spatial Comfort / Discomfort Rating

A commonly used approach is to ask participants about their level of comfort in indirect human-robot interactions [5, 11, 16, 17, 20] Conversely, subjects can also be asked if they feel uncomfortable [13, 37]. The natural assumption is that a distance perceived as comfortable or not uncomfortable is appropriate.

2.1.3 Perceived Safety

The logic of this measurement is that it is assumed that people are more likely to accept an autonomous motion if they feel safe [19]. Specifically, a situation that is objectively safe (i.e., it is not harmful in a physical way) but not perceived as such can lead to negative feelings and stress [38]. Hence an appropriate distance should be perceived as subjectively safe.

2.1.4 Expectation Conformity

People unconsciously infer expectations from the implicit rules for spatial behavior. People who do not meet these expectations cause an unpleasant affective state (e.g., fear) in their counterpart and are therefore perceived as less desirable, e.g., [39]. As mentioned, these behavioral rules and respectively the derived expectations can, to some extent, be transferred to human-robot interactions [7, 40]. Thus, the more likely a robot is to behave according to the expectations of the human interaction partner, the more appropriate it should be perceived.

2.1.5 General Motion Acceptance

It has been postulated that acceptance is a crucial metric to assess the quality of human-robot interactions [41]. Acceptance has been used in numerous studies to evaluate human interaction with social robot systems [42, 43]. As components of the general acceptance of robotic motion behavior, Oestreicher [43] identified predictableness, trustworthiness, and subjectively good motion as positive and surprisingness, strangeness, and uncomforted as negative predictors. Thus this measurement instrument, in contrast to those presented so far, by definition exists of multiple items, as acceptance as such is already a fanned out construct.

2.1.6 Direct Distance

Even though this approach differs from the others substantially, we want to discuss it, due to its relevance. Several studies were conducted in which researchers allowed their subjects to set their own preferred distance from a robot by controlling it with a remote control. In a study by Walters et al. [7], the robots were moved to within 60 cm of the subjects by default before participants could adjust its position forward and backward to indicate what they felt to be an appropriate distance. In similar studies researchers had participants steer the robot completely freely as close to them as they preferred [14] or stop the approach at an appropriate distance [44]. While this is a suitable method to form reference values for appropriate human-robot proxemics, these findings have limited practical use for this study, since (a) the control of the robot rests within the subjects, which has restricted predictive validity about autonomous robotic motion, and (b) the interaction is direct, not indirect, due to the steering control. The work-around for this study is to ask people directly right after the interaction whether a certain pre-programmed distance is perceived as too close, without placing the robots control in the participants' hands.

2.2 Robot Features

Finally, we would like to discuss appropriateness differences with respect to different robots. While many robots are endowed with human-like attributes like arms or bodies [45], there is still little research on how human characteristics influence the appropriate proximity of a robot. It has been shown that a higher level of a human-like appearance leads to more positive impressions of the robot, while at the same time, distance preferences towards humanoid robots are found to be slightly larger than for mechanoids in these studies [8, 11, 43, 46–48]. In one of the first studies, Syrdal et al. [9] found a robot equipped with a anthropomorphic robot face was kept at a greater distance than the same robot without the face. Similar results have been presented more recently by [48, 49]. This could be explained by the non-linear relationship between human-likeliness of an artificial object and its acceptance [13, 50]. Robots that have too human-like features can appear as creepy and thus result in larger distance preferences [13]. Additionally, it is worth noting that while anthropomorphizing robots with masks or other physical features has been explored, transmitting an actual human face via a monitor has not yet been investigated as a potential method of traversing the Uncanny Valley [50, 51]. Furthermore, research on the effect of robot height on human comfort levels has yielded mixed results. For example, Syrdal et al. [9] found no preference difference between a height of 1.2 and 1.4 m when participants were standing, but Koay et al. [8] found that participants preferred shorter robots (1.2 m vs. 1.4 cm) when sitting. Studies have also shown that humans feel more comfortable getting closer to robots that are lower than knee height (around 51.6 cm on average) compared to taller robots [46, 52, 53]. However, it appears that there is a gap in research

comparing robots with a height of around one meter, which have been studied independently, to taller, almost human-sized robots [48]. A direct comparison between the two has yet to be conducted.

Before we proceed with the method section, it is imperative to reiterate that our main objective is to demonstrate the associations between the various measurement instruments and the appropriateness of distances, rather than to devise an optimal appropriateness scale. The utilization of various robots serves primarily to consider potential interactions between different robots and the measurement instruments, and hence, to broaden our empirical basis. We do not criticize the contribution of previous studies, but rather want to pinpoint a methodological issue which should and can easily be overcome in order to foster building a common understanding in and of proxemics research. Further, by considering the diverse measurement instruments as nuances of a common appropriateness (i.e., an ad-hoc scale), we anticipate that new insights can be gleaned regarding the impact of robot height and of displaying a realistic face on a robot's appropriateness perception.

In consonance with recent findings [5], we anticipate that the measurement values of all measurement instruments (i.e., nuances of appropriateness) used will exhibit a progressive increase before stabilizing for all three robots. Additionally, we posit that displaying a natural face on a digital screen, as opposed to sculptural body parts, will have a positive impact on the robot's appropriateness in relation to distance. Furthermore, we anticipate interactions between distance and measurement instrument and measurement instrument and robot.

Table 1 *Item list for measuring appropriateness of distance in human-robot interaction*

Variable	Items
Spatial discomfort rating	The autonomous transport assistant's/ Beam's motion behavior made me feel uncomfortable.
Direct distance	How would you rate the autonomous transport assistant's/ Beam's maintained lateral/frontal distance towards you?
Expectation conformity	The autonomous transport assistant's/ Beam's motion behavior was exactly like I expected it.
Perceived safety	How safe did you feel around the autonomous transport assistant/ Beam?
General motion acceptance	The autonomous transport assistant's/Beam's motion behavior was good.
	The autonomous transport assistant's/ Beam's motion behavior was surprising.
	The autonomous transport assistant's/ Beam's motion behavior was predictable.
	The autonomous transport assistant's/ Beam's motion behavior was strange.
	I would trust an autonomous transport assistant/ a Beam with such kind of distance behavior.
	The autonomous transport assistant's/ Beam's motion behavior was polite.

3 Methods[1]

3.1 Selection of Test Distances

It has generally been postulated that the personal zone (0.46 to 1.20 m; reserved for friends and family, as well as highly organized interactions) seems to be the most adequate spatial zone in human-robot interactions [54]. Further, in accordance with a recent study by Neggers and colleagues [11] we generally assume that the shape of the personal space with regard to human-robot proxemics is round-shaped (i.e., differences between right, left, front or rearward passing, do not seem to be significant). At the same time the appropriate distance seems to be larger for frontal approaches than for lateral (and frontal or rearward) passes, e.g., [14, 44, 49]. In the case of lateral passing, Pacchierotti and colleagues [16] demonstrated that 0.4 m is more comfortable than 0.3 m or 0.2 m. In a recent study, Neggers and colleagues [5] were able to show that the comfort of the participants improved with increasing lateral passing distance, stabilizing at about 1 m. Lauckner and colleagues [14] found that a mean passing distance below 0.46 cm for lateral passing and below 0.77 m for frontal encounters was not desired by participants in a hallway setting, see also [9]. As mentioned, closer distances may sometimes be necessary from a practical point of view. Thus, distances for lateral passing were set between 0.20 and 0.60 m (step 0.10 m) and somewhat larger for frontal encounters, i.e., between 0.50 and 1.10 m (step 0.15 m).

3.2 Measures of Appropriateness

In our study, we employed a set of measures as described above and operationalized them as survey items (see Table 1). Participants were asked to rate their level of discomfort, expectation conformity, perceived safety and general motion acceptance of the robot's motion using a 5-point bipolar Likert scale, where 1 represents "strongly disagree" and 5 represents "strongly agree". Additionally, participants were asked to rate the direct distance from the robot using a 5-point unipolar scale, where 1 represents "much too close" and 5 represents "not too close at all". The original study also included two hypothetical questions that aimed to investigate the extent to which the participants compared the robot's behavior to that of a human. These were not considered due to the research question of this study.

---

[1] This Study uses data from a Dissertation Thesis. The Experiments were Originally Conducted by Lauckner [13]▲

Fig. 1 Transport Assistant (Bosch Research Prototype)

3.3 Tested Robots

Throughout the two conducted experiments, two differently machine-like looking robots are employed. The first robot - the "transport assistant" (TA) - is a self-constructed and designed research prototype by Robert Bosch GmbH, Germany (see Fig. 1). It is internally used for a wide range of soft- and hardware tests, and provides a manually and autonomously maneuverable research platform for conducting human-robot interaction experiments.

In particular, the TA comprises a prototypic cuboid-like mock-up body attached to an omni-directional mobile platform provided by KUKA Roboter GmbH, Germany. The technical equipment is covered by a prototypic semi-transparent white shell (the mock-up body). In the front, a black display is attached which is without any function for all experiments in the present work and was constantly switched off. In total, the entire robot prototype is 0.73 m deep, 0.46 m wide and 1.05 m high. By localizing itself based on laser data, the TA can autonomously move in a
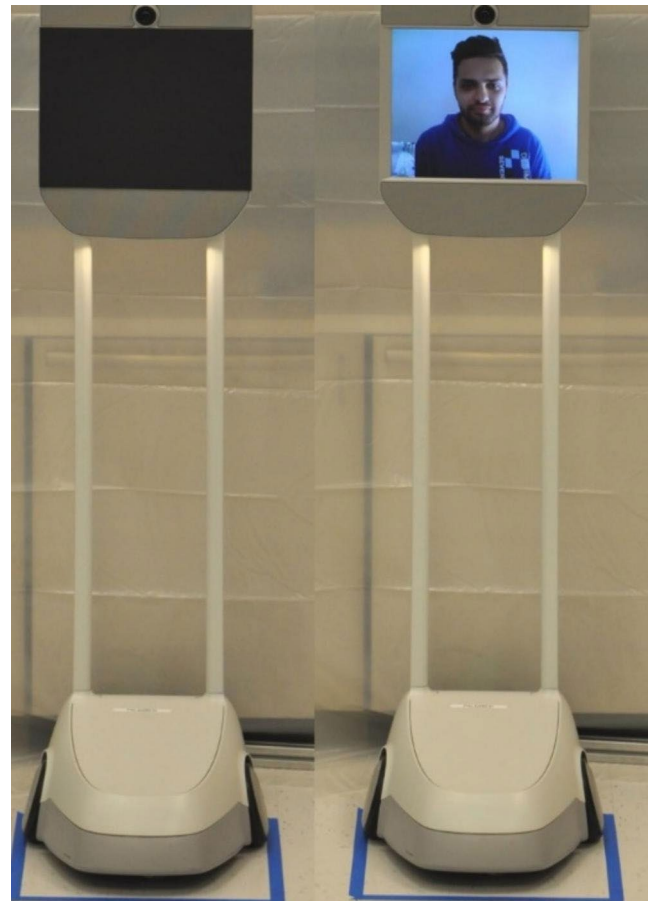

Fig. 2 Beam without face (left) and Beam with face (right)

hallway or in an open area. In addition, it is possible to remotely control its movements by using a Logitech wireless gamepad F710 with two analog control elements.

In contrast to the TA and its prototypic characteristic, the second employed robot – Beam (height = 158.7 cm, width = 50.8, depth = 66 cm) - is an already commercially available system by GoBe Robots [55]. Essentially, it is a semi-autonomous telepresence system and can neither localize nor move autonomously. However, Beam posed an ideal second test platform for the purpose of this research. Similar to the TA, Beam also has a machine-like appearance (see Fig. 2).

The system features a 17 inch screen, a six-microphone array enabling remote users to localize directions of sound, two wide-angle high resolution cameras (one front facing and one down facing), a digital zoom and two radio modules for seamless switching between access points on a wireless network. Importantly, Beam can be controlled remotely from personal computers using keyboard or mouse devices, but does not operate autonomously. A typical interaction involves seeing and talking with the remote user's face as presented in Fig. 2. This functionality is used for a manipulation of Beam's level of human-likeness.

3.4 Experiment I - Autonomously Approaching and Passing Mechanoid

3.4.1 Participants Experiment I

Forty Germans participated in this experiment. The 20 male and 20 female subjects had an average age of 29.2 years ($SD = 5.81$). All participants received a 30€ voucher for their participation. Further, all participants signed of a letter of consent.

3.4.2 Procedure Experiment I

The experiment was conducted in the robotics lab of the Robert Bosch GmbH in Schwieberdingen, Germany. The lab was divided by a wall covered with white film and had a door-like entrance. Entering induced a feeling of being in a hallway with white walls. The simulated hallway was 6 m long and 2.90 m wide. These dimensions were chosen to ensure a sufficient amount of space regarding the experimental variations. In addition, the chosen hallway width approximately resembled a common hallway size in a hospital or a larger office space. The laboratory hallway-like setting is shown in Fig. 3. In addition, a small round table and a chair were placed in the rear left corner of the hallway (from the TA's point of view). The table was provided to the participants for completing the questionnaires, and the chair was used by the examiner to put off additional material, such as an iPad. The iPad was needed for launching the autonomous movements of the TA.

The actual study began by familiarizing participants with the autonomous assistant, discussing data security concerns, and outlining the parameters of the experiment. In addition, participants were informed that the researcher had the ability to intervene during the mechanoid's autonomous behavior at any time. As soon as participants had no more questions, they were requested to exclusively observe the

Fig. 3 Hallway-like setting in Experiment I from the TA's view

autonomously approaching or passing mechanoid without changing their own position. For frontal approaches, subjects stood behind the short blue line (back right of Fig. 3). For lateral passing, they stood at the designated distance on the long straight blue line (in the left front half of Fig. 3). In order to explore subjects' sensations towards the robot's maintained frontal and lateral distances, in a first block of five trials, frontal distance was varied (0.5m / 0.65m / 0.8m / 0.95m / 1.1m). Accordingly, in a second block of five trials, the lateral passing distance was varied (0.2m / 0.3m / 0.4m / 0.5m / 0.6m). The resulting 10 experimental trials were completely randomized in order. This randomization aimed to take already proven habituation effects into account [7]. During all experimental trials the TA drove with a constant speed of 0.6m/s (acceleration: 2m/s², -2m/s²). Within experimental trials exploring frontal distances the robot started 4.5m in front of the subjects. In all trials exploring the lateral distances the TA's starting position was 3.2m in front of the and 1.2m to the left of the subjects. The experimenter remained beside the table and behind the subject during the autonomous drives of the TA. After each trial, the participants were instructed to spontaneously indicate their sensations towards the experienced proxemic behavior of the mechanoid in a questionnaire. The total experiment lasted around one hour.

3.5 Experiment II - Autonomously Approaching and Passing Mechanoid with a Face

3.5.1 Participants Experiment II

Among the 40 participating subjects were 22 (55%) females and 18 (45%) males with an average age of 35.2 years ($SD = 11.7$). The sample consisted mostly of US Americans as compared to mostly Germans in the first experiment. Furthermore, about half of the participants had a technical background. All participants received a $20 monetary compensation for their participation.

According to Hall [22] and Nanda and Warms [56], people from North America and Northern Europe both can be classified as non-contact cultures. Thus, the underlying spatial conventions of all participants were assumed to be comparable.

3.5.2 Procedure Experiment II

This experiment took place in a robotics lab in the Robert Bosch Research and Technology Center in Palo Alto, USA. However, the previously applied hallway-like setting was reconstructed meticulously. Thus, starting positions of Beam and participants comprised the same distances to each other. Moreover, as in the first experiment, a small round

table, which served the participants to complete the questionnaires, was placed in the rear left corner of the hallway (from Beam's point of view).

Beam's speed and acceleration values were identical to the first experiment (i.e., speed: 0.6 m/s and acceleration: 2 m/s², -2 m/s²). The essential manipulation in this experiment was the alteration of Beam. Either Beam's screen was turned off (blank) or a real human face was displayed. For the latter, a live video of a confederate was streamed on Beam's screen. This employee also served as the examiner's confederate, secretly operating the Beam throughout the whole experiment, i.e., the Wizard of OZ technique [54]. Both versions, Beam with face (fBeam) and without (Beam), are illustrated in Fig. 2.

The design of the experiment was an extended copy of the first experiment (see Fig. 4). Each block with five trials of the second experiment was extended by varying the version of the robot. Thus, the first block comprised a 5×2 mixed design with frontal distance (0.5 m / 0.65 m / 0.8 m / 0.95 m / 1.1 m) serving as a within-subjects factor and the two versions of Beam as a between-subjects factor. The second block also comprised a 5×2 mixed design with lateral passing distance (0.2 m / 0.3 m / 0.4 m / 0.5 m / 0.6 m) serving as a within-subjects factor and version of Beam serving as a between-subjects factor. 20 participants were randomly assigned to each block. As in the first experiment, the resulting 10 trials for each of Beam's versions were completely randomized in order. Questionnaires after each trial were identical to those used in the first experiment. However, further questions were included as a manipulation check to see if presenting a face had an impact on Beam's general visual impression. With semantic differential scales (GODSPEED I, II, III) [57], participants rated the animacy (whether the robot is perceived as lifelike, e.g. "dead" vs. "alive"), likeability (positive impressions towards the visual appearance or behavior of the robot, e.g., "unpleasant" vs. "pleasant") and human-likeness ("machine-like" vs. "human-like"). Furthermore, familiarity and uncanniness were assessed with one item (e.g., "The autonomous assistant somehow appears familiar to me.").
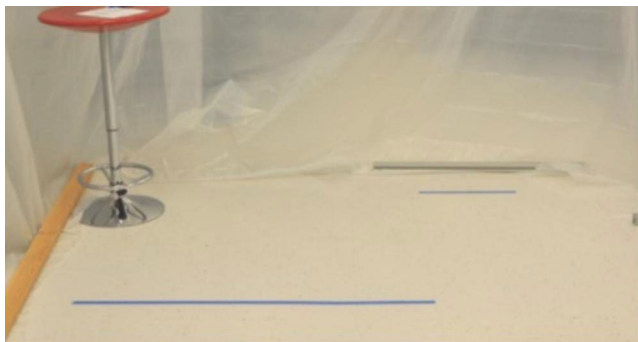

Fig. 4 Hallway-like setting in Experiment II from Beam's view

3.6 Data Handling

The data was processed with RStudio (version 2022.07.1). Packages used were car [58], catstats [59], corrplot [60], ez [61], ggpubr [62], Hmisc [63], janitor [64], openxlsx [65], RColorBrewer [66], remotes [67], rstatix [68], schoRsch [69] and tidyverse [70].

Data from the two experiments were transferred into a common data set for analysis and a score was calculated for the general acceptance scale consisting of the sum of the answers for all items (see Table 1) divided by the number of items (resulting possible range: 1–5). First, the general effect of presenting a face on Beam's screen on the robot's visual impression was tested by a series false discovery rate adjusted t-tests (GODSPEED-scale). Secondly, a correlation matrix was computed for the different measurement instruments. Thirdly, two separate mixed-model ANOVAs were computed for the analysis of appropriate lateral and frontal distances. The three robot types serving as the between factor and the five different measurements and the distances serving as the within factors. In situations where pairwise-comparison tests succeeded ANOVAs, the Benjamini-Hochberg correction was used to control the false discovery rate [71]. The single-item measures direct distance and spatial discomfort were inverted. In cases where the Mauchly's test of sphericity indicated that the assumption of sphericity was not met, the Greenhouse-Geisser sphericity correction was applied [72, 73]. Note that tables and figures not presented in the present article can be found as supplementary material in Online Resource 1.

3.6.1 Overall Appropriateness Scale

This scale is included in the ANOVA analysis as an interaction between distance and robot, since this interaction combines the different measurement instruments by design [73]. Mathematically the overall appropriateness scale is calculated by taking the total score of all items, divided by the number of all items (possible range from 1 to 5). The internal consistency was calculated for all distances and both scenarios (lateral and frontal). The different robot types were hereby not taken into account. Cronbach's α ranged from 0.72 to 0.89. A correlation heat map for the bivariate correlations of the measurements can be found in supplementary Fig. 4. Note, these are the mean correlations between the instruments across all three robots and all lateral and frontal distances. All measurement instruments used were found to correlate moderately to strongly with each other (range: $|r| = 0.42 - 0.81$). All correlations were significant $p < .05$.

4 Results

4.1 General Effect on the Visual Impression

Participants' ratings of the pictures presented in experiment two indicated no differences in likeability, human-likeness, familiarity and uncanniness between Beam and fBeam. However, perceived animacy was higher for latter fBeam ($t(38) = 1.99$, $p = .05$)). It can be inferred that the subjects did not perceive substantial visible variations between the two robots. In other words, the manipulation seemed not to have greatly affected the participants visual impression of the robot.

4.2 Lateral Passing Distance

A three-way mixed ANOVA was performed to evaluate the effects of measurement instrument, robot type (between factor) and distance (within factor) on the appropriate score assessed by the different measurement instruments. Since the three-way interaction of robot, distance and measurement instruments was not significant, the significant two-way interactions were deconstructed into simple main effects (see Table 2).

4.2.1 Simple Main Effects for Robot and Distance

The simple main effect of the type of robot on the combined appropriateness score (the total score across all measures), was statistically significant at the distances of 0.30 m ($F(2,397) = 10.50$, $p < .001$) and 0.60 m ($F(2,397) = 9.65$, $p < .001$). See supplementary Fig. 1 for a visualization. Consequently pairwise-comparison tests were conducted, while adjusting for familywise error [71]. For distance: 0.30 m they revealed significantly higher overall score of the fBeam compared to the Beam ($p < .001$, $g_{Hedges} = 0.64$) and the TA ($p = .015$, $g_{Hedges} = 0.31$) and a significantly higher overall score for the TA compared to the Beam ($p = .005$, $g_{Hedges} = 0.33$). For distance: 0.60 m the pair-wise tests revealed a significantly higher overall score for the TA compared to

the Beam ($p = .029$, $g_{Hedges} = 0.29$) and the fBeam ($p < .001$, $g_{Hedges} = 0.50$).

4.2.2 Simple Main Effects for Robot and Measurement Instrument

The simple main effect of robot type on the specific measurement score (i.e., the individual measurement instruments), while holding distance constant, was statistically significant for direct distance ($F(2,397) = 23.00$, $p < .001$) and for perceived safety ($F(2,397) = 3.73$, $p = .025$). See supplementary Fig. 2 for a visualization. The pairwise-comparison revealed a significantly higher direct distance score of the TA compared to the Beam ($p < .001$, $g_{Hedges} = 0.77$) and the fBeam ($p < .001$, $g_{Hedges} = 0.49$), as well as higher direct distance score for the fBeam compared to the Beam ($p = .027$, $g_{Hedges} = 0.37$). Further perceived safety score for the fBeam was significantly higher than for the TA ($p = .036$, $g_{Hedges} = 0.35$).

4.2.3 Simple Main Effects for Measurement Instrument and Distance

The simple main effect of the type of measurement instruments on measurement score, while ignoring the robot type, was statistically significant at all distances. For 0.20 m ($F(2.98,235.14) = 42.78$, $p < .001$), for 0.30 m ($F(2.81,222.11) = 24.51$, $p < .001$), for 0.40 m ($F(2.91,230.21) = 6.471$, $p < .001$), for 0.50 m ($F(3.01,237.44) = 8.31$, $p < .001$) and for 0.60 m ($F(3.06,242.04) = 11.17$, $p < .001$). See Fig. 5 for a visualization. The pairwise-comparison tests can be found in the supplementary Table 1.

4.3 Frontal Approach Distance

A three-way mixed ANOVA was performed to evaluate the effects of measurement instrument, robot type (between factor) and distance (within factor) on the appropriate score assessed by the different measurement instruments. Since the three-way interaction robot, distance and measurement instruments was not significant, the significant two-way interactions were deconstructed into simple main effects (see Table 3).

4.3.1 Simple Main Effects for Robot and Measurement Instrument

The simple main effect of robot type on the specific measurement score (i.e., the individual measurement instruments), while holding distance constant, was statistically significant only for direct distance ($F(2.397) = 20.9$, $p < .001$). See supplementary Fig. 3 for a visualization. The

Table 2 *Three-way mixed ANOVA results for lateral passing*

Effect	GG ε	df	F	partial η²
Robot		2,77	0.75	0.019
Distance	0.727	2.91, 223.92	67.03***	0.465
Measurement	0.586	2.34, 180.36	45.65***	0.372
Robot : Distance	0.727	5.82, 223.92	2.30*	0.056
Robot : Measure	0.586	4.68, 180.36	10.97***	0.222
Distance : Measure	0.626	10.02, 771.3	13.78***	0.152
Robot : Distance : Measure	0.626	20.03, 771.3	1.14	0.029

Note. *** p < .001, ** p < .01, * p < .05

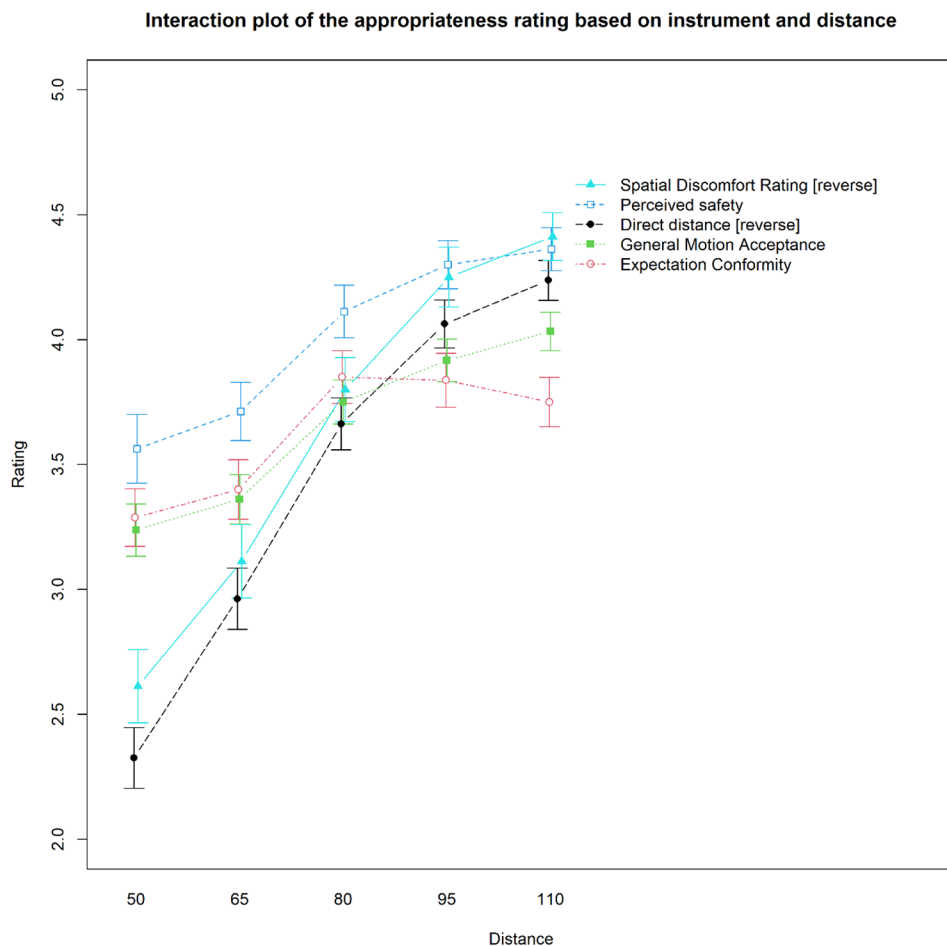**Fig. 5** Interaction plot distance and measurement instrument for lateral passing

Interaction plot of the appropriateness rating based on instrument and distance

Table 3 *Three-way mixed ANOVA results for frontal distances*

Effect	GG ε	df	F	partial η²
Robot		2,77	0.28	0.007
Distance	0.743	2.97, 228.75	71.25***	0.481
Measurement	0.749	3.00, 230.63	25.48***	0.249
Robot:Distance	0.743	5.94, 228.75	1.11	0.028
Robot:Measure	0.749	5.99, 230.63	7.79***	0.168
Distance:Measure	0.649	10.39, 799.9	14.27***	0.156
Robot:Distance:Measure	0.649	20.78, 799.9	1.09	0.028

Note. *** p < .001, ** p < .01, * p < .05

pairwise-comparison revealed a significant higher direct distance score for the TA compared to the Beam ($p < .001$, $g_{Hedges} = 0.65$) and the fBeam ($p < .001$, $g_{Hedges} = 0.64$).

4.3.2 Simple Main Effects for Measurement Instrument and Distance

The simple main effect of measurement instruments on measurement score, while ignoring the robot type, was statistically significant all distances. For 0.50 m ($F(3.38, 266.75) = 37.35$, $p < .001$), for 0.65 m ($F(3.20, 252.62) = 12.88$, $p < .001$), for 0.80 m ($F(3.14, 248.07) = 6.471$, $p < .001$), for

0.95 m ($F(3.02, 238.41) = 7.03$, $p < .001$) and for 1.10 m ($F(3.20, 252.96) = 14.60$, $p < .001$). See Fig. 6 for a visualization. The pairwise-comparison tests can be found in the supplementary Table 2.

5 Discussion

The objective of this study was to investigate variations in appropriate frontal approach and lateral passing distances of autonomously moving robots resulting from the application of various measurement approaches, or in other words, the operationalization of an appropriateness measure - with a focus on distances below and around one meter. Additionally, the appearance of three different robot types were considered: the shorter mechanoid (TA), taller mechanoid (Beam), and taller mechanoid with a real-life face (fBeam).

We neither found a significant three-way interaction between robot, distance and specific measurement instrument in lateral passing nor in frontal approach scenarios. Furthermore, we did not find a significant main effect for the appearance of the robot (TA, Beam or fBeam). The
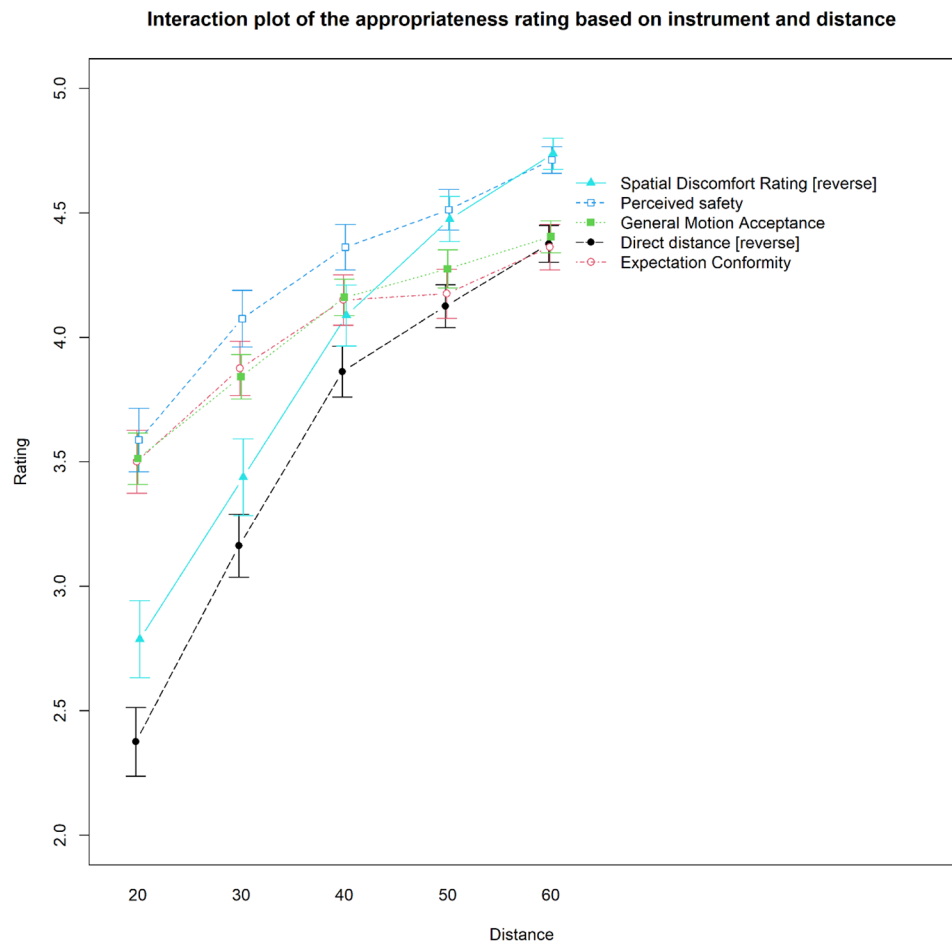
significant main effect of distance revealed that the appropriateness, regardless of the robot or the measurement instrument, increases with higher distances. The plotted results reveal a consistent trend, which is also observed in recent studies such as Neggers et al. [5, 11, 74]. Initially, the slope is steep, but it gradually becomes less pronounced over time (see Figs. 5 and 6). These results, replicating common findings, suggest that humans' underlying understanding of spatial conventions regarding human-human proxemics guide their feelings, sensations and expectations towards robots too [5, 11, 16, 17, 54]. This was true for lateral passing as well as for frontal approach. More importantly for this study, the choice of measurement instruments had a significant effect on participants responses. This confirms our assumption, that the different operationalizations of appropriateness in relation to proximity leads to different findings. However, since the interactions between measurement instrument and distance and between type of robot and measurement were significant we will focus our interpretation on them, starting with the earlier [73].

5.1 Determinants of Appropriateness

The main objective of this article was to demonstrate how various operationalizations of appropriateness and measures thereof can lead to different results relevant for human-robot proxemics (in indirect interactions). As expected, a significant interaction between measurement instrument and measurement specific scores was found for lateral passing as well as frontal approach. Specifically significant differences were recorded for all distances with respect to the underlying measurement instrument. Furthermore, these differences varied depending on the distance. At certain distances, one measurement instrument was more strict (i.e., indicated a lesser appropriateness), at certain other distances, the other. Similar variations were observed in the interactions between the robot and the specific measurement instrument. For instance, while the TA received the highest score for direct distance, it had the lowest score for spatial comfort (both for lateral passing and frontal approach). This means that if direct distance were used as the sole operationalization of appropriateness measure, the interpretation would be that the TA behaved most appropriately in terms of

distance. If the interpretation were to be based on the spatial comfort measure the TA would be deemed most *inappropriate*. Hence, without delving into the specifics of individual pairwise comparisons, it is evident that the subjective determination of the appropriateness of a distance and a robot is dependent on the method of measurement, even when the measures are highly correlated, as in this case.

We do not propose to replace any of the conceptual approaches, as they inherently represent real world constructs that are meaningful to people and have an intrinsic meaning for the assessment of the appropriateness of human-robot proxemics. However, combining them into a more general and practically widely applicable measure will help to get a more common measure of robots' appropriate movement and distancing behavior. After all, appropriate robotic motion behavior – just like human motion behavior, should reflect trustworthiness, predictability, no surprises, no strangeness and comfortable distances [40, 43, 75]. At the same time, it should take into account the subjective sense of safety [38] and well-being of humans [5, 11]. Finally, it also seems sensical (if possible) to directly query whether a distance is perceived as too close [13]. In our study we therefore took this step of unifying the measurement instruments and created an ad hoc scale of general appropriateness that includes all the nuances of appropriateness presented. Thus, the current scale, although with room for improvement, but with its high level of internal consistency enables comparison of the three robots on an overall scale.

5.2 Effects of Robot Features on General Appropriateness

One objective of this study was to study if robotic appearance differences manifest themselves in response variation. The ad-hoc formed appropriateness scale was used to compare the robots with each other, yielding mixed results. The interaction of robot and distance was significant for lateral passing scenarios, but not for frontal approaches. However, our findings for lateral passing align well with previous research [46, 48]. Specifically, the TA, which is the shortest robot in our study, was perceived as more appropriate for use at very close distances (30 cm) compared to its taller counterpart, the Beam robot. In comparison to the studies conducted by Syrdal et al. [9] with a height difference of 20 cm, it is likely that the 50 cm difference between our robots had an impact on the assessment of the robots' appropriateness. However, the fact that the fBeam was also considered more appropriate than the Beam at very short distances contrasts somewhat with previous results [52]. In studies which used sculpted human-like features to alter the appearance of a mechanoid, it was classically found that

they tended to be kept at a distance, e.g., [8]. However, it is also worth noting that viewing a real face on a screen is different from viewing a human-like "component." While the latter can be perceived as unsettling or creepy, which may explain previous results, the former should not elicit the Uncanny Valley effect. Interestingly, the differences found between the Beam and fBeam occurred even though their visual impression only differed in terms of animacy, but not in terms of familiarity, uncanniness, likability, or human-likeliness, as noted in reference [57]. This may be because the GODSPEED scale used in the study did not fully mask the participants' perceptions or was not sensitive enough, and/or the perception differences were only subliminally perceived by the participants. The authors suspect that this could be due to the fact that the robot's shell has more influence on how it is perceived visually than the face on the screen. In addition, the human-likeliness items in particular have come under some criticism recently [76]. Nevertheless, presenting an actual human face had an effect, even if it did not translate fully to the reported visual perception of the robot. If possible it can be thus useful to equip a mechanoid with human-like features. However, these should be a face presented on a screen at best and sculptural body parts should be avoided [48]. It's important to note that the above results should be interpreted with caution, as the measurement instruments used in the study, when combined, do not represent an established scale, despite yielding good reliability.

5.3 Methodological Implications and Future Research

The current study demonstrates that using different measures of appropriateness of distance between a robot and a human can lead to unreliable and widely varying results and hence, in difficulty comparing these results. Such diverging and unreliable findings can hinder theory building themselves – and even more important – find their way into real life robotic motion algorithms. To increase reliability and improve the practical usefulness of further proximity studies, we suggest to combine known measures of human-robot proxemics into a broader and more reliable one. Our first attempt that resulted in a highly reliable scale – not just because of the higher number of items – [77, 78] – is encouraging[2]. Nevertheless, we see additional room for

---

[2] Because of the single-item nature of the specific measurement instrument, we could not calculate their reliability (e.g., Cronbach's Alpha) or in other words, their measurement error. However, even if we would assume a rather good reliability of 0.7 (which is most likely not the case for singe-items measures), the bi-variate correlations between the specific measurement instruments – corrected for the measurement error – would be close to unity. This would again justify their unification.

improvement. For one, using similar response options as far as possible for all items (i.e., the same Likert scale) will increase ease of use for participants. To value attraction and aversion in persons equally, negative and positive items should be considered in a balanced number [79–81]. Furthermore, in addition to classical questionnaires – that are often used due to their economic nature, utilizing a broader mix of methods would help to increase especially the validity of such studies further and even help to overcome the common method bias [82, 83]. In the context of proxemics, the "think-aloud" technique is especially adequate. This qualitative approach involves asking the participant to verbally express their thoughts and considerations as they occur during the course of an interaction. This enables researchers to gain insight into the individual's internal states in real-time, as opposed to retrospectively through questionnaires [84]. Especially in the case of cognitively unchallenging tasks (such as standing next to a robot), this method is useful because the subjects can fully concentrate on their sensations [85]. For successful applications of this method see [86, 87]. Another, more objective, approach is to utilize physiological measurements. One promising physiological measurement is skin conductance, which can provide valuable insights into the subject's physiological responses to robotic presence. Relationships between human-proxemics and physiological responses have been known for a long time, and skin conductance in particular is a simple measure to implement [88]. In addition to skin conductance, and also relatively easy to implement, is the measurement of heart rate [89]. It should be noted, however, that physiological measurements alone can also lead to erroneous conclusions. In order to gain the most reliable insights, a mixed-method approach, that combines questionnaires, introspection and physiological measures would be the highest standard.

Finally, we argue that adequate measurement of appropriate distances is necessary for smooth social integration of robots. This helps provide a solid foundation for scientists' claims and allows practitioners to make informed programming decisions, resulting in robots that are more likely to behave appropriately when interacting with humans.

6 Limitations

Some limitations of this work are to be pointed out. Strictly speaking, one could argue that when the Beam was presented with a face, the human interaction partners no longer perceived the frontal interaction as indirect. They might have expected a direct interaction, as they would have with a human approaching them. However, since they underwent several trials with the robot, it can be assumed that this initial violation of expectations, which may have been present in the first trial, lost its impact on the responses. Even though for a study that examines human-robot interactions, see e.g., [17, 90] our sample was comparably large, the sample size was still small in general terms, and thus, especially the absolute distances might not be generalizable. In addition to the generally small sample, only participants from WEIRD countries were interviewed. Since it is known that the nature of the personal space can differ between cultures [91, 92], this also limits the transferability, see [93] specifically for inter-cultural differences in human-robot proxemics. Furthermore, in the first experiment, the visual appearance of the TA was not assessed, unlike in the experiment involving Beam. As a result, the differences between the robots could not be attributed to visual perceptual differences, and interpretation could only be based on previous literature. Finally, it is known that not only the proximity of an object, but also its speed and acceleration influence the perception of appropriate distances [19, 46, 74, 94]. Even though this article was not concerned with explicitly examining the nature of personal space, future studies should take this into account. Specific attention should be paid to conducting studies in different cultural settings and to general diversity of the sample. Only then can reasonable conclusions be drawn for appropriate distances and, ultimately, programming decisions that are representative of society as a whole, see [95, 96, 97].

Data Availability The datasets (including the R-script) generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Statements and Declarations

Ethics The experiments conducted for the present research apply to the ethical principles stated by the American Psychological Association for psychological research with human participants. Before the experiment all individuals were informed about the basic intent of the study and gave their informed consent to participate.

Competing Interests The authors have no relevant financial or non-financial interests to disclose. The authors have no competing interests to declare that are relevant to the content of this article. All authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript. The authors have no financial or proprietary interests in any material discussed in this article.

References

1. Kanda A, Arai M, Suzuki R et al (2014) Recognizing groups of visitors for a robot museum guide tour. In: 2014 7th International Conference on Human System Interactions (HSI). IEEE, pp 123–128

2. Kopka M, Krause K (2021) Can you help me? In: Schneegass S, Pfleging B, Kern D (eds) Mensch und Computer 2021. ACM, New York, NY, USA, pp 64–68

3. Transforma Insights Autonomous Robotic Systems including drones, personal assistance robots and factory machinery to see ten-fold growth in the next decade. https://transformainsights.com/news/autonomous-robotic-systems-ten-fold-growth#:~:text=In%20total%20we%20expect%20345,fold%20increase%20over%20the%20decade. Accessed 19 Sep 2022

4. Mumm J, Mutlu B (2011) Human-robot proxemics. In: Billard A, Kahn P, Adams JA (eds) Proceedings of the 6th international conference on Human-robot interaction - HRI '11. ACM Press, New York, New York, USA, p 331

5. Neggers MME, Cuijpers RH, Ruijten PAM (2018) Comfortable passing distances for Robots. In: Ge SS, Cabibihan J-J, Salichs MA et al (eds) Social Robotics, vol 11357. Springer International Publishing, Cham, pp 431–440

6. Walters ML, Koay KL, Woods SN et al (2007) Robot to Human Approaches: Preliminary Results on Comfortable Distances and Preferences. In: AAAI Spring Symposium: Multidisciplinary Collaboration for Socially Assistive Robotics, p 103

7. Walters ML, Oskoei MA, Syrdal DS et al (2011) A long-term Human-Robot Proxemic study. In: 2011 RO-MAN. IEEE

8. Koay KL, Syrdal DS, Walters ML et al (2007) Living with Robots: Investigating the Habituation Effect in Participants' Preferences During a Longitudinal Human-Robot Interaction Study. In: RO-MAN 2007 - The 16th IEEE International Symposium on Robot and Human Interactive Communication. IEEE, pp 564–569

9. Syrdal DS, Kheng Lee Koay, Walters ML et al (2007) A personalized robot companion? - The role of individual differences on spatial preferences in HRI scenarios. In: RO-MAN 2007 - The 16th IEEE International Symposium on Robot and Human Interactive Communication. IEEE, pp 1143–1148

10. Rios-Martinez J, Spalanzani A, Laugier C (2015) From Proxemics Theory to socially-aware Navigation: a Survey. Int J of Soc Robotics 7:137–153. https://doi.org/10.1007/s12369-014-0251-1

11. Neggers MME, Cuijpers RH, Ruijten PAM et al (2021) Determining shape and size of Personal Space of a human when passed by a Robot. Int J of Soc Robotics 14:561–572. https://doi.org/10.1007/s12369-021-00805-6

12. Sardar A, Joosse M, Weiss A et al (2012) Don't stand so close to me. In: Yanco H, Steinfeld A, Evers V. (eds) Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction - HRI '12. ACM Press, New York, New York, USA, p 229

13. Lauckner M (2016) Human-Robot Spatial Interaction in a Hallway. Dissertation, TU Berlin

14. Lauckner M, Kobiela F, Manzey D (2014) 'Hey robot, please step back!' - exploration of a spatial threshold of comfort for human-mechanoid spatial interaction in a hallway scenario. In: The 23rd IEEE International Symposium on Robot and Human Interactive Communication. IEEE, pp 780–787

15. Rossi S, Staffa M, Bove L et al (2017) User's personality and activity influence on HRI comfortable distances. In: Kheddar A, Yoshida E, Ge SS et al (eds) Social Robotics. ICSR 2017. Springer International Publishing, Cham, pp 167–177

16. Pacchierotti E, Christensen HI, Jensfelt P (2005) Human-robot embodied interaction in hallway settings: a pilot user study. In: ROMAN 2005. IEEE International Workshop on Robot and Human Interactive Communication, 2005. IEEE, pp 164–171

17. Pacchierotti E, Christensen HI, Jensfelt P (2006) Evaluation of Passing Distance for Social Robots. In: ROMAN 2006 - The 15th IEEE International Symposium on Robot and Human Interactive Communication. IEEE

18. Obaid M, Sandoval EB, Zlotowski J et al (2016) Stop! That is close enough. How body postures influence human-robot proximity. In: 2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN). IEEE, pp 354–361

19. Neef NE, Kastner K, Schmidt M et al (2022) On optimizing driving patterns of Autonomous Cargo Bikes as a function of Distance and Speed—A psychological study. IEEE Open J Intell Transp Syst 3:592–601. https://doi.org/10.1109/OJITS.2022.3198120

20. Yamazoe H, Chun J, Kim Y et al (2022) Analysis of impressions of robot by changing its motion and trajectory parameters for designing parameterized behaviors of home-service robots. Intel Serv Robotics. https://doi.org/10.1007/s11370-022-00447-1

21. Sommer R (1969) Personal Space. The behavioral basis of design. Prentice-Hall, Inc., Englewood Cliffs, New Jersey. ($2.75)

22. Hall ET (1966) The hidden dimension: Man's use of space in public and private. The Bodley Head., London, UK

23. Ogden B, Dautenhahn K (2000) Robotic etiquette : structured interaction in humans and robots. In: Proceedings of the 8th Symposium on Intelligent Robotic Systems, pp 353–361

24. Burgoon JK, Jones SB (1976) Toward a theory of Personal Space Expectations and their violations. Hum Commun Res 2:131–146

25. Kanaga KR, Flynn M (1981) The Relationship between Invasion of Personal space and stress. Hum Relat 34:239–248. https://doi.org/10.1177/001872678103400305

26. Khan AY, Kamal A (2010) Exploring reactions to invasion of personal space in university.Journal of Behavioural Sciences20

27. Takayama L, Pantofaru C (2009) Influences on proxemic behaviors in human-robot interaction. In: 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems. IEEE, pp 5495–5502

28. Lindner F, Eschenbach C (2011) Towards a formalization of Social Spaces for socially aware Robots. In: Springer, Berlin, Heidelberg, pp 283–303

29. Bergkvist L, Rossiter JR (2007) The predictive validity of multiple-item versus single-item measures of the same constructs. J Mark Res 44:175–184. https://doi.org/10.1509/jmkr.44.2.175

30. Gardner DG, Cummings LL, Dunham RB et al (1998) Single-item Versus multiple-item Measurement Scales: an empirical comparison. Educ Psychol Meas 58:898–915. https://doi.org/10.1177/0013164498058006003

31. Fuchs C, Diamantopoulos A (2009) Using single-item measures for construct measurement in management research: conceptual issues and application guidelines.Die Betriebswirtschaft69

32. Oshagbemi T (1999) Overall job satisfaction: how good are single versus multiple-item measures? J Managerial Psychol 14:388–403. https://doi.org/10.1108/02683949910277148

33. Allen MS, Iliescu D, Greiff S (2022) Single item measures in Psychological Science. Eur J Psychol Assess 38:1–5. https://doi.org/10.1027/1015-5759/a000699

34. Leung SO, Xu ML (2013) Single-item measures for subjective academic performance, Self-Esteem, and Socioeconomic Status. J Social Service Res 39:511–520. https://doi.org/10.1080/0148376.2013.794757

35. McIver JP, Carmines EG (1981) Unidimensional scaling. Sage university papers, no.07–024. SAGE, Beverly Hills, London

36. Döring N, Bortz J, Pöschl S et al (2015) Forschungsmethoden und evaluation in den Sozial- und Humanwissenschaften, Fifth edition. Springer-Lehrbuch. Springer, Berlin, Heidelberg

37. Li R, van Almkerk M, van Waveren S et al (2019) Comparing Human-Robot Proxemics Between Virtual Reality and the Real World. In: 2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI). IEEE, pp 431–439

38. Hasegawa Y, Dias C, Iryo-Asano M et al (2018) Modeling pedestrians' subjective danger perception toward personal mobility vehicles. Transp Res Part F: Traffic Psychol Behav 56:256–267. https://doi.org/10.1016/j.trf.2018.04.016

39. McDowell KV (1972) Violations of personal space. Can J Behav Sci / Revue canadienne des Sci du comportement 4:210–217. https://doi.org/10.1037/h0082306

40. Kastner K, Gehlmann F, Salzer S et al (2021) Determinants of the acceptability of autonomous (cargo) mobility. Transp Res Interdisciplinary Perspect 11:100448. https://doi.org/10.1016/j.trip.2021.100448

41. Mizoguchi H, Sato T, Takagi K et al (1997) Realization of Expressive Mobile Robot. In: Proceedings of International Conference on Robotics and Automation. IEEE

42. Lohse M, Hanheide M (2008) Evaluating a social home tour robot applying heuristics. Robots as Social Actors Workshop: International Symposium on Robot and Human Interactive Communication (RO-MAN 08)

43. Oestreicher L (2007) Cognitive, Social, Sociable or just Socially Acceptable Robots? In: RO-MAN 2007 - The 16th IEEE International Symposium on Robot and Human Interactive Communication. IEEE, pp 558–563

44. Torta E, Cuijpers RH, Juola JF (2013) Design of a Parametric Model of Personal Space for Robotic Social Navigation. Int J of Soc Robotics 5:357–365. https://doi.org/10.1007/s12369-013-0188-9

45. Fong T, Nourbakhsh I, Dautenhahn K (2003) A survey of socially interactive robots. Robot Auton Syst 42:143–166. https://doi.org/10.1016/S0921-8890(02)00372-X

46. Butler JT, Agah A (2001) Psychological Effects of Behavior patterns of a Mobile Personal Robot. Auton Robots 10:185–202. https://doi.org/10.1023/A:1008986004181

47. Hancock PA, Billings DR, Schaefer KE et al (2011) A meta-analysis of factors affecting trust in human-robot interaction. Hum Factors 53:517–527. https://doi.org/10.1177/0018720811417254

48. Bhagya SM, Samarakoon P, Viraj MA et al (2019) An Exploratory Study on Proxemics Preferences of Humans in Accordance with Attributes of Service Robots. In: 2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN). IEEE, pp 1–7

49. Walters ML, Dautenhahn K, te Boekhorst R et al (2009)An empirical framework for human-robot proxemics

50. Mori M, MacDorman K, Kageki N (2012) The Uncanny Valley [From the field]. IEEE Robot Automat Mag 19:98–100. https://doi.org/10.1109/mra.2012.2192811

51. Walters ML, Syrdal DS, Dautenhahn K et al (2008) Avoiding the uncanny valley: robot appearance, personality and consistency of behavior in an attention-seeking home scenario for a robot companion. Auton Robot 24:159–178. https://doi.org/10.1007/s10514-007-9058-3

52. Samarakoon SMBP, Muthugala MAVJ, Jayasekara AGBP (2022) A review on Human–Robot Proxemics. Electronics 11:2490. https://doi.org/10.3390/electronics11162490

53. Teichtahl AJ, Wluka AE, Strauss BJ et al (2012) The associations between body and knee height measurements and knee joint structure in an asymptomatic cohort. BMC Musculoskelet Disord 13:19. https://doi.org/10.1186/1471-2474-13-19

54. Huettenrauch H, Eklundh K, Green A et al (2006) Investigating Spatial Relationships in Human-Robot Interaction. In: 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems. IEEE, pp 5052–5059

55. GoBe R (2022) Beam is moving to GoBe Robots. https://gobe.blue-ocean-robotics.com/beam-to-gobe. Accessed 15 Sep 2022

56. Nanda S, Warms RL (2011) Cultural anthropology, 10th edn. Wadsworth Cengage, Belmont, CA

57. Bartneck C, Kulić D, Croft E et al (2009) Measurement Instruments for the Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety of Robots. Int J of Soc Robotics 1:71–81. https://doi.org/10.1007/s12369-008-0001-3

58. Fox J, Weisberg S (2019) An {R} companion to Applied Regression, 3rd edn. SAGE, Thousand Oaks, CA, USA

59. Greenwood M, Hancock S, Carnegie N (2022) catstats: Statistics for Montana State University

60. Wei T, Simko V (2021) Corrplot. Visualization of a Correlation Matrix

61. Lawrence MA (2016) ez: Easy Analysis and Visualization of Factorial Experiments

62. Kassambra A (2020) ggpubr: 'ggplot2'. Based Publication Ready Plots

63. Harrell F Jr (2022) Hmisc: Harrell Miscellaneous

64. Firke S (2021) janitor: Simple Tools for Examining and Cleaning Dirty Data

65. Schauberger P, Walker A (2021) Openxlsx: read. Write and Edit xlsx Files

66. Neuwirth E (2022) RColorBrewer:ColorBrewer Palettes

67. Csárdi G, Hester J, Wickham H et al (2021) remotes: Installation from Remote Repositories, Including 'GitHub'

68. Kassambra A (2021) rstatix: Pipe-Friendly Framework for Basic Statistical Tests

69. Pfister R, Janczyk M (2020) schoRsch:Tools for Analyzing Factorial Experiments

70. Wickham H, Averick M, Bryan J et al (2019) Welcome to the Tidyverse. JOSS 4:1686. https://doi.org/10.21105/joss.01686

71. Benjamini Y, Hochberg Y (1995) Controlling the false Discovery rate: a practical and powerful Approach to multiple testing. J Royal Stat Soc Ser B (Methodological) 57:289–300

72. Greenhouse SW, Geisser S (1959) On methods in the analysis of profile data. Psychometrika 24:95–112

73. Field A, Miles J, Field Z (2013) Discovering statistics using R. SAGE, Los Angeles

74. Neggers MME, Cuijpers RH, Ruijten PAM et al (2022) The effect of robot speed on comfortable passing distances. Front Robot AI 9:915972. https://doi.org/10.3389/frobt.2022.915972

75. Krause K, Assmann T, Schmidt S et al (2020) Autonomous driving cargo bikes – introducing an acceptability-focused approach towards a new mobility offer. Transp Res Interdisciplinary Perspect 6:100135. https://doi.org/10.1016/j.trip.2020.100135

76. Weiss A, Bartneck C (2015) Meta analysis of the usage of the Godspeed Questionnaire Series. In: 2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN). IEEE, pp 381–388

77. Schrum ML, Ghuy M, Hedlund-Botti E et al (2022) Concerning Trends in Likert Scale usage in Human-Robot Interaction: towards improving best Practices. https://doi.org/10.1145/3572784. J Hum.-Robot Interact

78. Schwarz N, Knauper B, Hippler H-J et al (1991) Rating scales: numeric values may change the meaning of Scale Labels. Pub Opin Q 55:570. https://doi.org/10.1086/269282

79. Gu H, Wen Z, Fan X (2015) The impact of wording effect on reliability and validity of the core self-evaluation scale (CSES): a bi-factor perspective. Pers Indiv Differ 83:142–147. https://doi.org/10.1016/j.paid.2015.04.006

80. Steinberg L, Rogers A (2022) Changing the Scale: the Effect of modifying response scale labels on the measurement of personality and affect. Multivar Behav Res 57:79–93. https://doi.org/10.1080/00273171.2020.1807305

81. Weng L-J (2004) Impact of the number of response categories and Anchor Labels on Coefficient Alpha and Test-Retest Reliability. Educ Psychol Meas 64:956–972. https://doi.org/10.1177/0013164404268674

82. Otto S, Kröhne U, Richter D (2018) The dominance of introspective measures and what this implies: the example of environmental attitude. PLoS ONE 13:e0192907. https://doi.org/10.1371/journal.pone.0192907

83. Podsakoff PM, MacKenzie SB, Lee J-Y et al (2003) Common method biases in behavioral research: a critical review of the literature and recommended remedies. J Appl Psychol 88:879–903. https://doi.org/10.1037/0021-9010.88.5.879

84. Hertzum M, Borlund P, Kristoffersen KB (2015) What do thinking-aloud participants say? A comparison of Moderated and Unmoderated Usability Sessions. Int J Hum Comput Interact 31:557–570. https://doi.org/10.1080/10447318.2015.1065691

85. Jääskeläinen R (2010) Think-aloud protocol. In: Gambier Y, van Doorslaer L (eds) Handbook of translation studies, vol 1. John Benjamins Publishing Company, Amsterdam, pp 371–373

86. Weiss A, Bernhaupt R, Tscheligi M et al (2009) Addressing User Experience and Societal Impact in a User Study with a Humanoid Robot. In: Adaptive and Emergent Behaviour and Complex Systems - Proceedings of the 23rd Convention of the Society for the Study of Artificial Intelligence and Simulation of Behaviour, pp 150–157

87. Weiss A, Igelsbock J, Pierro P et al (2010) User perception of usability aspects in indirect HRI - a chain of translations. In: 19th International Symposium in Robot and Human Interactive Communication. IEEE, pp 539–545

88. Llobera J, Spanlang B, Ruffini G et al (2010) Proxemics with multiple dynamic characters in an immersive virtual environment. ACM Trans Appl Percept 8:1–12. https://doi.org/10.1145/1857893.1857896

89. Weistroffer V, Paljic A, Fuchs P et al (2014) Assessing the acceptability of human-robot co-presence on assembly lines: A comparison between actual situations and their virtual reality counterparts. In: The 23rd IEEE International Symposium on Robot and Human Interactive Communication. IEEE, pp 377–384

90. Lichtenthäler C, Lorenzy T, Kirsch A (2012) Influence of legibility on perceived safety in a virtual human-robot path crossing task. In: 2012 IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication. IEEE, pp 676–681

91. Hayduk LA (1983) Personal space: where we now stand. Psychol Bull 94:293–335. https://doi.org/10.1037/0033-2909.94.2.293

92. Beaulieu C (2004) Intercultural study of Personal Space: a Case Study. J Appl Social Pyschol 34:794–805. https://doi.org/10.1111/j.1559-1816.2004.tb02571.x

93. Eresha G, Haring M, Endrass B et al (2013) Investigating the influence of culture on proxemic behaviors for humanoid robots. In: Duncan BA, Murphy RR (eds) Comfortable Approach Distance with small Unmanned Aerial Vehicles. IEEE, Piscataway, NJ, pp 430–435

94. Peters A, Spexard TP, Hanheide M et al (2011) Hey robot, get out of my way - A survey on a spatial and situational movement concept in HRI. In: Gottfried B, Aghajan H (eds) Behaviour Monitoring and Interpretation – BMI, pp 147–165

95. Kim S (2022) Equitable science for all: Q&A with Dr. Flavio Azevedo on open science practices. https://science.nasa.gov/open-science/transform-to-open-science/stories/dr-flavio-azevedo

96. Zabel S, Otto S (2021) Bias in, bias out – The similarity-attraction effect between chatbot designers and users. Lecture Notes in Computer Science. International Conference on Human-Computer Interaction.:184–197. https://doi.org/10.1007/978-3-030-78468-3_13

97. zabel, S., Schlaile, M. P., – Otto, S. (2023). Breaking the chain with individual gain? Investigating the moral intensity of COVID-19 digital contact tracing. Computers in Human Behavior, 143, 107699.https://doi.org/10.1016/j.chb.2023.107699

Nicolas E. Neef is a PhD student at the University of Hohenheim. His research interests are in human-robot interaction and the use of motive-driven algorithms in the ecological domain. He currently is working to pursue an academic career.

Sarah Zabel is a PhD student at the University of Hohenheim. Her Research focuses on human-technology interaction, with a special emphasis on the aspects of social and ecological sustainability.

Mathis Lauckner is a freelance UX and design researcher, based in San Francisco. His research intrests include the development of intellient robotic systems, the exploration of socially acceptable robotic behavior and robot proxemics in general.

Siegmar Otto is a professor at the University of Hohenheim. He researches the digital transformation from a sustainable angle. His research is highly empirical and aims to uncover biases in digital decision systems and on their interfaces, but also to find ways to utilize digital change in order to enable sustainable development. This incorporates research on the human-machine interface, how to shape the outcome of algorithms and how they in turn affect human cognition and behavior.