



Article

# When Robots Fail—A VR Investigation on Caregivers' Tolerance towards Communication and Processing Failures

Kim Klüber \*  and Linda Onnasch 

Faculty of Psychology, Department of Engineering Psychology, Humboldt-Universität zu Berlin, 10099 Berlin, Germany

\* Correspondence: kim.klueber@hu-berlin.de

**Abstract:** Robots are increasingly used in healthcare to support caregivers in their daily work routines. To ensure an effortless and easy interaction between caregivers and robots, communication via natural language is expected from robots. However, robotic speech bears a large potential for technical failures, which includes processing and communication failures. It is therefore necessary to investigate how caregivers perceive and respond to robots with erroneous communication. We recruited thirty caregivers, who interacted in a virtual reality setting with a robot. It was investigated whether different kinds of failures are more likely to be forgiven with technical or human-like justifications. Furthermore, we determined how tolerant caregivers are with a robot constantly returning a process failure and whether this depends on the robot's response pattern (constant vs. variable). Participants showed the same forgiveness towards the two justifications. However, females liked the human-like justification more and males liked the technical justification more. Providing justifications with any reasonable content seems sufficient to achieve positive effects. Robots with a constant response pattern were liked more, although both patterns achieved the same tolerance threshold from caregivers, which was around seven failed requests. Due to the experimental setup, the tolerance for communication failures was probably increased and should be adjusted in real-life situations.

**Keywords:** anthropomorphic communication; communication failure; failure tolerance; VR study; caregiver; human–robot interaction



**Citation:** Klüber, K.; Onnasch, L. When Robots Fail—A VR Investigation on Caregivers' Tolerance towards Communication and Processing Failures. *Robotics* **2022**, *11*, 106. <https://doi.org/10.3390/robotics11050106>

Academic Editor: Thierry Chaminade

Received: 31 August 2022

Accepted: 6 October 2022

Published: 7 October 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The current global shortage of healthcare professionals [1], which is expected to increase in the next years [2], is countered by the expanded use of technology and robotics. Conventional social robots support caregivers in healthcare facilities by performing cognitive and emotional stimulating tasks in interactions with patients. The social robot *Paro*, for example, is a great help when dealing with patients with dementia [3,4], and the human-like robot *Pepper* when entertaining patients [5]. Additionally, service robots support caregivers in functional tasks [6,7]. According to the International Standardization Organization, a service robot is defined as a robot “that performs useful tasks for humans or equipment, excluding industrial automation applications” [8]. Similar to industrial robots, service robots support humans in physically demanding tasks and therefore have a great potential for the healthcare sector as there are still not enough support options for caregivers to compensate and reduce the serious health consequences they face. A few example tasks that such robots can be used for in healthcare are disinfection, logistics, monitoring, and moving patients (e.g., patient positioning) [7]. In contrast to social robots, service robots' primary interaction partners are caregivers, who hand over specific tasks, load the robot, or reposition the patient together with the robot, the latter requiring a great deal of coordination. To ensure a successful interaction with the caregivers, no additional (cognitive) demand should be placed on the care personnel, but certain requirements are

placed on the robot. One of the most important, addressing the interaction itself, is that the robot should adopt familiar social communication scripts grounded on human–human dialogue strategies to simplify the human–robot interaction (HRI) for caregivers [9,10]. For this reason, a verbal communication seems to be most appropriate. Verbal communication, and more precisely, oral communication (speech), is a natural tool for humans, which allows us to exchange information quickly and effortlessly at the same time. In robotic systems, however, understanding and producing speech is very prone to failures. Robots working in care facilities have to interact and communicate with very different users (e.g., caregivers, patients, visitors), which requires the robot to constantly adapt. This opens up the possibility for failures that cannot be completely counteracted in advance. In particular, the interaction with humans makes it impossible to identify all possible types of robotic failures beforehand and complicates an error-free task execution [11]. Since robots are expected to make mistakes, it is important to address how caregivers respond to faulty robots and what is needed to achieve greater tolerance for failures.

This work aims to answer how tolerant caregivers are with failure-prone robots and how communication failures influence caregivers' behavior and perception towards them. In the following, the advantages of an anthropomorphic communication (speech) are given. Next, the influence of failures during communication and interactions with robots is presented. For this, we follow the taxonomy of Honig and Oron-Gilad [11]. Since the study was conducted as a virtual reality (VR) experiment, we provide a detailed examination of this research method before presenting our research question and hypothesis.

## 2. Related Work

### 2.1. Anthropomorphic Communication

An anthropomorphic communication, usually referring to a verbal, spoken communication, represents a simple way for caregivers to interact with robots. Robots produce speech by text-to-speech systems and even convey emotions by further including prosody in the speech production [12]. According to the media equation theory [9], technological devices with anthropomorphic features should automatically trigger already familiar interaction schemes. An anthropomorphic communication thereby enables caregivers to mindlessly recall familiar social scripts and transfer them to the interaction with the robot. This in turn makes the interaction with the robot more intuitive. In addition, robotic verbal communication is one of the most effective features when considering the positive influence of an anthropomorphic design such as an increase in likeability and trust [13]. Since service robots are often restricted in their appearance and movement by their function, the implementation of an anthropomorphic communication is also the easiest way to include anthropomorphic features into service robots. Interacting with spoken language has even more advantages [10–12]. A few reasons are the fast and most efficient exchange of information by speech [14], the real-time coordination of physical actions [14], the social potential of spoken language [15], and that speech is the most preferred communication channel by caregivers compared to communicating via sound or text [16]. Furthermore, people expect the robot to speak as robots become more social and capable [14]. All these advantages support the implementation of a verbal spoken communication by robots in the healthcare setting.

### 2.2. Robotic Failures

In HRI research, the term “failure” refers to “a degraded state of ability which causes the behavior or service being performed by the system to deviate from the ideal, normal, or correct functionality” [17] (p. 9). This definition includes both the actual and the subjectively perceived failure [11]. Honig and Oron-Gilad have developed a taxonomy to structure human–robot failures [11]. According to their taxonomy, failures can be divided into technical and interaction failures. Whereas interaction failures include problems that are caused by humans, social norms, or the environment, technical failures primarily include problems that are caused by the robot. When adapting robots for use in care facilities,

adjustments and countermeasures should be implemented on the robotic device's side, and it is necessary to focus on technical failures in particular. A main component of technical failures is software failures, which are further divided into design, communication, and processing failures.

Software failures are especially important in the verbal interaction with the user and affect how the robot is perceived and evaluated by humans. Processing failures reduce, for example, the perceived reliability, trustworthiness, understandability, and competence of robots [11,18]. Salem et al. showed that processing failures that led to a wrong robot behavior significantly decreased the robot's trustworthiness [18]. Beyond that, failures furthermore influence the behavior of users. In terms of communication failures, unexpected answers from a voice assistant, for example, cause users to adjust their responses by speaking louder, more clearly, rephrase the question, or repeat the question with small modifications to vocabulary or grammar [19,20].

Mavrina and colleagues conducted a long-term study with five families on the use of a voice assistant [21]. The number of requests made by the families was assessed and divided by successful and failed requests. Furthermore, the satisfaction with the voice assistant was queried. The authors found that satisfaction with the voice assistant was significantly lower the higher the number of abandoned, failed requests was. However, satisfaction was only surveyed once after the study. Thus, it cannot be concluded from the results whether successful requests improved satisfaction after failed requests occurred or whether the timing of failed requests affected the level of satisfaction. In addition to a reduced satisfaction, failed interactions negatively affect the frequency of use [22]. However, this seems to be modulated by the technical savvy of users, as the study by Luger and Sellen showed that technically experienced users were more tolerant of communication failures and aborted interactions with voice assistants after a greater number of attempts compared to less technically savvy users [22]. The interviews by Luger and Sellen were, however, conducted with only 14 participants, who additionally used different voice assistants. This poses the question of generalizability of results.

To minimize such failure consequences, it is important to examine different recovery strategies that can be applied after an occurred failure. Kim et al. have investigated whether apologies are suitable as a recovery strategy [23]. More specifically, they examined whether trust rehabilitation differs when failures are attributed either to internal (full responsibility lies with the individual) or external causes (responsibility also lies with other persons). They found that internal attributions rehabilitated trust better than external attributions. However, the study was not conducted in the HRI domain. Instead, participants watched videos of job applicants who were accused of incorrectly filing a tax return and whose hiring was to be decided. It is thus unclear whether the results also apply to communication with robots.

In addition to apologies, various recovery strategies, such as ignoring, blaming, justifying/explanation, etc., have already been examined by researchers [24] within the field of HRI [24–26]. Choi and colleagues compared apologies with explanations given by a robot after a service failure [25]. The authors showed that both strategies had positive effects on recovery. This effect was, however, only present for humanoid robots and not for non-humanoid ones. Choi et al. concluded that the observed difference for different types of robots was due to a lack of social capabilities by non-humanoid robots. To be successful as a recovery strategy, other parameters are important. The purpose of an explanation is to reveal the reason or cause for a failure [25]. The effectiveness of an explanation, for example, is driven by perceived adequacy and the truthfulness of information [26].

### *2.3. Conducting HRI Research in VR*

In recent years, VR has become a popular tool for conducting HRI user studies [27,28]. VR offers an alternative to provide visual cues that are similar to the real world and creates realistic and immersive environments. Badia and colleagues stated that VR systems that elicit a realistic feeling and appear to be plausible can even create the same behavioral and

psychophysiological responses as a real-world interaction [29]. VR has several advantages, but also raises new challenges [30]. Human safety, for example, is crucial when interacting with robots [29]. VR can be used to explore new forms of interactions, as it provides a safe tool for testing HRI without jeopardizing the safety of humans. Furthermore, VR allows the testing of multiple virtual robots with different designs in various environments. This does not have to be limited to existing robot systems, as hypothetical robot appearances and behaviors can be implemented as well [29]. Overall, VR provides a less resource-consuming tool (i.e., time and cost) compared to studies with real robots [28].

When conducting a VR study, the main concern is whether participants respond realistically or whether they are influenced by the virtual nature of the study. It is therefore necessary to control if the interaction evokes a high level of presence (actually being in the environment) [31]. In addition to the environment, the presentation of the robot influences the perception and evaluation of robots and the effects on humans [13]. Badia and colleagues have identified variables that can be manipulated and measured in a VR experiment [29]. They concluded that HRI studies in VR offer the assessment of subjective and objective metrics, thereby providing comparable options as real experiments. With regard to the manipulating variables, a distinction was made on three categories: collaborative robot (cobot), environment, and user. In the present study, the variation of the robot (equivalent to cobot) is most relevant. A property mentioned by Badia et al. that can be manipulated on the robot's side is the degree of anthropomorphism [29]. A meta-analysis by Roesler and colleagues examined the influence of anthropomorphism in social HRI [32]. They analyzed embodied and depicted robots separately, with virtual robots belonging to the latter group. Human-related outcomes such as robot perception (subjective measure) or behavior (objective measure) were considered as dependent variables. They found that anthropomorphism investigated via physically embodied robots positively influenced subjective and objective measures whereas depicted robots failed to show a positive effect on the objective outcomes. However, subjective outcomes such as perception and attitude showed a consistent positive effect of anthropomorphism using depicted robots. This suggests that behavioral data especially are more difficult to capture without real robots.

Further empirical results on the comparability between VR and lab-based physically embodied HRI studies provide mixed results. Weistroffer and colleagues, for example, studied the co-presence of humans and robots and found no differences in questionnaire answers between real and virtual situations [33]. The study was conducted within an industrial setting, in which participants had to work side-by-side with the robot on a car door. In contrast, Li and colleagues found differences in proxemics showing that participants preferred a closer interaction with real robots instead of virtual robots [31]. For their user studies, the social robot Pepper was used, once with the real Pepper and once with its virtual counterpart. The authors suggested that one reason for the greater distance in VR was because the virtual robot was perceived as more discomforting compared to the real Pepper. To achieve the same results between a virtual scenario and laboratory setting the basic requirements should not differ.

It can be assumed that the type of robot exposure in studies influences the observed human-related outcome variables. Although this does not apply to all outcomes, it should be considered when generalizing findings. Furthermore, the discrepancies between results indicate that certain control variables (e.g., immersion) should be gathered to formulate statements for transferability. Overall, advantages such as the ecological benefits and safety aspects show that VR is a valid tool to obtain initial results related to HRI.

### 3. Research Questions and Hypotheses

Based on the presented literature, it was shown that VR is a less resource-consuming and less risky research tool for conducting HRI user studies compared to studies with real robots [28]. It is therefore a valid tool to investigate HRI-related questions. Although most studies either investigate HRI in an industrial context [28,31] or with social robots [3,4,6],

the use of service robots in healthcare is a new field that is just starting to be increasingly researched [7]. With our study, we aimed to address this research gap. Moreover, service robots in the healthcare sector provide great assistance in functional tasks (e.g., cleaning, transportation tasks) [7]. As a result, caregivers become the primary interaction partners, compared to social robots having the patients as primary interaction partners. Previous studies, however, have included students as participants [23,24] or a random sample [25,26]. A major benefit of our study is the inclusion of caregivers. This allows us to derive implications relevant to this specific target group. In studies with caregivers, previous research revealed that robots communicating via speech are beneficial for a successful interaction [14–16]. However, robotic verbal communication is prone to failures in terms of processing speech input from various users in unstructured environments and providing accordingly appropriate answers and actions [11]. Hence, it is necessary to consider failure consequences (e.g., how caregivers respond to communication failures of robots) and possible countermeasures (e.g., recovery strategies) to ensure the long-term use of robots [11,24]. Accordingly, we investigated which type of explanation is more suitable in care settings for justifying failures. According to the failure taxonomy of Honig and Oron-Gilad, the narrated failures of the robot in our study belonged to processing and communication failures [11]. In our study, the robot was equipped with a face and thus human-like characteristics. It could therefore be assumed that recovery strategies did not fail due to a lack of social capabilities [25]. We expected that justifications for processing and communication failures should have a positive effect. We assumed, on the one hand, that explanations based on human-like properties are more understandable and comprehensible, because humans can apply these explanations to themselves and identify with them [9,34]. On the other hand, explanations that involve technical terms can create a more realistic impression of failures caused by the robot, which is perceived as more truthful [26]. Our exploratory research question was therefore:

**R1.** *Which failure justification has a more positive impact on the evaluation of robots by caregivers?*

Since a failed interaction reduces the frequency of use [22], we were also interested in how tolerant caregivers are towards a robot that fails in communication. The failed interaction was caused by the robot not being able to process the speech input (processing failure) [11]. How long would caregivers try to interact with the robot? What was their tolerance threshold?

**R2.** *What is the tolerance threshold for caregivers to repeat voice prompts to a robot?*

Studies have already shown that people adjust their response pattern in case of a failed interaction [19,20]. We assumed that a robot that gives concrete suggestions for an adaptation would be evaluated better than a robot that always answered the same way. We therefore hypothesized that a greater variance in responses from the robot would lead to a better evaluation and a greater tolerance among caregivers for robot failures and thus higher repetition rates.

**H1.** *A variable response pattern leads to more repetitions by the caregivers (higher error tolerance) and a better evaluation of the robot as a constant response pattern.*

#### 4. Materials and Methods

The study was preregistered at the Open Science Framework (OSF) where the raw data of the study are available. All subjects gave their informed consent for inclusion before they participated in the study. The study was conducted in accordance with the Declaration of Helsinki, and the protocol was approved by the Ethics Committee of the Humboldt-Universität zu Berlin (2022-09). The study consists of six parts in total (1. front selection, 2. design selection, 3. proxemics, 4. failure justification, 5. error tolerance, 6. interview). This paper focuses only on parts addressing robot communication failures (4. and 5.).

#### 4.1. Participants

Various care facilities within the area of Berlin were contacted via e-mail to recruit participants. A prerequisite for participation was employment as a nursing/care specialist, nursing/care assistant, everyday helper, service worker, or therapist in inpatient care. Further prerequisites were not suffering from impaired gait or clinical balance disorders, legal age, and meeting the requirements of the Coronavirus regulation (recovered from COVID-19 or fully vaccinated with an additional negative test result). We recruited 30 participants who worked in one of the aforementioned professions. The average work experience of the participants was 17 years ( $SD = 10$ ; ranging from 2 to 39 years). The mean age of the participants was 40 years ( $SD = 9$  years), ranging from 24 to 55 years, and the majority of the participants were female ( $N_{female} = 21$ ;  $N_{male} = 9$ ). Only two participants stated having previous experience with robots, but not with care robots. For taking part in the study, participants were financially compensated with EUR 100.

#### 4.2. Design

The study comprised two subsequent tasks that participants performed within the VR environment. The first was the failure justification task. In this part of the study, we used a within-subject design. During the task, the robot justified its failures with a human and a technical reason, respectively.

The second task was the error tolerance task and was implemented as a between-subject design. The robot asked the participants to repeat a previously posed question, either always with the same request (constant response pattern) or with a slightly rephrased request (variable response pattern).

#### 4.3. Materials and Measures

The study was conducted as a VR experiment, created with Unreal Engine 4.7. The VR environment resembled a kitchen in a care facility (see Figure 1). For tasks in which the robot had to communicate, audios were recorded upfront with the Amazon Polly (<https://aws.amazon.com/de/polly/>; accessed on 22 February 2022) Natural Text-To-Speech (NTTS) software. The full script of the audios is available at the OSF.

For the failure justification task, two audios were recorded upfront. In the audios, first the robot introduced itself, then described its tasks and functions, and lastly described a situation of a failed interaction. In this interaction, a patient had a request but the robot made some mistakes (e.g., did not find the patient's room again). In the human-like condition, the robot justified its mistakes with the fact that it was new in the facility and had difficulties remembering routes. In the technical condition, the justification was based on a not fully calibrated map of the facility. The exact scripts are presented in Table 1.

For the error tolerance task, five audios were recorded upfront. When speaking with a constant response pattern, the robot always said "Excuse me, I didn't understand you. Could you please repeat that?". In the variable response pattern, instead of "Could you please repeat that?", the robot said "Could you please speak more slowly/loudly/clearly?" or "Could you please rephrase that?". All audios are included as supplementary materials.

To determine how caregivers evaluated the robotic communication, they were asked to rate their attitude towards the use of the robot [35], their failure forgiveness towards the robot (adapted from [36]), how reliable they perceived the robot [37], and how much they liked the robot (Godspeed III; [38]). Except for likeability (semantic differential), we measured all items on a 5-point Likert scale anchored from 1 (totally disagree) to 5 (fully agree). A customized item was added to determine to whom the caregivers attributed the failed interaction in the second task. The selection options were the robot, themselves, or both. The Negative Attitudes towards Robots Scale (NARS; [39]) and the Igroup Presence Questionnaire (IPQ; [40]) were further collected on a scale of 1-5 to control for factors that might influence the results. The IPQ is divided into four subgroups: the spatial presence, which measures the sense of being physically present in the VR; involvement, which measures the attention devoted to VR; experienced realism, which measures the subjective

experience of realism in the VR; and general presence, which assesses the general “sense of being there”. The original items were presented in German. A detailed description of the questionnaires in the failure justification task can be found in the supplementary materials section.



Figure 1. VR environment and virtual robot. The yellow x shows the position of the participants.

Table 1. Script for the failure justification task.

Condition	Script
technical	<p>Hello, my name is <b>Kali</b> and I am the new robot on the station since 5 days. My task is to bring support and relief to your everyday care. One example is the use as calling system. Requests are recorded and forwarded to you or carried out independently.</p> <p>Three days ago, the following errors happened during task execution:</p> <p>A patient had asked for <b>sausage</b>, so route navigation to the kitchen was started. Since my system was still incompletely calibrated for localization in the station, the route back to the patient could not be calculated. Full calibration was not completed for 96 h.</p> <p>The current localization status is finalized, and a complete map of the station is saved.</p> <p>The order <b>sausage</b> was also incorrect because the speech recognition system had categorized the word as <b>thirst</b>. As a consequence, <b>a bottle of water</b> was taken from the kitchen. My speech processing system is still error prone with some words. Software updates continue to improve my system.</p>
human-like	<p>Good day, I am <b>Ali</b> the new robot in the facility since one week. I try to support and relieve you in your daily work. For example, you can use me as calling system.</p> <p>Thereby I take requests and execute them independently or forward them to you.</p> <p>Recently, the following mishaps unfortunately happened to me:</p> <p>A patient had asked me for a piece of <b>bacon</b>, so I went to the kitchen. However, since I have such a hard time remembering directions, I got lost on the way back to the patient.</p> <p>It took me a few more days to find my way around the facility. In the meantime, I already know my way around.</p> <p>By the way, I didn't have any <b>bacon</b> with me then either, but <b>a piece of pie</b>. Instead of <b>bacon</b>, I heard <b>pastry</b>. Due to the many new impressions at the beginning, I was mentally distracted and had probably misunderstood. However, I'm always trying to improve</p>

Note. Original script was recorded in German. The German words sausage and thirst (Wurst and Durst) and bacon and pastry (Speck and Gebäck) rhyme. The words in bold were swapped between conditions and represent the two versions of the scripts to avoid hearing the same story twice. The versions were balanced across the participants.

#### 4.4. Procedure

Prior to study participation, caregivers who met the study prerequisites were sent the informed consent form, which could either be returned by mail or brought to the study appointment, and a questionnaire, in which demographic data and the NARS were collected. In the informed consent, participants were informed about the procedure and the study's purpose. However, they were not informed that their tolerance towards the robot was assessed, to avoid influencing their behavior by that information. At the study appointment, participants were again informed about their rights and risks before putting on the VR equipment. The participants performed five tasks in the VR environment, followed by an interview. The present paper only describes the two VR tasks, the failure justification task and error tolerance task, in their exact procedure. In the failure justification part, the robot stood in front of the participants and justified failures either with human-like or with technical reasons. The order of justification type was balanced between participants. After each failure justification, the attitude towards using the robot, the failure forgiveness, the reliability, and the likeability of the robot were questioned. After this, the error tolerance task followed. Participants were instructed to ask the robot for the current time. After participants had asked the question, the experimenter pressed a button on the keypad so that the recorded audio played and it seemed as if the robot had answered to the question. Participants in the constant condition listened always to the same request to repeat the question. In the variable condition, the different audios were played in random order. If the participants did not stop the interaction themselves at some point, it was stopped after 15 repetitions. The error tolerance was therefore measured by the number of repetitions. After the failed interaction with the robot, the likeability questionnaire was surveyed again and a customized item was added to determine to whom participants attributed the failed interaction. At the very end, participants were asked to answer the IPQ questionnaire, and then the VR glasses could be taken off. The time spent in the VR was 45 min on average (all five VR tasks).

#### 4.5. Statistical Analysis

Mean ( $M$ ) and standard deviation ( $SD$ ) were calculated for all collected variables. For normally distributed data,  $t$ -tests were calculated; otherwise, the Wilcoxon signed rank test was applied. The significance level was set to  $p < 0.05$ . For analyses including more than one factor, mixed analyses of variance (ANOVAs) were calculated. For tests with categorical variables, the Chi-Square test of independence ( $\chi^2$ ) was used.

### 5. Results

#### 5.1. Control Variables

Overall, the participating caregivers showed a medium negative attitude towards robots ( $M = 2.7$ ,  $SD = 0.6$ ), which did not differ between gender ( $M_{males} = 2.7$ ,  $SD_{males} = 0.7$ ;  $M_{females} = 2.7$ ,  $SD_{females} = 0.6$ ;  $t(28) = 0.213$ ,  $p = 0.833$ ) nor experimental group ( $M_{variable} = 2.9$ ,  $SD_{variable} = 0.6$ ;  $M_{constant} = 2.5$ ,  $SD_{constant} = 0.6$ ;  $t(28) = 1.612$ ,  $p = 0.118$ ). According to the IPQ, the spatial presence was rated high with  $M = 4.5$  ( $SD = 0.8$ ) as well as the general presence with  $M = 4.5$  ( $SD = 0.8$ ). The involvement ( $M = 3.1$ ,  $SD = 1.0$ ) and the experienced realism ( $M = 3.6$ ,  $SD = 0.7$ ) were rated on a medium level. All in all, participants experienced a strong sense of presence in the VR, which did not differ between gender (all  $p > 0.05$ ). When checking for group difference in the error tolerance task, we found that the group with the constant response pattern gave significant higher ratings in terms of experienced realism ( $M = 4.0$ ,  $SD = 0.7$ ) than the group with the variable response pattern ( $M = 3.3$ ,  $SD = 0.7$ ;  $t(28) = -2.491$ ,  $p = 0.019$ ). No differences were found for the other subgroups (all  $p > 0.05$ ).

#### 5.2. Failure Justification

The results of the failure justification tasks can be seen in Table 2. Overall, we found no significant difference between the technical and the human-like failure justifications for any surveyed questions (Wilcoxon signed rank test, all  $p > 0.05$ ). A descriptive examination of



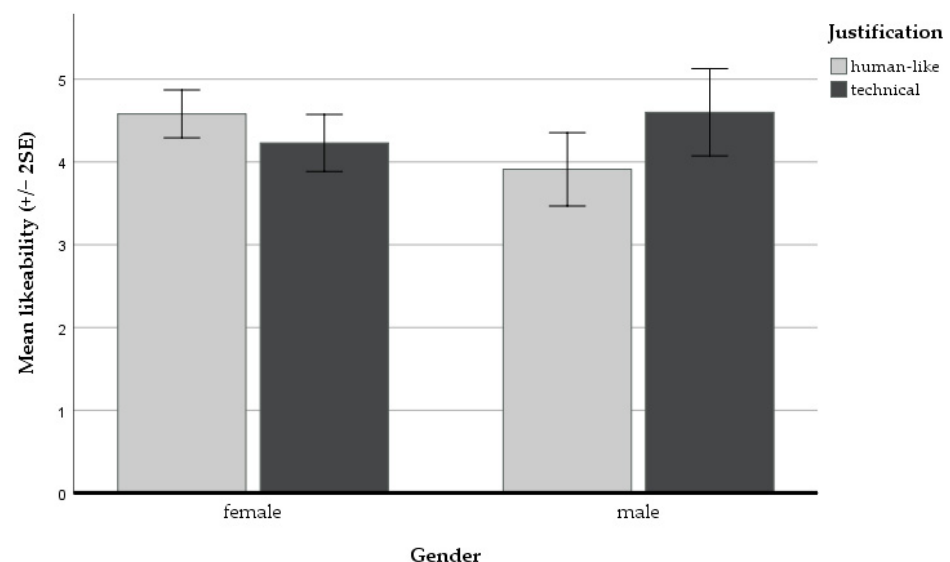
the results, including gender, revealed differences. Females rated the human-like justification higher on all variables than the technical failure justification. For males, the opposite pattern appeared. They rated the technical justification higher than the human-like condition.

**Table 2.** Mean and standard deviation in brackets for the two failure justifications divided by the participants' gender.

Factors	Gender	Failure Justification	
		Technical	Human
attitude to use	female	3.68 (0.91)	3.71 (0.94)
	male	4.30 (0.87)	3.67 (1.32)
failure forgiveness	female	3.79 (0.98)	3.86 (1.11)
	male	4.03 (1.21)	3.75 (1.35)
reliability	female	3.23 (0.83)	3.31 (0.95)
	male	3.39 (1.02)	3.23 (1.09)
likeability	female	4.23 (0.79)	4.58 (0.44)
	male	4.60 (0.78)	3.91 (1.03)

Note. Female:  $N = 21$ , male:  $N = 9$ .

To test statistically for gender effects, a mixed ANOVA was calculated. With regard to reliability, forgiveness, and attitude, no main effect nor interaction effects were found. With regard to the attitude towards using the robot, the interaction just missed the conventional level of significance ( $F(1,28) = 4.021$ ,  $p = 0.055$ ,  $\eta^2 = 0.126$ ). For the likeability ratings, a significant interaction was found ( $F(1,28) = 9.266$ ,  $p = 0.005$ ,  $\eta^2 = 0.249$ ). Females liked robots with human-like justifications more; males liked robots with technical justifications more (see Figure 2).

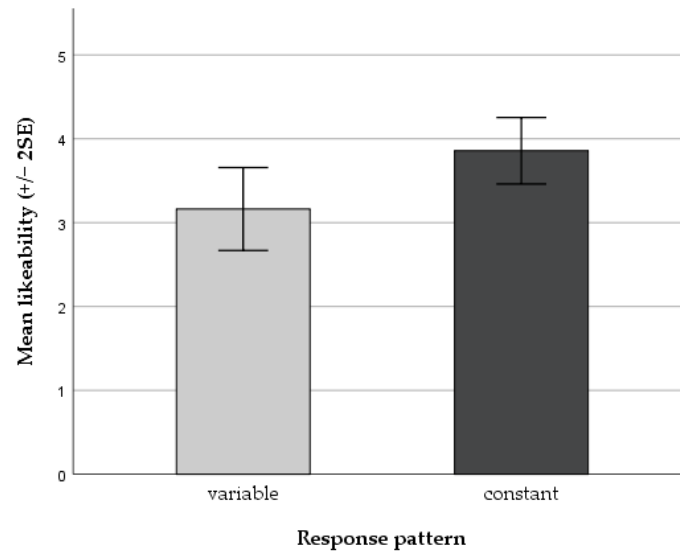


**Figure 2.** Likability results for the two failure justifications divided by the participants' gender.

### 5.3. Error Tolerance

Overall, 19 participants stopped the interaction with the robot by saying something similar to "For how long should I continue doing that?". This stop criteria was labeled as self-determination. Participants who stopped with self-determination repeated the question on average seven times ( $SD = 3$ ) and rated the likeability of the robot as  $M = 3.5$  ( $SD = 0.9$ ). The remaining eleven participants continued until the experimenter stopped the interaction after participants had repeated the question 15 times. In this group, participants rated the robot's likeability as  $M = 3.4$  ( $SD = 1.1$ ). No significant difference on likeability was found between

the different stop criteria ( $t(28) = -0.220, p = 0.827$ ) nor between gender ( $t(28) = 0.327, p = 0.746$ ). However, a significant difference between the response patterns was found ( $t(28) = 2.151, p = 0.040$ ). The constant response pattern was liked more (see Figure 3).



**Figure 3.** Likability results for the two response patterns.

As we found differences with regard to the experienced realism, we included this subscale in a further analysis as a covariate. This caused the significant difference in likeability between the two response patterns to disappear ( $F(1,27) = 3.127, p = 0.088, \eta^2 = 0.104$ ).

In Table 3, the number of participants and the rated likeability of the two response pattern groups divided by the used stop criteria are shown. In terms of the distribution of participants, we found no significant relation between response pattern and stop criteria ( $\chi^2(1) = 0.741, p = 0.389, \phi = 0.157$ ).

**Table 3.** Number of participants and likeability results for the two response patterns divided by the used stop criteria.

Response Pattern	Stop Criteria	No. of Participants (%)	Likeability (SD)
<b>variable</b> ( $N = 16$ )	max. repetition	7 (44%)	3.17 (1.17)
	self-determination	9 (56%)	3.16 (0.89)
<b>constant</b> ( $N = 14$ )	max. repetition	4 (29%)	3.90 (0.81)
	self-determination	10 (71%)	3.84 (0.76)

Note.  $N = 30$ .

With regard to the failure attribution, we found that either the robot or both the robot and the participant were considered responsible (see Table 4), which was independent of the response pattern ( $\chi^2(1) = 0.386, p = 0.534, \phi = -0.115$ ; note: the cell “participant” was excluded for the calculation).

**Table 4.** Distribution of the failure attribution divided by the used stop criteria.

Attribution to	Response Pattern	
	Variable	Constant
robot	8 (50%)	8 (57%)
participant	0 (0%)	1 (7%)
both	8 (50%)	5 (36%)

Note.  $N = 30$ .

## 6. Discussion

The aim of our research was to investigate how caregivers respond to communication failures of robots and whether there are ways to positively influence the caregivers' perceptions and behaviors towards an erroneous robot.

### 6.1. The Impact of Justifications

Our first research question addressed the impact of failure justifications. We assumed that justifying failures either in a human-like or a technical manner would be assessed differently by caregivers. To our surprise, we found no difference in the results of the two failure justifications and, furthermore, that both justifications provided relatively high ratings. An effective explanation should provide truthful and adequate reasons [26]. We believe the high scores obtained for the technical justification were because it fit well with the nature of the agent—as robots are technical devices—and was therefore plausible. However, the human-like explanation, which also scored high, fit very well too. The provided human-like justifications were applicable to one's own experiences and therefore seemed credible.

However, when including gender as a factor in the analyses, we found some differences. Females rated the human-like justification higher and significantly liked this type of justification more. Males showed the opposite evaluation. The technical failure justification was favored. These stereotypical findings indicate that males seem to be more attracted to technological terms than females. In a literature review by Widder, it was shown that people of different gender react differently towards robots [41]. More generally, it was stated that males tend to like and engage more with robots than females. However, some contrary findings were mentioned, too. For example, females showed more positive attitudes towards the idea of robots having emotions. These findings are in line with our results, which likewise indicate that men tend to prefer technical traits and women tend to prefer human-like traits. These preferences could result from a matching effect of gender and gender-specific characteristics. However, it should be noted that some studies have proven this effect, while others have found the exact opposite [42,43]. Since the existing body of research is still ambiguous, further research is needed on this topic. Independent of the different gender preferences, it should be considered whether they should be included in the robot design at all, or whether it should explicitly be omitted. Weßel and colleagues have analyzed ethical problems of gender stereotyping in social robotics and identified possible solutions [44]. Two of the solution strategies they mentioned were neutralization and queering. In this context, neutralization refers to a gender-neutral behavior (speaking and acting). In contrast, queering proposes a certain level of gender fluidity, rather than following a binary concept. With regard to the current study, using both types of justifications simultaneously might accordingly create a mixed or somehow neutral response behavior. In this way, stereotypes can be avoided even with different justifications.

Comparing the likeability results from the failure justification task with the error tolerance task showed that the first task with explanations (justifications) led to higher likeability results than the second task without giving an explanation. In the failure justification task, the robots' likeability was rated with an average of 4.3. In the error tolerance task, which did not include an explanation or any other recovery strategy, the same robot was only rated with 3.5. This indicates that regardless of the particular type, it is generally beneficial to provide an explanation. Our results are therefore consistent with other studies that found a positive influence of recovery strategies on robot perceptions [23–26].

To answer our first research question (R1), we can conclude that a recovery strategy is useful to reduce failure consequences. Regardless of whether human-like or technical justifications were provided, both justifications yielded overall good results. Small differences in the type of justification only resulted from different preferences among men and women, which, however, were only found in relation to likeability.

### 6.2. Tolerance Threshold of Caregivers

Our second research question (R2) aimed to address how tolerant caregivers are with robots in a failed communication and whether there is a threshold for repeating a prompt. With regard to the error tolerance of caregivers, we revealed that the threshold for repeating a request was around seven repetitions when caregivers stopped the interaction self-determined. Seven repetitions still seem very high and are not feasible in the daily nursing practice. It should be noted that, due to the study situation, the participants probably interacted with the robot for a longer time than in real life. Caregivers usually are under time pressure and have to cope with all kinds of demands. Of course, this was not given in the study. Nevertheless, it was interesting to observe what limit emerged in a relaxed situation.

The interaction between caregiver and robot is always mutual. The question is therefore not only how long the user is interacting with the robot but also how long the robot tries to interact with its counterpart before stopping on its own accord. This should not happen too early in the interaction. If the robot aborts the interaction by itself, it takes on the leading part. However, robots should serve caregivers more as a tool [45]. This implies that the decision-making power should remain with the humans. In this way, the distribution of roles between humans and robots can be ensured with a clearly assigned responsibility [46]. A maximum repetition rate of about seven times before the robot independently aborts the interaction seems therefore appropriate. Luger and Sellen reported a similar amount (2–6 repetitions) for users to set their expectation about a system [22]. Overall, it can be stated that the tolerance range for failed interactions lies within the single-digit range and expectations are quickly established. It is important to be aware of this low threshold. Systems or robots that are highly error-prone should prepare solution approaches and recovery strategies to overcome set expectations and support ongoing interactions.

### 6.3. The Influence of the Robot's Response Pattern

We hypothesized that robots speaking with a variable response pattern would be liked more and achieve a greater number of repetitions by the caregivers (H1). Interestingly and against our expectation, we found that the constant response pattern was significantly liked more than the variable pattern. However, this effect disappeared when the experienced realism was included as a covariate. Furthermore, both patterns revealed the same number of repetitions. We therefore have to reject our hypothesis. A reason why the variable response pattern did not achieve better results might be the uncertainty aroused in the participants by providing several options for the failed request. Speaking more slowly, loudly, clearly, or completely rephrasing the question might have resulted in not knowing what really mattered. The results of the IPQ questionnaire indicate that the variable pattern was considered less realistic. Randomly issuing different reasons seemed unlikely for the participants. Overall, the results showed that it is not necessarily worth the effort to implement a variable response behavior in the robot. However, if the reason for a misunderstood communication is indeed, for example, that a person is speaking too quietly, that should be addressed in the request.

We additionally queried who was responsible for the failed interaction. Except for one participant, the majority did not hold themselves solely responsible for the failed interaction. Nevertheless, about half of the participants felt that both parties, i.e., themselves and the robot, were responsible for the failed interaction. Similar attributions have been found in other studies [21]. Mavrina and colleagues found that participants attributed communication breakdowns least to themselves and then to the voice assistant [21]. In their study, an option to attribute the breakdown to the programmer was included, to whom the errors were most frequently attributed. Badia and colleagues stated that the degree of robot autonomy is decisive for how much blame is assigned to the robot in work tasks [29]. With a higher autonomy, more blame is assigned. To get a more detailed understanding of failure attributions, future studies could repeat our study but include further options such as the VR setting or the experimenter. Although in our study the participants were

not responsible for the failure, it is important to note that they also blamed themselves. When designing HRI, this should be considered. In situations where the reason for a failed interaction is known, concrete and transparent feedback should be provided (e.g., stressing the real sources of misunderstanding). Human-, robot-related, and environmental factors can be considered (e.g., [29,47]). This would allow the user to estimate whether the error was caused by him-/her-self (e.g., because of the used voice volume), by the robot (e.g., because of a lack of vocabulary), or due to environmental conditions (e.g., because of ambient noise). This would encourage the users with greater confidence in their actions. Otherwise, blaming oneself unfounded could elicit feelings of being stupid or lacking in technical savvy [22].

#### 6.4. Limitations, Strengths and Future Studies

The VR experiment brought many advantages compared to studies with depicted robots [13,30]. The size of and the proximity to the robot could be sensed, and the spoken words could directly be assigned to the robot by lip movements. Nevertheless, the study lacked a true interaction. In the failure justification task, the failures were only narrated by the robot. Caregivers did not experience the failures themselves. This could be one reason why our participants rated the robot, in general, very high in the failure justification task. In the error tolerance task, participants experienced the failure, but the given answers were initiated by pressing a button operated by the experimenter. Of course, the participants were not aware of that, but this was the reason why the interaction was in general very easily structured. This assumption is supported by the IPQ results. We found an overall high perceived presence in the virtual environment, but the score for involvement was the lowest compared to the other subscales. The results should be replicated with self-experienced failures and in a real interaction.

In the present study, we focused solely on failures. Caregivers did not experience any successful verbal interaction with the robot. Extending the failure-prone with successful sessions would create a more realistic interaction. For future studies, it would be interesting to see what influence failures have when successful interactions have already been experienced. In the study by Mavrina and colleagues, satisfaction was queried after a combination of failed and successful requests [21]. However, not only is the overall assessment important, but also the evolution of specific effects (e.g., satisfaction, trust, forgiveness). Future studies could therefore examine whether the timing of occurring failures has an influence (e.g., failures in the beginning vs. failures at the end of an interaction).

In order to conclude statements on communication patterns of care robots, it is advantageous that we specifically surveyed the group of caregivers. This allowed us to make explicit predictions for this target group. However, this group was highly occupied, especially in times of the pandemic, and the acquisition of participants was difficult. Thus, a disadvantage is the small sample size and the not evenly distributed gender of the participants. Future studies should seek for a greater sample size and acquire more male caregivers as participants.

In conclusion, this study gave an initial insight into how caregivers in particular react to robotic communication failures. Robot designers should generally ensure that justifications are provided in the event of a failed interaction, as the satisfaction with the robot will be less reduced, and due to a transparent explanation, users will become more confident in their behavior.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/robotics11050106/s1>, Table S1: Script for the failure justification task.; Table S2: Used questionnaires of the failure justification task; Audio S1: failure justification\_human-like1; Audio S2: failure justification\_technical1; Audio S3: request\_louder; Audio S4: request\_more clearly; Audio S5: request\_repeat; Audio S6: request\_rephrase; Audio S7: request\_slower.

**Author Contributions:** Conceptualization, K.K. and L.O.; methodology, K.K.; software, K.K.; validation, K.K. and L.O.; formal analysis, K.K.; investigation, K.K.; resources, K.K.; data curation, K.K.; writing—original draft preparation, K.K.; writing—review and editing, L.O.; visualization, K.K.; supervision, L.O.; project administration, L.O.; funding acquisition, L.O. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the German Federal Ministry of Education and Research (BMBF) within the project RoMi “Roboterunterstützung bei Routineaufgaben zur Stärkung des Miteinanders in Pflegeeinrichtungen”, grant number 16SV8436.

**Institutional Review Board Statement:** The study was conducted in accordance with the Declaration of Helsinki, and the protocol was approved by the Ethics Committee of the Humboldt-Universität zu Berlin (2022-09; 10 February 2022).

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** Publicly available datasets were analyzed in this study. Those data can be found here: [https://osf.io/k5sqt/?view\\_only=60aaf5f5bd7b483788eb2cc38942158b](https://osf.io/k5sqt/?view_only=60aaf5f5bd7b483788eb2cc38942158b).

**Acknowledgments:** We would like to express our sincere gratitude to Markus Ahrendt and Claudius Lotz for the great implementation and programming of the virtual environment and the virtual robots. A big thank you goes to Robert Klebbe, Christopher Friese, and Saskia Mischke for their assistance in conducting the study. Many thanks to all participants for their valuable time and to Lars Meese who supported us in contacting the caregivers. The article processing charge was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—491192747 and the Open Access Publication Fund of Humboldt-Universität zu Berlin.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

## References

1. World Health Organization. *State of the World's Nursing 2020: Investing in Education, Jobs and Leadership*; World Health Organization: Geneva, Switzerland, 2020; ISBN 978-92-4-000327-9.
2. Buchan, J.; Catton, H.; Shaffer, F.A. Ageing Well? Policies to Support Older Nurses at Work. *Int. Cent. Nurse Migr.* **2020**, *1*, 1–48.
3. Chang, W.-L.; Šabanović, S.; Huber, L. Situated Analysis of Interactions between Cognitively Impaired Older Adults and the Therapeutic Robot PARO. In *Social Robotics*; Herrmann, G., Pearson, M.J., Lenz, A., Bremner, P., Spiers, A., Leonards, U., Eds.; Lecture Notes in Computer Science; Springer International Publishing: Cham, Germany, 2013; Volume 8239, pp. 371–380. ISBN 978-3-319-02674-9.
4. Papadopoulos, I.; Koulouglioti, C.; Ali, S. Views of Nurses and Other Health and Social Care Workers on the Use of Assistive Humanoid and Animal-like Robots in Health and Social Care: A Scoping Review. *Contemp. Nurse* **2018**, *54*, 425–442. [[CrossRef](#)]
5. Carros, F.; Meurer, J.; Löffler, D.; Unbehaun, D.; Matthies, S.; Koch, I.; Wieching, R.; Randall, D.; Hassenzahl, M.; Wulf, V. Exploring Human-Robot Interaction with the Elderly: Results from a Ten-Week Case Study in a Care Home. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, Honolulu, HI, USA, 25–30 April; ACM: Honolulu, HI, USA, 2020; pp. 1–12.
6. Broekens, J.; Heerink, M.; Rosendal, H. Assistive Social Robots in Elderly Care: A Review. *Gerontechnology* **2009**, *8*, 94–103. [[CrossRef](#)]
7. Holland, J.; Kingston, L.; McCarthy, C.; Armstrong, E.; O'Dwyer, P.; Merz, F.; McConnell, M. Service Robots in the Healthcare Sector. *Robotics* **2021**, *10*, 47. [[CrossRef](#)]
8. International Organization for Standardization ISO 8373:2012(En). Available online: <https://www.iso.org/obp/ui/#iso:std:iso:8373:ed-2:v1:en> (accessed on 25 May 2022).
9. Reeves, B.; Nass, C. The Media Equation: How People Treat Computers, Television, and New Media like Real People and Places. *Choice Rev. Online* **1997**, *34*, 34-3702. [[CrossRef](#)]
10. Severinson-Eklundh, K.; Green, A.; Hüttenrauch, H. Social and Collaborative Aspects of Interaction with a Service Robot. *Robot. Auton. Syst.* **2003**, *42*, 223–234. [[CrossRef](#)]
11. Honig, S.; Oron-Gilad, T. Understanding and Resolving Failures in Human-Robot Interaction: Literature Review and Model Development. *Front. Psychol.* **2018**, *9*, 861. [[CrossRef](#)]
12. Bonarini, A. Communication in Human-Robot Interaction. *Curr. Robot. Rep.* **2020**, *1*, 279–285. [[CrossRef](#)] [[PubMed](#)]
13. Roesler, E.; Manzey, D.; Onnasch, L. A Meta-Analysis on the Effectiveness of Anthropomorphism in Human-Robot Interaction. *Sci. Robot.* **2021**, *6*, eabj5425. [[CrossRef](#)]
14. Marge, M.; Espy-Wilson, C.; Ward, N.G.; Alwan, A.; Artzi, Y.; Bansal, M.; Blankenship, G.; Chai, J.; Daumé, H.; Dey, D.; et al. Spoken Language Interaction with Robots: Recommendations for Future Research. *Comput. Speech Lang.* **2022**, *71*, 101255. [[CrossRef](#)]

15. Bainbridge, W.A.; Hart, J.W.; Kim, E.S.; Scassellati, B. The Benefits of Interactions with Physically Present Robots over Video-Displayed Agents. *Int. J. Soc. Robot.* **2011**, *3*, 41–52. [[CrossRef](#)]
16. Klüber, K.; Onnasch, L. Appearance Is Not Everything—Preferred Feature Combinations for Care Robots. *Comput. Hum. Behav.* **2022**, *128*, 107128. [[CrossRef](#)]
17. Brooks, D.J. A Human-Centric Approach to Autonomous Robot Failures. Ph.D. Thesis, University of Massachusetts Lowell, Lowell, MA, USA, 2017; p. 229.
18. Salem, M.; Lakatos, G.; Amirabdollahian, F.; Dautenhahn, K. Would You Trust a (Faulty) Robot? Effects of Error, Task Type and Personality on Human-Robot Cooperation and Trust. In Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction, Portland, OR, USA, 2–5 March 2015; ACM: Portland, OR, USA, 2015; pp. 141–148.
19. Cho, J.; Rader, E. The Role of Conversational Grounding in Supporting Symbiosis Between People and Digital Assistants. *Proc. ACM Hum.-Comput. Interact.* **2020**, *4*, 1–28. [[CrossRef](#)]
20. Jiang, J.; Jeng, W.; He, D. How Do Users Respond to Voice Input Errors? Lexical and Phonetic Query Reformulation in Voice Search. In Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval, Dublin, Ireland, 28 July–1 August 2013; ACM: Dublin, Ireland, 2013; pp. 143–152.
21. Mavrina, L.; Szczuka, J.; Strathmann, C.; Bohnenkamp, L.M.; Krämer, N.; Kopp, S. “Alexa, You’re Really Stupid”: A Longitudinal Field Study on Communication Breakdowns Between Family Members and a Voice Assistant. *Front. Comput. Sci.* **2022**, *4*, 791704. [[CrossRef](#)]
22. Luger, E.; Sellen, A. “Like Having a Really Bad PA”: The Gulf between User Expectation and Experience of Conversational Agents. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, San Jose, CA, USA, 7–12 May 2016; ACM: San Jose, CA, USA, 2016; pp. 5286–5297.
23. Kim, P.H.; Dirks, K.T.; Cooper, C.D.; Ferrin, D.L. When More Blame Is Better than Less: The Implications of Internal vs. External Attributions for the Repair of Trust after a Competence- vs. Integrity-Based Trust Violation. *Organ. Behav. Hum. Decis. Process.* **2006**, *99*, 49–65. [[CrossRef](#)]
24. Zhang, X. “Sorry, It Was My Fault”: Repairing Trust in Human-Robot Interactions. Master’s Thesis, University of Oklahoma, Norman, OK, USA, 2021.
25. Choi, S.; Mattila, A.S.; Bolton, L.E. To Err Is Human(-Oid): How Do Consumers React to Robot Service Failure and Recovery? *J. Serv. Res.* **2021**, *24*, 354–371. [[CrossRef](#)]
26. Bradley, G.; Sparks, B. Explanations: If, When, and How They Aid Service Recovery. *J. Serv. Mark.* **2012**, *26*, 41–51. [[CrossRef](#)]
27. Dianatfar, M.; Latokartano, J.; Lanz, M. Review on Existing VR/AR Solutions in Human–Robot Collaboration. *Procedia CIRP* **2021**, *97*, 407–411. [[CrossRef](#)]
28. Etzi, R.; Huang, S.; Scurati, G.W.; Lyu, S.; Ferrise, F.; Gallace, A.; Gaggioli, A.; Chirico, A.; Carulli, M.; Bordegoni, M. Using Virtual Reality to Test Human-Robot Interaction During a Collaborative Task. In Proceedings of the Volume 1: 39th Computers and Information in Engineering Conference, Anaheim, CA, USA, 18–21 August 2019; American Society of Mechanical Engineers: Anaheim, CA, USA, 2019; p. V001T02A080.
29. Badia, S.B.I.; Silva, P.A.; Branco, D.; Pinto, A.; Carvalho, C.; Menezes, P.; Almeida, J.; Pilacinski, A. Virtual Reality for Safe Testing and Development in Collaborative Robotics: Challenges and Perspectives. *Electronics* **2022**, *11*, 1726. [[CrossRef](#)]
30. Pan, X.; de C Hamilton, A.F. Why and How to Use Virtual Reality to Study Human Social Interaction: The Challenges of Exploring a New Research Landscape. *Br. J. Psychol.* **2018**, *109*, 395–417. [[CrossRef](#)]
31. Li, R.; van Almkerk, M.; van Waveren, S.; Carter, E.; Leite, I. Comparing Human-Robot Proxemics Between Virtual Reality and the Real World. In Proceedings of the 2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI), Daegu, South Korea, 11–14 March 2019; IEEE: Daegu, Korea, 2019; pp. 431–439.
32. Roesler, E.; Manzey, D.; Onnasch, L. Embodiment Matters in Social HRI Research: Effectiveness of Anthropomorphism on Subjective and Objective Outcomes. *ACM Trans. Hum. Robot Interact.* **2022**, 3555812. [[CrossRef](#)]
33. Weistroffer, V.; Paljic, A.; Fuchs, P.; Hugues, O.; Chodacki, J.-P.; Ligot, P.; Morais, A. Assessing the Acceptability of Human-Robot Co-Presence on Assembly Lines: A Comparison between Actual Situations and Their Virtual Reality Counterparts. In Proceedings of the The 23rd IEEE International Symposium on Robot and Human Interactive Communication, Edinburgh, UK, 25–29 August 2014; IEEE: Edinburgh, UK, August, 2014; pp. 377–384.
34. Roesler, E.; Onnasch, L.; Majer, J.I. The Effect of Anthropomorphism and Failure Comprehensibility on Human-Robot Trust. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* **2020**, *64*, 107–111. [[CrossRef](#)]
35. Heerink, M.; Krose, B.; Evers, V.; Wielinga, B. Measuring Acceptance of an Assistive Social Robot: A Suggested Toolkit. In Proceedings of the RO-MAN 2009—The 18th IEEE International Symposium on Robot and Human Interactive Communication, Toyama, Japan, 27 September–2 October 2009; IEEE: Toyama, Japan, 2009; pp. 528–533.
36. Hur, J.C.; Jang, S.S. Is Consumer Forgiveness Possible?: Examining Rumination and Distraction in Hotel Service Failures. *Int. J. Contemp. Hosp. Manag.* **2019**, *31*, 1567–1587. [[CrossRef](#)]
37. Kidd, C.D. Sociable Robots: The Role of Presence and Task in Human-Robot Interaction. Master Thesis, Massachusetts Institute of Technology, Cambridge, MA, USA, 2003.
38. Bartneck, C.; Kulić, D.; Croft, E.; Zoghbi, S. Measurement Instruments for the Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety of Robots. *Int. J. Soc. Robot.* **2009**, *1*, 71–81. [[CrossRef](#)]
39. Syrdal, D.S.; Dautenhahn, K.; Koay, K.L.; Walters, M.L. The Negative Attitudes towards Robots Scale and Reactions to Robot Behaviour in a Live Human-Robot Interaction Study. In *Adaptive and Emergent Behaviour and Complex Systems*; SSAISB: Brighton, UK, 2009.

40. Schubert, T.; Friedmann, F.; Regenbrecht, H. The Experience of Presence: Factor Analytic Insights. *Presence Teleoperators Virtual Environ.* **2001**, *10*, 266–281. [[CrossRef](#)]
41. Widder, D.G. Gender and Robots: A Literature Review 2022. *arXiv* **2022**, arXiv:2206.04716.
42. Siegel, M.; Breazeal, C.; Norton, M.I. Persuasive Robotics: The Influence of Robot Gender on Human Behavior. In Proceedings of the 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems, St. Louis, MO, USA, 10–15 October 2009; IEEE: St. Louis, MO, USA, 2009; pp. 2563–2568.
43. Eyssel, F.; Hegel, F. (S)He’s Got the Look: Gender Stereotyping of Robots. *J. Appl. Soc. Psychol.* **2012**, *42*, 2213–2230. [[CrossRef](#)]
44. Weßel, M.; Ellerich-Groppe, N.; Schweda, M. Stereotyping of Social Robots in Eldercare: An Explorative Analysis of Ethical Problems and Possible Solutions. In *Frontiers in Artificial Intelligence and Applications*; Nørskov, M., Seibt, J., Quick, O.S., Eds.; IOS Press: Oldenburg, Germany, 2020; ISBN 978-1-64368-154-2.
45. Cifuentes, C.A.; Pinto, M.J.; Céspedes, N.; Múnica, M. Social Robots in Therapy and Care. *Curr. Robot. Rep.* **2020**, *1*, 59–74. [[CrossRef](#)]
46. Sharkey, A.; Sharkey, N. We Need to Talk about Deception in Social Robotics! *Ethics Inf. Technol.* **2021**, *23*, 309–316. [[CrossRef](#)]
47. Hancock, P.A.; Billings, D.R.; Schaefer, K.E.; Chen, J.Y.C.; de Visser, E.J.; Parasuraman, R. A Meta-Analysis of Factors Affecting Trust in Human-Robot Interaction. *Hum. Factors J. Hum. Factors Ergon. Soc.* **2011**, *53*, 517–527. [[CrossRef](#)]