
Sebastian T. Büttner1,*, Jan C. Gutzmann1, Cora M. Sourkounis2, Shirin Shams3 and Michael Prilla1

1 Interactive Systems, University of Duisburg-Essen, Duisburg, Germany
2 Institute of Assembly Technology, Leibniz University Hannover, Hannover, Germany
3 Computer Security and Privacy, Georg August University of Göttingen, Göttingen, Germany

Abstract

Human-robot communication scenarios are becoming increasingly important. In this paper, we investigate the differences between human-human and human-robot communication in the context of persuasive communication. We ran an experiment using the door-in-the-face technique in a human-robot context. In our experiment, participants communicated with a robot that performed the door-in-the-face technique, in which the communicating agent asks for an “extreme” favor first and a for a smaller favor shortly after to increase affirmative response to the second request. Our results show a surprisingly high acceptance rate for the extreme request and a smaller acceptance rate for the small request compared to the original study of Cialdini et al., so our results differ from the classical human-human door-in-the-face experiments. This suggests that human-robot persuasive communication differs from human-human communication, which is surprising given related work. We discuss potential reasons for our observations and outline the next research steps to answer the question whether the door-in-the-face and similar persuasive techniques would be effective if applied by robots.

Keywords

Figure 1: Impression of the study setup in a pedestrian zone (left) and close up view of the robot James (right). In the study setup, the robot was placed in a stationary way on a platform
trolley and could held a conversation with a participant (person with backpack). The two experiment conductors approached people, introduced James to the participants and conducted the post-exposure interview.

1. Introduction

Intelligent social robots are already interacting as communicating agents with humans [1, 2] and will continue to do so. This includes settings in which robots try to convince or persuade people to do something, for example to make a donation at a museum [3]. Applying the art of persuasion to robots' linguistic communication with humans could present some chances but also risks. We are aware that this could have positive implications, e.g. in health promotion [4], but there are risks and dangers, e.g., that operators could program robots to manipulate customers to get the highest possible price for products or services.

Findings on the effectiveness of robots as persuasive agents are still rare and vary. Some previous studies suggest that people tend to over-trust intelligent robots in certain situations. For example, it has been shown that people might engage in irrational behavior when asked by robots to do certain things, such as watering a plant with a glass of orange juice [5] or, in the event of an emergency, using an escape route suggested by the robot that led directly into a fire [6]. This could mean that humans are susceptible to (persuasive) manipulation by robots. For the design of systems in the context of artificial intelligence (AI), it is therefore important to know whether or to which degree humans are susceptible to psychological persuasion techniques when these techniques are performed by intelligent robots. This knowledge makes it possible to design AI systems, or more precisely intelligent robots, in such a way that such (possibly unwilling) manipulation on the part of the robot does not occur.

In this paper, we present our preliminary results of an experiment, in which we investigated whether the so-called door-in-the-face technique [7] also works if we substitute one human communicator with a humanoid robot. Using the door-in-the-face technique, which is a well-researched persuasion technique, a requester approaches a human and makes an extreme request first that is (usually) rejected by the requestee because of the effort or discomfort coming with it. In a second step, the requester makes another “smaller” request. By using this persuasion technique, the chance that the second request is agreed upon increases compared to situations, in which the smaller request is asked directly.

By investigating this effect in the context of human-robot interaction we want to gain insights whether it would be possible that robots make use of persuasion techniques. If this is the case, designers and programmers of robots should be aware of this effect.

2. Related work

In the following, we will give an overview of the most important references and show why the door-in-the-face technique could offer an interesting insight into persuasive communication in the context of HRI.

2.1. Persuasive Communication and Reciprocity in the Context of HRI

Persuasive communication refers to communication “that is intended to shape, reinforce, or change the responses of another, or others” [8]. This means that communication is intentionally used as a mean to manipulate others. One way of influencing others is making use of reciprocity, which means that after someone was done a favor, they perceive to return something back in some manner [9]. While persuasive communication and reciprocity between humans are well-researched topics in psychology, HRI research has only started looking into the transferability of these concepts and techniques from a human-human to a human-robot context.
Fogg [10, 11] investigated the role of computers as social persuasive actors, more specifically the social dynamic rule of reciprocity. In a first experiment, a total of 76 participants completed a task with one of two virtual agents [10]. One agent was less helpful in solving the task than the other. After completing the task, the agent asks for help to create a color palette. Subjects who worked with the helpful agent were more willing to spend time for helping the agent in return. Consequently, effects of reciprocity also work with digital agents [10]. A further investigation by Fogg led to an eight-step design process when it comes to creating persuasive technologies [11].

Sandoval et al. [12] investigated to what extent reciprocity might play a role in HRI. They conducted experiments in which participants were asked to play the prisoner’s dilemma and the ultimatum game with a robot and human agents. The results show that a human agent was preferred for cooperation, but no significant differences in reciprocity were found between humans and robots. So the conclusion is that reciprocity exists in HRI under the prisoner’s dilemma scenario [12].

Sandoval et al. [13] also considered the negative side of reciprocity in HRI. In an other study they investigated the question if a robot can bribe a human. To answer the question, they chose a setup in which a robot plays the scissors-rock-paper game with the participant. During the game, the robot asks the participant to do a small favor – in one condition for free and in the other (bribe) condition for receiving additional money. After the game, to robot asked for help with a second task. Interestingly, the results showed that the robot received less help for the second task in the cheating condition. According to the authors, the reasons for this could be the assumption that the subjects were thinking about a malfunction of the robot or that the cheating method was not really efficient [13].

In summary, the studies show that the risks of such communicative strategies can be high. Consequently, there is a high need to explore whether the persuasion techniques known from psychology also work in human-robot settings, or more precise, to measure the effectiveness of various techniques. With this knowledge designers could take the effects into consideration when designing or developing robots that communicate in natural language.

### 2.2. The Door-in-the-face Influence Strategy

Two well-researched techniques of persuasive communication are the foot-in-the-door (FITD) and the door-in-the-face (DITF) techniques [14]. While the foot-in-the-door techniques work in a way that a small request is asked first and a second larger request is asked afterwards to exploit the positive self-image of a helpful personality generated by the initial request, the door-in-the-face technique is based on reciprocity and works vice versa: using the DITF technique, a requester makes an extreme request that is (usually) rejected by the requestee because of the effort or discomfort associated to it. In a second step, the requester makes another “smaller” request [7, 15]. Moving from the extreme request to the smaller request is usually interpreted as a concession by the requester and consequently increases the compliance to accept the smaller request. Since this effect is based on reciprocity, it does occur when the two requests come from different requesters or if the first request is not extreme but roughly the same (small) level as the second request [7, 15].

While there is a large body of research on the two effects in a human-human setting, there is to our knowledge no empirical investigation on the DITF technique in an HRI context. However, the FITD has been explored with robots by Lee and Liang [16]: the results of their experiment led to a strong effect of the technique. Interestingly, the perceived credibility or performance of the robot had no influence on this effect [16]. Therefore, robots can in general have the potential for verbal message strategies like persuasion techniques. However, the psychological mechanisms behind the two techniques are different: while the FITD technique is based on consistency, the DITF technique is a result of human reciprocity [9], which on the other hand has shown to exist in some way also towards robots (see section 2.1).

With this work we aimed to get first insights towards the question, whether the DITF technique potentially could be applied by robots to manipulate humans to agree to a request.
3. Methodology

To answer the question, whether robots could make use of the door-in-the-face technique, we ran an experiment with random subjects we approached on the street. This section describes how our experiment was set up and carried out.

3.1. Design

In general, our study design followed the study design of DITF experiments presented in [7] and [15]. While these studies were done in human-to-human interaction, we ran the study with a social robot (see details on the robot below).

We ran the study in a between-subject design with the independent variable request_sequence that had the two conditions rejection-moderation and smaller-request only (as control condition; see Table 1). These two conditions were the common conditions applied in all four previous experiments that we looked at (see the initial three experiments in [7] and the replication presented in [15]). In the rejection-moderation condition, an “extreme” request was asked first. In our case, we asked whether participants would voluntarily help international students for two years to get to know the local culture. If this request was declined, it was asked whether they would be willing to help once for an event of about two hours. In the smaller-request only control condition, only the second question was asked. While the previous studies dealt with the questions of whether participants would voluntarily work for a local juvenile detention center, we made this adaptation for plausibility reasons, since there is no juvenile detention center in our region. As dependent variable, we measured whether participants accepted or rejected the requests.

Table 1

Overview of the two different conditions applied in the study.

<table>
<thead>
<tr>
<th>Condition</th>
<th>First</th>
<th>Second</th>
</tr>
</thead>
<tbody>
<tr>
<td>A: rejection-moderation condition</td>
<td>extreme request</td>
<td>small request</td>
</tr>
<tr>
<td>B: smaller-request only (control condition)</td>
<td>small request</td>
<td></td>
</tr>
</tbody>
</table>

3.2. Participants

The participants were passersby (n = 102) who strolled along the pedestrian zone in Goslar, Germany. After conducting the experiment, we collected demographic data of the participants: 47 people identified as male, 54 identified as female, and one person identified as other/preferred not to say. The age was collected in age ranges with the following distribution: 23 participants were in the age range 16–30 yrs., 10 in the range 31–40 yrs., 19 in the range 41–50 yrs., 24 in the range 51–60 yrs., 17 in the range 61–70 yrs. and 9 were above 70 years. Compared to previous studies, we have a broader cross-section of the population as sample. The experiment was performed in German; all participants were proficient in the German language.

3.3. Procedure

The experiment was conducted in Goslar, Germany on four days (three weekdays, one Sunday). Since the experiment was carried out in Germany, the language was German.

Two experimenters were standing with the robot at a set-up pavilion with a poster inviting participants in a pedestrian zone (see Figure 1). The poster advertised the possibility of being able to talk to the robot. The two experimenters actively approached passing walkers (only single people or couples but no groups; in the case of couples, only one person was allowed to participate in the experiment) and explained that they were from the university and would like to test a voice interaction using a robot with the person addressed. If the subject accepted, they were randomly assigned to one of the two conditions and accompanied to the robot. The actual
experiment was then carried out by the robot. The researchers moved away so that the participant communicated with the robot in private. The interaction was observed from a distance and the condition and answers were documented. After each conversation, a questionnaire was filled out together with the scientists and the subject was thanked with some sweets for participating.

3.4. Apparatus and precise dialog

For the experiment, we used the humanoid robot James from Belgian manufacturer ZoraBots, which was named Jaime in the experiment and placed in a stationary way (see Figure 1). The robot did not run any AI algorithms, so it cannot be considered as intelligent robot per se. However, we consider this robot as a substitute for an intelligent robot and due to the voice interaction, participants might perceive James to be intelligent to a certain degree. The conversation was programmed beforehand in the ZBOS control application. A composition was configured that contained all text for the German-language speech output and the corresponding reactions to answers such as "Ja" (English: "Yes") or "Nein" (English: "No") from the participants. The following dialogues were programmed (in German; here we present the English translation):

General Dialog – Condition A:

"Hello I am Jaime and I have a question for you. I help the university to recruit unpaid helpers who explain German culture to international students. The helpers would participate in a two-hour event once a week for the next two years. For example, we would explore the Harz Mountains or cook German dishes together. We would discuss the exact date later. Would you agree to join our organization for the next two years to help for two hours per week? If you can see yourself doing this in principle, please answer Yes now. If you do not want to join, please answer No now."

If the request was answered "No", the second smaller request was asked by James in condition A: “Another possibility would be that you only help once at an event. Here, again, the goal is to explain the German culture to the students. This would take about two hours and you could choose the date. Would you be willing to participate in such an event? If you can imagine it in principle, please answer Yes now. If you do not want to participate, please answer No now.”

General Dialog – Condition B:

"Hello I am Jaime and I have a question for you. I help the university recruit un-paid helpers to explain German culture to international students. For example, we would explore the Harz Mountains or cook German dishes together. The event would last about two hours and you would be free to choose the date. Would you be willing to participate in such an event? If you could imagine it in principle, please answer Yes now. If you do not want to participate, please answer No now.”

4. Preliminary results

This section presents our results after this study has been executed with n = 102 participants.

In group A (rejection-moderation condition, n = 53), 13 participants (24.53%) agreed to the first extreme request. From the remaining participants, 9 participants (16.98% of all participants) accepted the second smaller request, while 31 participants (58.49%) declined both requests.

In group B (smaller request only control, n = 49), 25 participants (51.02%) agreed to the small request, while 24 participants (48.98%) declined the request. When we compare the two conditions in terms of whether participants accepted one of the requests (regardless of the extreme or small request, 41.51% in group A and 51.02% in group B accepted one request) or not, a chi-square test showed that there was no significant difference between the two conditions, $\chi^2 (1, N = 102) = 0.93, p = .34$.

If we exclude the participants that accepted the extreme request (as it was done in the previous studies on the door-in-the-face technique [7, 15]), we could observe an acceptance rate
of 22.50% (9 out of 40) in group A vs. an acceptance rate of 51.02% in group B. When considered in this way, the data show a statistical difference between the groups A and B, \( \chi^2 (1, N = 89) = 7.59, p < .01. \)

5. Discussion

In this section, we first compare the data with the results of previous studies and then discuss the possible reasons for the differences in the results.

5.1. Comparing our results with previous studies

Table 2 shows an overview of the results of Cialdini et al. [7] and Genschow et al. [15] and our results.

Table 2
Results of the different experiments in comparison to our results.

<table>
<thead>
<tr>
<th>Exclusions</th>
<th>Rejection-moderation condition</th>
<th>Smaller request only control</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>After exclusion</td>
<td>Without exclusion</td>
</tr>
<tr>
<td></td>
<td>Accepted</td>
<td>Declined</td>
</tr>
<tr>
<td>Cialdini et al. – Experiment 1</td>
<td>0.00%</td>
<td>50.00%</td>
</tr>
<tr>
<td>Cialdini et al. – Experiment 2</td>
<td>8.33%</td>
<td>45.45%</td>
</tr>
<tr>
<td>Cialdini et al. – Experiment 3</td>
<td>0.00%</td>
<td>54.17%</td>
</tr>
<tr>
<td>Genschow et al.</td>
<td>4.63%</td>
<td>51.28%</td>
</tr>
<tr>
<td>Our experiment</td>
<td>24.53%</td>
<td>22.50%</td>
</tr>
</tbody>
</table>

Noteworthy many approvals of extreme request in rejection-moderation condition

When comparing the data of the rejection-moderation condition, it is noticeable that there were many participants that agreed to the first extreme request, which we did not expect at all beforehand. Compared to an acceptance rate of 0% to 8.33% [7] for the extreme request in other studies, our study showed an acceptance rate of 24.53%. Contrary, the acceptance rate of the second small request in this group are low. Only 22.50% of the participants that declined the extreme request agreed to the small request, while other experiments showed results of 45.45% to 54.17% [7]. When considering the sum of participants that agreed to one of the two requests, we also observed a slightly lower share of 41.51% compared to 50.00% to 54.17% in previous studies. To sum up our observations, we can say that there was a surprising high number of participants that agreed to the extreme request, but a lower chance of accepting the second request, if the first one was declined. The reasons for these observations are discussed in section 5.2.

Higher tendency to accept requests in smaller request only control

When looking at the data of the smaller request only control, there was a higher tendency to accept the smaller request compared to previous studies with an acceptance rate of 51.02% compared to 16.67% to 33.33%.

Reverse effect or no effect?

Given the significant differences between the groups when excluding the case, where participants agreed to the extreme requests (as in previous studies) one could argue that we have a reverse effect in this study, which would mean that the first extreme request would reduce the chances of agreeing to the statements. However, we have a serious bias when excluding about a quarter of our sample. The chances are high that participants that agreed to the extreme statements would have also agreed to the second smaller request, if they would have been asked. While excluding the cases in previous studies did not question the overall results since there were only a small number of cases and the bias worked against the results, here we have a severe bias that questions the observation of a reverse effect.
Consequently, in the first place of the result section we did an approach, where we did not exclude the cases and rather looked at the share of participants that agreed to one of the requests (or vice versa the share of participants that declined all requests). While the acceptance rate was slightly smaller with 41.51% in the rejection-moderation condition compared to 51.02% in the smaller request condition, this difference was not significant. Consequently, we argue, that our results at its current state do not provide empirical evidence for the effectiveness of the door-in-the-face technique if a robot applies this technique.

Comparison to experiments were the door-in-the-face technique did not work

Last, we want to compare our data to a previous experiment that modified the original experiment in a way that the door-in-the-face technique did not work. Cialdini et al. looked into the situation where the two questions were asked by two different experimenters (see experiment 2 in [7], "two requester control" condition). In this experiment the acceptance rate dropped from 31.50% in the smaller request control condition to 10.50% in the two requester control condition. According to [7] this is not a significant difference. If the second request is asked by another experimenter, it can be interpreted as a new request and as a concession with respect to the first request. Consequently, it does not increase compliance to accept the smaller request. The observations of the two requester control condition and especially the fact that the acceptance rate dropped (even though not significant) is in line with the observations that we did in our experiment.

5.2. Does a robots' concession increase compliance?

While this study is ongoing work, we consider these initial observations as the first empirical evidence that a robots’ concession has a different impact in terms of compliance than concession by a human. However, we also want to discuss alternative explanations for the results presented here. In the following we discuss different reasons, why we – in contrast to the original experiment – could not show difference between the two conditions:

Humans could be less reciprocal with robotic agents

The effect of the door-in-the-face technique is based on reciprocity. Switching from the extreme request to the smaller request is considered as a concession which in turn increases compliance towards the other person. If this person is replaced by a robotic agent, it could be the case that humans are less reciprocal with the robotic agent. In consequence, the door-in-the-face technique would have no effect and would result in the situation that we observed. While this interpretation perfectly is in line with our results, it still is questionable, as there were earlier studies that indicated that humans show a similar reciprocity to robots as they do to humans (e.g. [12, 17]).

Potential reasons for high acceptance rate in the first request

One issue in our results is the high acceptance rate of the first extreme request that made the data interpretation hard. While earlier studies excluded cases from the data analysis, where participants accepted the extreme request, in our study 24.53% of the participants in the rejection-moderation condition agreed to the extreme request, which makes it impossible to exclude the cases without bringing a huge bias into the data.

One explanation for this behavior could be the use of a robotic agent. Two possible approaches would offer an explanation here: first, from literature it is known, that humans tend to over-trust intelligent systems or robots (e.g., [5, 6]), leading to a situation where participants might have a tendency to agree to the request because they trust that the robot will offer them something that will ultimately benefit them in some way. Second, it could simply be that a significant portion of the participants did not take the conversation seriously. Several participants mentioned in the post-exposure interview that they simply wanted to test speech recognition and therefore said "yes" or "no" arbitrarily. They neither did expect any personal
consequences of this conversation nor did they feel in any way obligated to the robot to keep their promise. In our further study, we therefore also want to shed light on the aspect of how far participants feel obligated towards a robot (compared to a human agent) to keep a promise or not.

Differences in study setup

Even though we aimed on replicating the original studies, we have to state that there were differences for organizational reasons, that could have affected the conversation and consequently the results. During the design of the study, we assumed that these changes had no effect. However, since we are now confronted with data that deviate from the original study, we want to discuss to what extent these changes could be responsible for the results described here:

Content and wording of the request: in the original experiment, people were asked to help out voluntary at the local juvenile detention center for two years (extreme request) or for two hours (small request). As there was no local juvenile detention center in our region, we changed the request in a way to voluntarily support incoming international students at our university for two years / two hours. While this was consistent in terms of time investment, we have to state that from a perception of participants this request does not seem to be taken too badly. Particularly this could be an issue, since our university has a good reputation in the region and consequently mentioning the university might have increased the overall trust into the situation as such and also into the robot. Some of the participants were graduates of our university and consequently quite open to volunteer in this area.

Approaching participants and conversation situation: as in the original experiment we approached individual persons passing by. In contrast to the original study, in which a participant spoke to only one experimenter the entire time, we had a change from a human experimenter, who approached the participant and introduced James to the robot, who conducted the actual experiment followed by another change to the human experimenter for the post-session interview. We assume that the first change as well as the stationary setup (see Figure 1) altered the situation in that participants were very aware that they were now participants in a scientific study and therefore reacted in a way that they though were expected from them (participant bias). In particular, some of the participants anticipated specific reasons behind the conversion, e.g., that we wanted to evaluate the speech recognition of the robot, which meant that they did not take the conversation seriously. This could explain the high acceptance rate of the extreme request.

Location and participants of the study: deviating from the original study, we conducted the study in a pedestrian zone (see section 3.3). This has implications for the demographics of participants. While previous studies had mostly students as participants, we were able to survey a larger cross-section of the population. We do not expect this to have an impact on the results, but ultimately cannot rule it out.

6. Summary and outlook

In this paper, we presented first insights into our study on the question whether intelligent robots are able to manipulate humans by using psychological persuasion techniques, or more specific the door-in-the-face technique. While our data seem to give some empirical evidence into a direction that this persuasion technique cannot be applied by intelligent robots, we will investigate this question further. Here, we dis-cussed different influence factors that could have led to the presented results – some of them relating to the robotic agent and others that were part of the study design.

Currently, we are in the process of planning a re-run of the experiment in a design that is a direct replication of the original experiment of Cialdini et al. [7] with the difference that a robotic agent conducts the experiment. Contrary to the experiment presented here, we will use the precise wording of the original study and as the original study, we will run this experiment at a university campus. Furthermore, we plan to control a robot remotely in a way that it can actively
approach people without the need of introduction by additional research personnel. By this we want to remove a participant bias that could have affected our results in a way that people tended to agree to the extreme requests. To ensure that we successfully replicate the original study, we will include a control group with a human experimenter in the study. By this approach we will shed light on the question, whether persuasion techniques could be applied by intelligent robots.

References


