ABSTRACT
It is inevitable that robots will make mistakes in social situations with humans. However making mistakes is part of the social aspect of humans. Important when making a social mistake is how the mistake is acted upon, whether or not the error is remedied. This research has investigated if and to what extent a negative first impression of a robot can be remedied by the same robot that caused it.
To investigate this, a series of controlled experiments have been conducted. A robot approached a participant and then performed a task with the participant. The experiment had 3 conditions (regular, mistake and recovery) to compare the effect of making a social mistake and the attempt to recover from it. After the experiment the impression the participants had of the robot was measured in the qualities of warmth and competence using a questionnaire.
A small difference was found in combined warmth and competence rating, albeit not a significant one. It was found that for one attribute, sensitivity, the robot significantly scored better in the recovery condition compared to both the regular and mistake condition. These results should be able to help design future robots.

Keywords
first impression, social recovery, social robotics, social mistake, HRI, sensitivity

1. INTRODUCTION
Robots are in increasing amounts among humans. Being among humans, these robots need to interact with humans. For this interaction to feel natural and comfortable these robots need to abide certain social laws. These social laws can be implicit or explicit and might be different from the ones humans follow.
As the nature of these social laws is very complex and possibly even contradicting at points, it is inevitable that robots will make mistakes with their implementation of these laws. Making mistakes, however, is part of the social aspect of humans; humans make mistakes as well. Important when making a social mistake is how the mistake is acted upon, i.e. if the error is remedied. It is even imaginable that intentionally making a mistake, giving the opportunity to recover from it, can have a positive effect regarding human-robot relations.
While this research focussed primarily on mistakes made when initially approaching a person, the conclusions drawn might be applicable for other types of mistakes as well or at the very least provide a guideline for handling other types of mistakes. In doing so the findings of this experiment should be able to help design future robots, especially with regards to error recovery.

1.1 Research Questions
The problem introduced gives rise to the following research questions:

1. Does, after making a negative first impression in approaching a human, acknowledging the mistake and then backing off and repeating the approach yield a better impression of the robot then when ignoring the mistake?
2. Does intentionally making a mistake and then attempting to recover from it yield a better impression of the robot then when the mistake is never made?

Aside from simply answering these questions with yes or no, this research also aimed to provide insight into the magnitude of the possible changes to the impression the robot makes.

1.2 Theoretical Background
1.2.1 Robots and Proxemics
Existing research showed that robots are capable of changing their proxemics, the interrelated observations and theories of man’s use of space as a specialized elaboration of culture [3], and thus their relation with humans. Mumm & Mutlu [6] showed that physical distance with people, a more specific part of proxemics, can be influenced by gaze behaviour dependent on how likeable the robot is. Furthermore, Kuzuoka et al.[4] showed this for a different area of proxemics. They experimented with a robot museum guide that was able to actively reconfigure the F-formation, a positional model based on proxemics, it had with a person or even a group of people.
Lohse et al.[5] showed that robots can improve communicating their intentions by letting a robot emit a sound congruent to its approach velocity. Based on the discussed proxemics research and other findings related to proxemics, it is expected that even minor details in the way a robot approaches can have an impact on how the robot is perceived. This translates into the expectation that the different experiment conditions should also have an impact on how the robot is perceived. This change in
perception in turn should have an impact on the first impression of the robot and furthermore the relation between the robot and the participant.

Additionally, a solid baseline on human-robot interaction and thus including proxemics is given by the media equation [7], which states that humans perceive other media and robots as social actors, similar to how they perceive other humans. Since humans can make up for mistakes by apologizing, influencing their relation, this theory would suggest apologizing would have a similar effect for robots.

1.2.2 First Impressions

When people meet for the first time, both sides quickly form a first impression about the other. Research done in psychology [2] showed that impressions humans have of other objects and in this case social actors, can be universally measured in warmth and competence. Here warmth mostly influences whether the impression is positive or negative, while competence mostly affects the extremity of that impression. The same research showed that these dimensions are judged at the initial impression and that they can be hard to recover from later on.

Although there is little research completed with robots attempting to recover from their own mistakes, let alone recovering from a negative first impression, there was research done for Interactive Virtual Agents (IVAs). Bergmann, Eysel & Kopp [1] researched how IVAs are judged based on the first impression they give, and showed that this initial impression made by an IVA can be changed. While this does not have to translate directly into robots, the media equation [7] does make this likely. This would in turn suggest that there is some kind of apology for robots that can remedy their mistakes made while approaching a person.

1.3 Hypotheses

Based on the existing literature discussed in section 1.2, the following hypotheses describe the possible results that are expected.

1. After making a negative first impression in approaching a human, backing off and repeating the approach without making a mistake yields a better impression of the robot then ignoring the mistake.

2. Intentionally making a mistake and then attempting to recover from it yields a better impression of the robot when the mistake is never made.

Following is the method used, results gained and the conclusion based on those results. The paper closes with a discussion section evaluating this paper’s findings.

2. RESEARCH METHOD

To research whether or not robots can recover from making a negative initial impression while approaching, an experiment was conducted in a controlled environment, with a set-up similar to experiments done in previous research [5, 4, 9].

2.1 Protocol

Participants were given instructions to wait at a certain spot in the room for a robot to approach the participant and then perform a small task with a robot. When the participant agreed to have understood the instructions the robot approached the participant and then performed the small task with this participant. The small task mostly consisted of the participant listening to a short dialogue by the robot and is explained in more detail in section 2.3. After the task was done the participant was asked to fill in a questionnaire.

The experiment was conducted with 3 conditions. Since the experiment relies heavily on the first impression made, the experiment was conducted between subjects. In all 3 conditions the instructions and task given to the participants were the same, the only difference between conditions being the way the robot approached. The 3 conditions were as follows:

1. **Regular-condition**: The robot approached the participant and performed the task without explicitly making a mistake in initially approaching the participant.

2. **Mistake-condition**: The robot attempted to make a negative first impression while initially approaching the participant. The robot then proceeded normally with the task, ignoring the mistake made.

3. **Recovery-condition**: The robot attempted to make a negative first impression while initially approaching the participant. The robot then tried to recover from this negative impression before continuing normally with the task.

2.2 Set-up

An overview of the set-up used in this experiment is given in Figure 1. The participants entered and were briefed on the left side of the room, indicated with ‘B’ in Figure 1. After the briefing they were told to wait at the location indicated with a ‘P’. An orange goblet was located on the edge of a table near where the participant was told to wait, indicated by the orange dot in the overview.

At the start of the experiment the robot was located at the position indicated with ‘E’ in Figure 1, this was also where the experimenter controlled the robot from. Standing within the briefing area or at the ‘P’, position ‘E’ was obstructed from view by a wall. During the experiment the robot traversed the path as indicated from ‘E’ to ‘P’. The questionnaire was filled in at the experimenter’s desk at position ‘E’.

![Figure 1. Overview of the setup used in the experiment.](image)

2.3 Task

Similar to the experiment done with the robot museum guide [4], the task the participants had to perform was to listen to the robot giving an explanation about a simple orange plastic goblet placed right next to the participant’s position.
The explanation was in English and divided in three parts, the robot asked the participant after each part if he or she wanted to know more information. The first audio clip was about goblets in general, the second audio clip gave a detailed definition of the colour orange, and finally the third clip was about orange as the Dutch colour. The audio clips were made using a free text-to-speech program using a male English voice.

In order to add some interactivity to the task, the robot asked the participant after explaining about the object to remove a post-it note with "Kick me!" written on it attached to the robot’s back.

2.4 Robot
The robot used for this experiment was a Giraff telepresence robot as seen in Figure 2. Where normally the telepresence robot would display its driver, in this case the experimenter, on its display, a black screen was used instead. Furthermore the display was covered up with a very simple robot’s face consisting only of 2 dots representing eyes. The robot also had some additional equipment attached to it, unrelated to this experiment. The robot was controlled by the experimenter, using the Wizard of Oz technique.

The negative impression on the participant was created by ignoring the participant’s personal space while approaching that participant, letting the robot stop at a distance less than 10 cm from the participant [8].

The recovery attempt made by the robot consisted of showing awareness of the mistake made by apologizing for it with an audio clip of: "I'm sorry, this is probably too close, let me try that again." Following this clip the robot backed off to about 2 meters, then approached again stopping at the same comfortable distance from the participant as the regular-condition, about 1.2 meters. The thought behind this being that the robot showed that it is capable of doing that what it did wrong the first time, correct the second time.

2.5 Measures
The questionnaire that the participants were asked to fill in after the experiment first attempted to measure the positiveness of the impression left behind by the robot by measuring warmth and competence. To measure warmth and competence the questionnaire asked participants to scale several aspects of the robot that can be categorized in either warmth (pleasant, friendly, helpful, likeable, approachable, sociable) or competence (intelligent, organized, expert, thorough) on a scale from 1 to 7, this is based on a questionnaire conducted in a previous research [1]. For each attribute the definition according to the Cambridge Dictionary\(^1\) of that attribute was given. After the attributes the questionnaire contained 4 demographic questions, asking for age, nationality, gender and robot experience. The questionnaire concluded with a check if the participant recognized the condition he or she has experienced, if the experiment was deemed fun and a direct question if he or she felt the robot’s recovery method (if applicable) was effective. For each page the order of questions was randomized for each participant.

In addition to the questionnaire, the amount of times the participant was interested in additional information was measured. Whether or not the participant actually went to the trouble of removing the post-it note attached to the robot’s back was also counted in an attempt to measure the amount of effort the participant was willing to spent on the robot.

2.6 Participants
A total of 45 participants participated in the experiment, 15 participants for each condition. The participants were randomly gathered from around the campus, where the experiment was conducted. No selection on the participants was performed. An effort was made to equalize the amount of male and female participants, yet due to the nature of the experiment location 78% of the participants were male, leaving 22% female participants. The age of the participants ranged from 17 to 60 years old, with an average age of 24 years old. Of the participants 91% had the Dutch nationality. All participants agreed with the consent form. 44 Out of 45 participants were filmed during the experiment.

3. RESULTS
The results gathered from the questionnaire attribute questions for warmth, competence and combined (warmth and competence) are shown in Table 1. A visual comparison of the combined scores can be found in Figure 3, the mean scores per attribute per condition are given in Figure 4.

Participants allowed the robot to give 84% of the information in the regular and recovery-condition, 80 % in the mistake-condition. Of the participants, 96% actually removed the post-it note attached to the robot’s back, both the regular and recovery method had 1 participant that did not remove the post-it note. The participant that did not remove the post-it note in the recovery-condition actually (gently, with respect to the equipment) kicked the robot.
Following the assumption that the data was normally distributed, statistical analysis was conducted using ANOVA to test for a significant difference of the combined scores between the conditions. In order to answer the research question, 2 tests were performed: the regular-condition against the mistake-condition and the regular-condition against the recovery-condition.

As a result of making Figure 4, a seemingly big difference for the attributes sensitive and possibly sociable and intelligent was spotted. Following this discovery, statistical analysis on these individual attributes was performed.

### 3.2 Expected Results

It was expected based on hypothesis 1 that the recovery-condition would leave a better impression (combined score) than the mistake-condition, yet results were not significant enough to confirm this hypotheses (p = 0.59). It was expected based on hypothesis 2 that the recovery-condition would leave a better impression (combined score) than the regular-condition, yet results were not significant enough to confirm this hypothesis (p = 0.15). Additionally, based on literature, it was expected that the mistake-condition would have a lower combined score than the regular-condition. Contrary to this expectation, no significant difference between the regular and mistake-condition was found (p = 0.36). These comparisons are visualised in Figure 3.

Confirming to expectations, most participants indicated that they found the robot’s recovery method used in this experiment to be effective when directly asked (67% yes, 27% somewhat, 7% no).

### 3.3 Unforeseen Results

As stated in the analysis, a seemingly big difference was found between conditions for the attributes sensitive and possibly sociable and intelligent. Table 2 shows the precise results for these attributes.

As with the data from the combined results, the data for these attributes was tested for normality using a Shapiro-Wilk test. The test rejected normality and thus disallowed the usage of ANOVA to check for significant differences. Instead a non-parametric Kruskal-Wallis test was used to compare mean ranks among the conditions of each attribute.

Following this analysis, a significant difference (p = 0.002) was found between the conditions for the sensitivity attribute. The difference in mean rank per condition suggests that the recovery-condition scored significantly higher than both other conditions (16.63; 19.93 and 32.43 for respectively the regular-, mistake- and recovery-condition).
Figure 4. Chart showing the scores for each condition for each robot attribute. The attributes are grouped by warmth (left) and competence (right).

No other attributes had a significant difference between their conditions.

4. CONCLUSION

Although it was hinted by the data that both hypotheses might be true, not enough significant data was found to provide significant answers to the posed research questions as to whether the recovery method used provided a beneficial boost in the robot-human relationship compared to not applying the recovery method. However, when looking at the individual attributes, as opposed to the overall combined results as stated by the research question, some significant results were found.

Based on the results found for the sensitivity attribute it is suggested that letting a robot recover from a mistake does significantly improve certain aspects of the impression made, namely the attribute sensitivity. On top of this it is even suggested that letting a robot intentionally make a mistake, just to be able to recover from it, can significantly improve the impression made by the robot.

5. DISCUSSION

The gathered combined results where quite different from the results that were hoped to be found. While the sample size might have simply been too small, there were a number of issues that might have impacted the results of the experiments.

Aside from the encountered issues, the found significant results for the sensitivity attribute were only discovered after conducting the experiments. For a full prove that the sensitivity attribute is positively influenced by recovery behaviour, further research with an adjusted goal to investigate the effect of recovery behaviour on the sensitivity attribute should be performed.

5.1 Limitations

There were several issues that might have limited the significance of the experiment. The robot’s look might have been one of the factors that influenced the outcome of the experiments. There was extra equipment, unused by this experiment, attached to the robot. As first impressions are among other things reliant on looks, this equipment could have impacted the first impression the robot gave to the participants. On the other hand, the robot’s look did not change during the period in which the experiments were conducted, meaning that this effect should be equalized across conditions and have minimal impact on the results found.

The period in which the experiments were conducted was relatively long, the experiments were spread out over 2.5 weeks. Some experiments were conducted in the morning, most experiments were conducted in the afternoon. The participants being mostly students, this difference in time could have influenced the eagerness of the participants to participate, since they most likely had lectures and or homework to do before or after the experiment.

The students that did participate were mostly gathered from the same area, the area close to the experiment location. Since this area is mostly used by technical studies such as computer science, electrical engineering and mathematics, the group of participants is likely not representative of a normal population. However, since only comparisons between the conditions were made, the effect of the group of participants not being fully representative should have been minimized.

The questionnaire was based on an existing questionnaire used by a previous research [1], yet was not completely copied. While adding definitions for the attributes, a duplicate definition was found. Since it seemed pointless to include 2 attributes that had the same definition, the duplication was resolved by not including the attribute affable in the list of attributes that had to be rated. Changing the questionnaire might have impacted its validity and thus that of the results found.

Aside from these issues that might have impacted the overall significance of the results found, there is a case to be made that the low volume of the robot as indicated by some of the participants resulted in the robot being better understandable from close up, the mistake-condition. This may in turn be part of the reason why the mistake-condition did not score significantly lower than the regular-condition, which was expected based on literature.

Despite no significant differences in results found between the regular and mistake-condition, most of the partici-
pants from the mistake-condition correctly identified their experiment condition. This confirms that the participants recognized the behaviour of the robot, yet this behaviour did apparently not transfer into significant differences when rating the robot’s attributes that were asked in the questionnaire.

5.2 Technical Issues

Several technical issues related to the robot were encountered during the experiments, preventing the experiments from being optimally conducted. The main technical issue was with the internet connectivity. The robot was controlled using the internet to connect the robot to the pilot program, as a result of this the encountered connectivity issues made it in some cases troublesome to control the robot. The problems that were caused by the connectivity issues ranged from a minor silence in the audio clip to the connection completely dropping, forcing the experiment to conclude.

In addition to the network issues there were some issues with the robot’s tilt sensors used to control the robot’s head. Since a lot of extra equipment was attached to the robot (not used by this experiment) the robot’s head was heavier than intended by its design. This resulted in sometimes failing head tilt sensors, leaving the robot’s head in a very weird looking position, likely impacting the impression the robot made.

Major network issues occurred in 6 experiments, 3 in the regular-condition and 3 in the recovery-condition. There were 3 cases in which the head tilt sensor malfunctioned, all 3 in the recovery-condition. This gave a total of 9 cases in which technical issues likely impacted the results, of which 6 in the recovery-condition. Due to limited time and in order to keep the amount of experiments at a reasonable number, it was decided to nevertheless include these experiments in the results.

Aside from these obvious technical issues the audio volume of the robot was slightly different between experiments. This was due to the experiments being spread out over a time period in which the equipment was used for other purposes as well. Some participants explicitly indicated that the volume was a bit low, even though the experimenter, who was further away from the robot, never had issues understanding the robot’s dialogue.

5.3 Miscellaneous Observations

On top of the results to the research question and the results for the individual attributes, some other interesting observations were made.

Despite a set-up in which one could argue it is pretty obvious that the Wizard of Oz technique is applied, most participants seemed to believe the robot was moving autonomously. Some participants questioned this and asked whether or not the experimenter was controlling the robot.

While the findings of the research seem to stick to the claim made by the media equation [7] that robots are treated as social actors, behaviour was observed that goes against this. Some participants had trouble hearing or understanding the robot’s voice, leaning towards the robot in an attempt to better understand the robot. This leaning was done in a manner that would be considered socially unacceptable among people.

When given Dutch instructions to answer clearly with “Ja” or “Nee”, many of the participants answered the robot’s questions with “Yes” or “No”. This is probably caused by the robot’s dialogue being in English itself.

5.4 Future Work

While the results of this experiment should already help in developing future robot behaviour, the focus of the research lied on change in the overall impression of the robot and the found significant results in individual attributes were only discovered after the experiment. Due to this, further research is needed to fully prove that certain attributes, and possibly in turn the overall human-robot relation, can be improved by applying a recovery method.

This future research should, besides finding a way to reduce the issues encountered in this research, focus more on the individual aspects, in particular sensitivity, of the impression made and how they are influenced by recovery behaviour. Aside from this shift in focus, which should provide more detailed findings per aspect of the impression made, further research with different recovery tactics could also provide interesting results and build on the discoveries made in this research.

6. REFERENCES


