How Socially Assistive Robotics Should Praise Users to Increase Motivation

B. Kurstjens
University of Twente,
P.O. Box 217, 7500AE Enschede
The Netherlands
b.kurstjens@student.utwente.nl

ABSTRACT
This paper describes research on motivating people with merely verbal praise. The ultimate goal is to improve robots in increasing the intrinsic motivation of people to complete tasks. In this research the Nao robot of Aldebaran will verbally compliment students after completing a puzzle. The research will focus on the effect of adjusting the heaviness of the praise to match the challenge of the task. The results suggest that adjusting the heaviness of the praise has an effect on intrinsic motivation, however the results are a significantly worse when the praise is adjusted to the difficulty of the puzzle.

Keywords
Socially Assistive Robotics, intrinsic motivation, verbal praise, challenging tasks, Nao of Aldebaran, mechanical voice

1. INTRODUCTION
The increasing amount of elderly around the world will have a huge impact on the economics and the social sphere and therefore adjustments will have to be made [11]. Socially Assistive Robotics (SAR) are made to motivate elderly to improve health conditions and the quality of life. Instead of physically helping elderly in their homes, the robots try to stimulate the elderly to stay fit. By keeping elderly fit, SAR can help them stay at their home and thus reduce costs at care centers [5]. As robots still are very expensive, research has to be done to make sure the robots live up to their full potential. This research is done to improve the robot motivational competences.

1.1 Motivation
There are two main types of motivation; extrinsic motivation and intrinsic motivation. Extrinsic motivation means doing something for a separable outcome. For example rewards like grades and money are extrinsic motivators. Intrinsic motivation means doing something that is inherently interesting or enjoyable [9].

Intrinsic motivation is effective for long-term task compliance and can be affected, among others, by the following three external factors; Praise, indirect competition and the optimal challenge of a task [3]. The effect of intrinsic motivation on long-term task compliance is very important, as elderly have to work for a long period with the robot to stay fit.

1.2 Verbal Praise
This paper will describe the effects of changing the verbal praise to match the challenge of the task, to find if there is a relation between the optimal challenge of a task and the heaviness of the praise to increase intrinsic motivation.

Verbal praise is rather difficult to measure and alter, as the perception depends on multiple factors like age, gender and culture. [7]

Moreover there are two types of verbal praise; mastery praise and social-comparison praise. Mastery praise aims to compliment the skill of the user instead of the result achieved. For example “You have great insight” compliments the skill ‘insight’. A person could have had great insight, regardless of the overall result. Social-comparison praise aims to compliment the result of the user in comparison to other results. For example “You did better than most others” compliments the user on the result achieved in comparison to other users. The mastery praise has usually a better impact on intrinsic motivation than Social-Comparison praise. [2] As noted before the verbal praise is also age, gender and culture dependent and the literature [2, 6, 14] is based on research with children.

1.3 Human robot interaction
Next to motivation and praise, the human-robot interaction is important to review. Humans have ‘common sense’ to know when to compliment another, but robots have to be programmed to know which compliment to use at what time to be credible. As the robots get more evolved each year, the opportunities in human robot interaction become more advanced. For example robots are now able to recognize movement, objects, speech and body language [10]. These abilities are important for the robot to be credible and to convince people that it is sincere.

Previous research by Fasola and Mataric show that a relational robot that praises the user encourages the motivation of a user more than a non-relational robot [4]. This research suggests that praise has a positive effect on the motivation of elderly; however it does not go into the best way of verbally praising the user.

This paper explores a part of the best way of verbally praising a human by a robot to increase their intrinsic motivation. It is written to add a piece to the puzzle and motivate others to complete the puzzle.

The next section describes the research question and the focus of the paper. The third section states the hypothesis to be proved by the results. In the fourth section the method is discussed. It starts with the robot used, continues with research on which praise is used and finishes with the actual experiment.
The fifth section shows the results and the sixth section discusses these results. The last section concludes the found facts and suggests future research.

2. RESEARCH QUESTION
The question to be answered in the research is:

“What is the effect of adjusting the robot’s praise to match the difficulty of the task?”

For example; if the robot praises very enthusiastically without regarding the difficulty of the task, would the user still be motivated? Or would the user be more motivated when the robot is more enthusiastic about more difficult tasks?

The answer on this research question will help the implementation of the robot to be optimal to motivate the user. It is a small step towards the larger question: “How should a robot motivate a human?”

There is little research to be found on robots complimenting humans and due to the factors that influence the perception of praise, this research field is difficult to investigate properly. For this reason this paper focuses on a seemingly very small part of the praise and with only two conditions as is described in section 3.4.2.

3. HYPOTHESIS
This research proposes that if the praise is matched to the difficulty of the task, then the intrinsic motivation will increase. Thus the specific praise should get better results (more liked) than the random praise.

Hypothesis: The intrinsic motivation of the participants will be higher in the specific praise condition than in the random praise condition

4. METHOD OF RESEARCH
4.1 The Robot
In this study the Nao robot of Aldebaran will be used (figure 1). The robot has a humanlike body; a head, chest, two arms and two legs. The robot is able to ‘see’ with a camera and ‘hear’ with a microphone. It has speakers to speak and is able to move the head, arms and legs to make gestures. This experiment however will only use the rotation of the head and the speakers.

Figure 1. Nao of Aldebaran

4.2 Participant Statistics
A total of 20 Dutch students participated in the experiment (evenly spread between two conditions). The students are between 19 and 25 years old with an average age of 22 (SD = 1.65). The population consisted of 17 male and 3 female students. Only two students had previous experience with a robot (one had seen a robot, the other had interacted with one). The students joined the experiment voluntarily and did not receive any compensation.

4.3 Verbal Praise research
Before being able to match the challenge of a task to the heaviness of the praise, there has to be some scale in heaviness of praise. The robot has a mechanical voice, which does not influence the enthusiasm of certain words. This means that the word ‘great’ is not pronounced with more enthusiasm than the word ‘good’. Moreover, the robot does not use any additional movements to support its verbal praise, as the movements might be seen by one participant who is less focused on the game, but not a participant who is focused. The heaviness of the praise is thus completely dependent on the content said and the perception of the participants.

Even though mastery praise has a better effect on intrinsic motivation than social-comparison, they will both be used. The game does not have enough skills to compliment, thus to avoid repetition some social comparison praise was used as well.

Research with children [2, 14] inspired this research to use certain words like ‘good’, ‘great’ and ‘best’. However as [7] suggests the praise is age specific and this research will not involve children, there has to be done more research before using the praise on students.

A small group of students (1 female, 3 male) were asked separately to listen to the robot and sort the praise in three categories. In Table 1 all the sentences the robot was programmed with are mentioned. The sentences actually used in the praise experiment are written in bold. The robot was sitting on a table next to a laptop, on which the student sorted the praise.

Based on the answers of the students, the sentences in Table 1 were put into three categories. The sentences picked for the actual praise experiment were sentences which the students voted unanimously in the same category.

<table>
<thead>
<tr>
<th>Category</th>
<th>#</th>
<th>Verbal Praise</th>
</tr>
</thead>
<tbody>
<tr>
<td>Light Praise</td>
<td>1</td>
<td>Well done, seems like you are getting the hang of it.</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Nice, you are a good Solver!</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>That’s good progress.</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>You are doing well, keep it up.</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>Good job, you seem to do better than most.</td>
</tr>
<tr>
<td>Medium Praise</td>
<td>6</td>
<td>Great job! You are better than the others!</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>I haven’t seen any better scores yet, keep it up!</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>You have scored in the top 3, stunning work!</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>I see you truly understand the game, nice!</td>
</tr>
</tbody>
</table>
As the robot’s mechanical voice can be hard to follow due to the pronunciation and speed of certain combination of words, some extra punctuation has been used in Nao Choregraphe (the programming software used in this research to program the Nao). To keep the praise readable, this punctuation has been left out of this paper.

4.4 The Experiment

The praise has been categorized, so the next step is a game that has increasingly difficult levels. The levels have to be short, challenging and enjoyable, as a long and boring game would demotivate the participants.

4.4.1 The Game

This research has used the game “Unblock me” by ‘GzDev’ (http://unblock-me.software.informer.com/), as it has different levels (from easy to hard) and all students in the pilot enjoy the game. During the experiment the participants play the game on a laptop. The game contains a dark block which has to be dragged to the exit to the right. To reach the exit other blocks have to be moved out of the way. The order to move the blocks is important; mindlessly dragging the blocks from one side to the other, usually does not result in a solution. However finding a solution without first moving the blocks around is often too hard. The game, as most games, can be trained and people who are used to playing it show better results than beginners.

Figure 2 is a screenshot of the game with additional arrows and an exit sign to show in which directions the blocks are able to move and where the darkest block has to go. The game has different levels (beginner, intermediate, advanced and expert).

<table>
<thead>
<tr>
<th>Level</th>
<th>Praise</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>You have definitely gotten the hang of it, great job!</td>
</tr>
<tr>
<td>11</td>
<td>That’s impressive, great job!</td>
</tr>
<tr>
<td>12</td>
<td>Very good! You have great insight!</td>
</tr>
<tr>
<td>13</td>
<td>Amazing! You must be king of this game! Your insight in this game is perfect!</td>
</tr>
<tr>
<td>14</td>
<td>Wow! You have beaten all odds on that one! No one will be able to beat that one!</td>
</tr>
</tbody>
</table>

Heavy praise

To test if the difficulty of the puzzle should be matched to the heaviness of the praise, two conditions have been tested.

- **Condition 1: Random praise.** Each task completed will be praised by the robot, but the robot will choose randomly from the list.
  For example the robot can use a very enthusiastic heavy praise (13 or 14) for the first simple puzzle, while use a light praise (1 or 5) for the last hard puzzle.

- **Condition 2: Specific praise.** Each task will be combined a sentence, selected on task difficulty and amount of enthusiasm in the praise. As difficulty increases with each level, the praise numbers simply do the same and increase with each level (1, 5, 6, 12, 13 and last 14).

4.4.3 The procedure

When the participant arrives he/she was asked to fill out a consent form, which states that he/she voluntarily joined and that it is allowed to record the experiment for academic use in papers.

Before the game can start, the robot gives an introduction of the game to the participant and wishes the participant good luck. This ensures that the robot is noticed at the beginning and gives the robot a leading role. After the introduction the participant starts playing the first level. The Nao notifies the participant to try the next level, once a level is completed.

The robot continues to ask the participant to try the next level until six levels have been finished. It does not motivate the participant during a level, as it could have an effect the results. The motivation during the level cannot be given equally, as participants do not always need it. So only at the end of a level a participant is praised. As all participants have to finish six levels, all the participants get the same amount of praise. At the end of the experiment the robot thanks the participant and asks them to fill out the questionnaires.

The pilot test was too hard, as it took the participants too long to solve 9 puzzles and thus the participants became demotivated. In the actual experiment the participants only needed to finish 6 puzzles. This meant that only 6 of the 9 sentences (1, 5, 6, 12, 13 and 14) of Table 1 in section 3.4.3 were used to avoid differences between conditions.

In figure 3 on the next page a participant is playing the game and the robot is watching the screen. Next to verbal praise, the robot only moved his head from the person to the screen and vice versa to catch the attention of the participant. It did not move anything else as other movements might have an effect on the results. This effect cannot be accounted for without extra conditions.
4.4.4 Measurements

To measure the motivation of the users, several indicators are used. The indicators are measured by a questionnaire given after the interaction with the robot. This questionnaire asked about the enjoyment of the puzzles and interaction with the robot, as enjoyment is strongly related to intrinsic motivation. The questionnaire was given to the participant at the end of the experiment.

The statements are derived from [3] and [13]. The questionnaire contained statements that are measured by a 5-point Likert scale with 1 meaning ‘Strongly Agree’, 3 meaning ‘Neither’ and 5 meaning ‘Strongly Disagree’ and two open questions. The questions are written in table 2 on the next page with an explanation why this question is asked.

As there are only two conditions and 20 participants, standard unpaired t-tests will be used to calculate if there is a statistically significant difference between the two conditions.

5. RESULTS

Participants gave lower, so better results to the random praise condition (RC) than on the specific condition (SC), except for question 2, 3 and 5. Question 2 had a mean of 2.30 for both conditions. This indicates that the participants liked the game equally in both conditions. This result suggests that the game did not influence any differences found in motivation. Question 3 had a mean of 1.90 for RC (SD = 1.20) and a 1.6 for SC (SD = 0.52) t(18) = 0.73, P = 0.48, this means that in both of the conditions the participants were able to understand the robot easily. Question 5 had a mean of 2.00 for RC (SD = 0.47) and a 1.70 for SC (SD = 0.48), t(18) = 1.08, P = 0.30, which indicates that participants in both conditions were focused on the game.

The results of question 5 in both conditions indicated that the robot did not get a lot of attention. The random praise condition scored a mean of 3.7 (SD = 0.67), which is almost statistically significant better than the specific praise condition with a mean of 4.3 (SD = 0.67), t(18) = 2.00, P = 0.06.

In both conditions the robot did not score well on “I think the Nao is a good motivator” and “The Nao motivated me a lot to solve the puzzle” as both conditions scored a 3 or higher and even though the random condition (respectively M = 3.0, SD = 1.05 and M = 3.1, SD = 0.88) scored better than the specific condition (respectively M = 3.30, SD = 0.95 and M = 3.70, SD = 1.25), thus the difference between the scores are not statistically significant. These scores show that the participants did not perceive the Nao as a motivator, regardless of the condition.

Three questions did return statistically significant results by using an unpaired t-test and all three in favor of the random condition. In the random condition question 8 scored a mean of 2.1 (SD = 0.88), while the specific condition only scored a 3.1 (SD = 0.88), t(18) = 2.54, P = 0.02. The results of question 9 also indicate a significant preference for RC (M = 2.3, SD = 1.06) over SC (M = 3.3, SD = 0.95), t(18) = 2.22, P = 0.04. Lastly the results of question 11 indicate a significant preference for RC (M = 2.4, SD = 1.26) over SC (M = 3.4, SD = 0.70), t(18) = 2.19, P = 0.04. These statistically significant results all indicate that the Nao did motivate the participants in the random condition more than in the specific condition.

Table 2. Questionnaire

<table>
<thead>
<tr>
<th>#</th>
<th>Closed questions</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>I enjoyed doing these puzzles.</td>
<td>To test whether the person was motivated by the game itself.</td>
</tr>
<tr>
<td>2</td>
<td>I would like to do these puzzles more often.</td>
<td>Big differences between participants might indicate a different game should be used.1</td>
</tr>
<tr>
<td>3</td>
<td>The robot was easy to hear and understand.</td>
<td>To test if the robot should change its pronunciation.</td>
</tr>
<tr>
<td>4</td>
<td>I focused a lot on the robot</td>
<td>To check if the robot has an active role in this experiment.</td>
</tr>
<tr>
<td>5</td>
<td>I focused a lot on the game</td>
<td>To check if the participant was motivated to play the game.</td>
</tr>
<tr>
<td>6</td>
<td>I think the Nao is a good motivator.</td>
<td>To test if the participant felt the robot was able to motivate in general.</td>
</tr>
<tr>
<td>7</td>
<td>The Nao motivated me a lot to solve the puzzle.</td>
<td>To test whether the participant felt the Nao motivated them in the game.</td>
</tr>
<tr>
<td>8</td>
<td>The robot made the game more fun</td>
<td>If the participant enjoys the game more with the robot and would like to play more games with the robot in the future, the robot has improved the intrinsic motivation.2</td>
</tr>
<tr>
<td>9</td>
<td>I would like to play more games with the Nao in the future</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>The robot was credible</td>
<td>Credibility is very important with praise, as incredible praise does not improve</td>
</tr>
</tbody>
</table>

1 If the motivation of the participants is very different due to the likability of the game, then the results could be hard to measure.
2 This enjoyment might be inflicted due to the new experience, if so; this would not have a long lasting effect.

Figure 3. Setup of the experiment
On question 10, which tests the credibility of the robot, the random condition scores an average of 3.0 (SD = 0.94), which is better than the specific condition with an average of 3.6 (SD = 0.84). These results indicate a slight trend towards the random condition with t(18) = 1.50, P = 0.15. These results indicate that the participants in general did not believe the compliments did not sound very credible and did not feel like they were complimented on their performance. This suggestion is important to investigate further, as credibility is crucial to good praise. 7 participants wanted that the robot would move more and were disappointed that it only used its head and speakers. 6 participants suggested that the robot should talk more during the game, like give tips or say motivational words (e.g. “You can do it!”). These two suggestions are chosen for on purpose (section 4.4.2). 6 participants thought the voice was either too childish or too mechanical.

6. DISCUSSION
The results suggest that the hypothesis is not supported. However the results also indicate that there is a difference between the specific praise condition and the random praise, as there are statistically significant results.

There might be several reasons why the results do not support the hypothesis and do partly support the opposite.

- The puzzles might have gotten too difficult at the later levels. This would mean that the participant did not feel like he/she were able to solve it properly. A very heavy praise would not feel credible, when the participant does not feel like he/she did well.
- The verbal praise used in the experiment was tested on only four students. A larger test is needed to see whether the praise truly is categorized as suggested in this experiment. The use of both mastery praise and social-comparison praise could also have an uncalculated impact on this experiment.
- The results show an almost statistically significance in favor for the random condition on focus on the robot. It might be possible that when the robot motivates the participant better; it is more liked and thus it gets more attention. However it could also indicate that the robot had a bigger impact on the students by coincidence and that would interfere with the results obtained on motivation.

As stated in the beginning of this section, this research does suggest there is a relation between the kind of praise and the task to be completed to motivate a user, however due to the focus of this research the reason for that relationship is not discovered.

As the robot is more liked in the random condition and the participant is more motivated to play the game again in the future, the timing of compliments seems to be relevant to the intrinsic motivation of the participant. When and how the robot should praise the user, has not been found though.

Last but not least it should be taken into account that this research has only used Dutch students. As noted in section 1.2 the compliments are culture, gender and age dependent. For example two of the participants wrote that the robot was childish.

The credibility of the robot also indicates another (possibly related) problem. Most students (6/10 of RC and 9/10 of SC) did not think the robot was credible, which has a major influence on this research. More research has to be done to find the right timing and the right words to motivate a person. This study focusses on the challenge of the task, instead of the performance of the participant. It would be interesting to research whether adjusting the compliments to the performance of the participant, would work better than adjusting them to the challenge of the task.

- The specific praise got more intense after each level. This could result in the feeling that the robot is working down a list of compliments (which it was), more than when the robot mixes the compliments. The idea that the robot is simply working down a list of compliments, would in turn enhance the feeling that the robot is not credible. The t-test on credibility however does not show a significant preference for either the random condition (M = 3.00, SD = 0.94) or the specific condition (M = 3.60, SD = 0.84), t(18) = 1.50, p = 0.15.
- Most participants argued they did not pay attention to the robot whatsoever and felt the robot was abundant. That could explain the negative results on ‘I think the Nao is a good motivator’ and ‘The Nao motivated me a lot to solve the puzzle’. The compliments might feel less credible while the robot is not actively involved in the game.
7. CONCLUSION
This paper discussed if there matching the verbal praise of a robot to the difficulty of a task increases the intrinsic motivation of a human. It describes a method to see if there is a relation between the difficulty of a task and kind of praise used.

“What is the effect of adjusting the robot’s praise to match the difficulty of the task?”

The results that the method produced suggest that the heaviness of the praise should not increase similarly to the challenge of the task, regardless of the performance of the user. Moreover the results indicate that increasing the heaviness of the praise and the challenge of the level equally, cause the user to be less motivated than randomly assigning praise to a task.

This means that it is important to choose which praise the robot uses and when it says the words carefully, to properly motivate a user. There should be further research to find the proper factors and use the right words. I would suggest to research whether compliments can be adjusted to facial expressions, as they will tell the robot whether the user is proud to finish a task or annoyed when he/she feels incompetent. Another factor that should be researched is combining the praise to the performance of a participant, for example giving the participant a heavy praise for doing a task quickly and a lighter praise for completing a task slowly.

8. REFERENCES


