Can Cognitive biases in Robots make more 'Likeable' Human-Robot Interactions than the Robots without such Biases: Case Studies using Five Biases on Humanoid Robot

Biswas M, Murray J

University of Lincoln, Lincoln, United Kingdom

Abstract

The research presented in the paper aims to develop long-term companionship between cognitively imperfect robots and humans. In order to develop cognitively imperfect robot, the research suggests to implement various cognitive biases in a robot's interactive behaviours. In our understanding, such cognitively biased behaviours in robot will help the participants to relate with it easily. In the current paper, we show comparative results of the experiments using five biased and one non-biased algorithms in a 3D printed humanoid robot MARC. The results from the experiments show that the participants initially liked the robot with biased and imperfect behaviours than the same robots without any mistakes and biases.

Keywords: Human-Robot Interaction, Human-Robot Long-term Companionship, Humanoid robot, Cognitive Bias, Imperfect Robots

Can Cognitive biases in Robots make more 'Likeable' Human-Robot Interactions than the Robots without such Biases: Case Studies using Five Biases on Humanoid Robot

The study presented in this paper seeks to better understand human-robot interaction and with selected 'cognitive biases' to provide a more human-preferred interaction. Existing robot interactions are mainly based on a set of well-ordered and structured rules, which can repeat regardless of the person or social situation. This can lead to interactions which might make it difficult for humans to empathize with the robot after a number of interactions. The research presented in this paper tests five cognitive biases, such as, misattribution, empathy gap, Dunning-Kruger effects, self-serving and humors effects on a life-size humanoid robot, see Fig. 1, and compare the results with non-biased interactions to find out participant's preferences to the interactions.

According to Breazeal (2001), a social robot should be socially intelligent and should have sufficient social knowledge. To develop social intelligence in social robots, researchers study various methods to allow a robot to adapt to human-like behaviour based social roles. Some of these more popular methods suggest developing human-like attributes in robots, such as, trait based personality attributes, gesture and emotions expressions, anthropomorphism. Dautenhahn (2009) investigated the identifying links between human personality and attributed robot personality where the team investigated human and robot personality traits as part of a human-robot interaction trial. Lee (2006) showed that developing cognitive personality and trait attributes in robots can make it more acceptable to humans, also expressing emotions and mood changing in interactions can help to make the attachment bond stronger between user and the robot. Meerbeek et al (2009) designed interactive personality process in robots which was based

on Duffy's anthropomorphism idea. Duffy (2003) suggested that anthropomorphic or lifelike features should be carefully designed and should be aimed at making the interaction with the robot more intuitive, pleasant and easy. Reeves and Nass (1996) argued that users usually show biased driven certain personality traits to machines (PC & others). Later in 2008, Walters et al investigated people's perceptions based on robot appearances and associated attention-seeking features in video-based Human Robot Interaction trials. In the recent years, Moshkina et al (2011) in Samsung Research Lab has developed a cognitive model which includes traits based personality, attitudes, mood and emotions in robot (TAME) to get humanlike responses.

The above studies discuss various approaches to making a robot more human-like so that it would be easy for people to interact with the robots. However, researchers argue that it is challenging for a robotic system to become relevant and highly individualized to the special needs of each user in the particular beneficiary population (Tapus A, 2007). However, to develop the robot more personal and make their responses much humanlike we propose to develop humanlike different cognitive biases in robots. As such, we investigate this different and unique approach, which is by applying selected cognitive bias to provide a more humanlike interaction. Scientists suggested that cognitive biases have reasonable amount of influence in human's characteristics and behaviours (A Wilke, 2012). Haselton (2005) suggested that people behaves uniquely which is largely influenced by individual's thinking, genetics, social norms, culture and needs. Kahneman (1972) suggested that human thinking is affected by a variety range of biases which influence humans in making various decisions and judgements which sometime can be fallible (making mistakes, forgetfulness, misunderstanding, arrogance, over excitement, idiotic behaviours and others). Such behaviours are common in people and we observe and experience such behaviours in daily life. In our understanding, the differences in cognitive thinking which

are influenced by various biases affect to make the individual's interactions unique, natural and humanlike. But in developing humanlike robots, we sometimes ignore such facts and make the robot without such humanlike behaviours. These cognitive biased behaviours (e.g. forgetfulness, making mistakes etc.) have not been tested in robots and long-term human-robot interactions yet. In our research, we develop few of the cognitive biases in the humanoid robot MARC and find out the effects of biases in human-robot interactions.

In the experiments presented in this paper, two different types of interactions were performed (e.g. conversational and game-playing) between the participant and the robot spanning a two-month time period in order to provide 'long-term time period' effects. In this paper we introduce a model demonstrating biased behaviours in the robot MARC (Figure 1) and, how such biased behaviours influences the interactions with the participant. At the end of the experiments, we compare participant's responses between the robot with a biased behaviour and the robot without the bias, showing the impact on long-term human-robot interaction the bias creates.

The study presented in this paper aims to find out if the cognitive biases helps to create the human-robot interactions better than a robot without such biases. In this case, the effects of cognitive biases and imperfectness have measured by comparing the interactions of both biased and non-biased algorithms of the robot. However, a robot's cognitive abilities depend on the robot itself, i.e. the way robot interact. To test our hypothesis, we use MARC - a 3D printed humanoid robot and the components of five cognitive biases. The cognitive biases were chosen from human's common behaviours, such as, misattribution, empathy, empathy gap, Dunning-Kruger with several characteristics imperfectness, such as, forgetfulness, wrong assumption, misinformation and others to test on the robots.

The Project: Cognitive Bias and Imperfectness in Robots

Current social robots are able to anthropomorphize and mimic human actions but their actions are limited and may not provide sufficient reasons for people to create and maintain social relationship (Baxter P, 2012). Baxter raised the memory issue of the robot where the robot could not carry out the previous interactions. In later section we show that the forgetfulness can also be important factor in human-robot interactions. However, it is likely in human interactions that people meet with other humans and are able to form different kinds of relationships. From that, we raise a simple question, 'What happens in human-human interaction which lacks in human-robot interaction that prevents a social relationship between the robot and human?'

To answer this question, it is necessary to investigate how humans interact and form relationships. As discussed earlier that human behaviours are unique based on the individual's thinking, genetics, social norms and culture (Haselton G, 2005). Unfortunately, human thinking is affected by a range of biases (Khaneman D, 1972) which influence humans in making decision and judgments.

In general, the cognitive biases among other factors have large influence to make human behaviours as the human-like cognitively imperfect (Wilke A, 2012). In our understanding, such differences in cognitive imperfectness among individuals are what make human interactions unique, natural and human-like. In existing social robotics, the robots imitate different social queues for example: eye-gazing, talking and body movements etc. But the other behavioural neutrality which includes faults, unintentional mistakes, task imperfectness are still absent in the current social robots. Sometimes a robot's social behaviours lack that of a human's common characteristics such as, idiocracy, humour and common mistakes. Many robots are able to present social behaviours in human-robot interactions but 'humanlike behaviours' are presented

in such manner that participant finds difficulties to relate with the robot. Hayoun (2014) asked an important question, “How can we interact with something 'more perfect' than we are?” As we know, making faults and misjudgements are common human characteristics, but we try to create something that’s more perfect than us. As discussed earlier, researches show that human-like traits based personality (Walters M, 2008) and emotion expressions can make human-robot interaction more enjoyable to human (Lee K, 2006). But cognitive biases in robots to allow the robots make common mistakes like humans have not been tested yet. If the robot shows tiredness, stressfulness after repeating works and complains to its owner, then what would the owner do? Get annoyed and replace the robot, or accept it and console in human-like manners? But, either way, it should make the user to think about, and that can lead to a certain type of relationship.

Studies suggest that various cognitive biases have a reasonable amount of influences on human thinking process to make misjudgments, mistakes and fallible activities (Baron, 2007). Such misjudgments, mistakes and fallible activities creates individual’s social behaviour human-like cognitively imperfect (Michael T, 1999). Bless et al (2004) suggested that cognitive biases can influence on human’s behaviours towards positive or negative ways. Biases effect on individual’s decision making, characteristics behaviours and social beliefs. In our understandings, such cognitive biases and common humanlike imperfection (e.g. misjudgments, mistakes and fallible activities) effect individual’s general communicative behaviours in human-human interaction and that makes the interaction humanlike natural. But, robots in the other hand, lack to present human-like cognitive characteristics in human-robot interaction and that might prevent the interaction to become human-like natural. We therefore expect that such ‘neutrality issue’ in human-robot interaction can be solved by using humanlike cognitive

imperfection (e.g. making mistakes, forgetfulness and fallible activities) in a robot's mode of interactions. Not only that, cognitive biases and imperfectness can make the robot-human interactions easy to understand and familiar to humans.

Hypothesis

The current research is based on the very important hypothesis, which is:

"Can the introduction of cognitive bias in robot influence Human-Robot Interaction in long period of time?"

The related research questions are:

1. Despite the robot's appearance and functions and features, can a robot demonstrates the important aspects of human-human and human-robot interactions by engaging with humans in short-term/long-term interaction based on cognitive biased behaviours?

2. By introducing cognitive biases in robot, is it possible to develop humanlike biased behaviours in robots which can influence human-robot interactions?

3. Will cognitive biases help humans to develop long-time human-robot interactions?

In the current study, two different experiments were done where in the 1st experiment MARC make conversation and in the 2nd experiment, MARC plays Roshambo (Rock-Paper-Scissors) with the participants. In each experiment, we maintained a parallel interaction experiment between the non-biased robot and participants to find out the difference between biased and non-biased interactions.

Experiments with MARC the humanoid robot

In the previous experiments (Biswas M, 14,15), we used ERWIN and MyKeepon robots to test misattribution and empathy gap biases. In both cases, participants favored for the biased versions of the interactions. In these experiments a humanoid robot MARC was used to test misattribution, empathy gap, Dunning-Kruger, self-serving and humors effects biases.

The robot MARC

We used a 3D printed humanoid robot MARC for this experiments. The reasons behind using humanoid robot is that, research suggests, humanlike body of a humanoid robots help users to understand the robot's gestures intuitively (Kanda T, 2005). The reason could be that the actions of general gestures which evolved in our socio culture for human-human interactions allow also for intuitive human-robot interactions. In the experiments, MARC used common gestures and such gestures were designed from various studies (Gross M, 2010). MARC's voice was created using text-to-speech software and then edited using Audacity to make it more robotic voice.

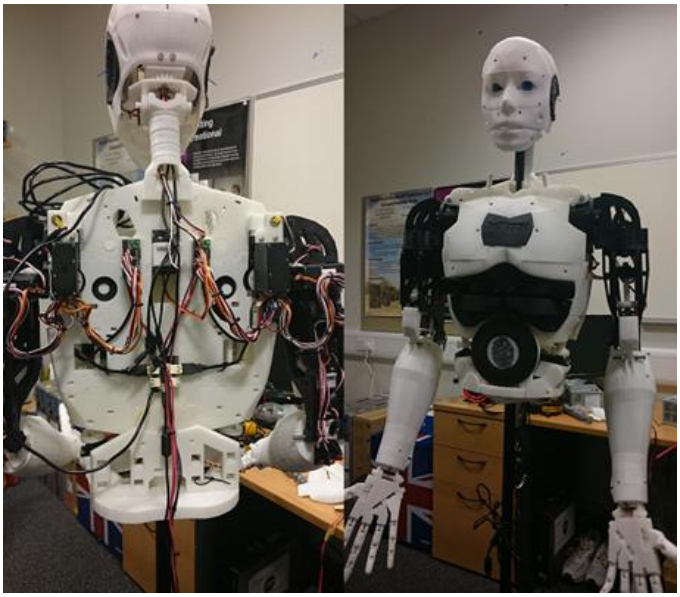


Figure. 1. The back and front of MARC the Humanoid Robot.

Selection of Cognitive biases

Five cognitive biases were selected in two experiments, such biases are, misattribution, empathy gap and Dunning-Kruger for the 1st experiments and, self-serving and humors effects biases are for the 2nd experiments.

Misattribution happens when someone remembers something accurately in part, but misattributes some details. This memory bias explains cases of unintentional plagiarism, in which a writer passes off some information as original when he or she actually read it somewhere before (Aronson E, 2005). In our experiments, the biased behavioural attributes were taken from misattribution are false memory, source of confusion and total forgetfulness.

Empathy gap is a cognitive bias which influences people to misunderstand the power of urges and feelings, such as, pain, hunger, fatigue (Nordgren, P, 2006, 2009), sexual arousal (Ariely, 2006), and cravings (Sayette, 2008) - on their behaviour. For example, when one is angry, it is difficult to understand what it is like for one to be happy, and vice versa. In our experiment, we tested two attributes of empathy gap bias, such as, hot state of empathy gap and cold state of empathy gap.

In 1999 psychologist David Dunning and Justin Kruger described the Dunning-Kruger effect as "…incompetent people do not recognize—scratch that, cannot recognize—just how incompetent they are," (Kruger, 1999). The Dunning-Kruger effect bias happens when an unskilled individual mistakenly suffers from illusory superiority and set their skill level much higher than actual. We tested three main behavioural characteristics of Dunning-Kruger bias, such as, unable to understand own lack of knowledge, unable to recognize other's true knowledge and, recognize and understand own lack of knowledge.

The self-serving bias is important for an individual to maintain and enhance their self-esteem, but in the process an individual tends to ascribe success to their own abilities and efforts, but failures to take into account external factors (Campbell & Sedikides, 1999). Individual's motivational processes and cognitive processes influence the self-serving bias (Shepperd, 2008). The general characteristics of such biased behaviours could be over confidence, lack of knowledge, ignorance and sometimes perusing. For example, in exams, students attribute earning good grades to themselves but blame poor grades on the teacher's poor teaching ability or other external causes.

Humors effects is a memory bias. People tends to remember humors incidents better than regular. This bias is common and could affect interpersonal relationship. In our experiments, for the humors effects algorithm, the robot make conversation with casual humors while playing games with the participants. The reason is to find out how such humors conversation helps to relate the participants with the robot and make long-term companionship. The current experiment uses specific components of the selected biases to make biased algorithms for the robot to interact with the participants.

The 1st Experiment – the conversational interactions based on misattribution, empathy gap and Dunning-Kruger biases effects

Methodology

The current experiments compare robot's three different biased behaviours with the robot's 'general communicative behaviours' in three different set of interactions. To do that, we develop the robot's general communicative behaviours without the effects of the selected biases. We call such general behaviours as the robot's 'baseline' interactive behaviours. As our study is in the field of friendly companion robot, so our 'baseline' algorithm is basically friendly simplistic version of the interaction which has made without the effects of the selected cognitive biases. For the biased algorithms, the components of the selected biases were developed based on the interactions of basic algorithm. Figure 2 shows the 1st experiment structure in details.

The set of biased interactions algorithms are to study the influences of the cognitive biases in the participant's preferences in the human-robot long-term interaction. To do that, the interaction dialogues were created based on the main components of the selected bias. For example, to study the influence of the misattribution bias the dialogues were developed based on the 'false memory', 'source confusion' and total forgetfulness. Similarly, the Dunning-Kruger bias was developed based on the main three components of the bias, such as, the 'robot fails to recognize own lack of knowledge', the 'robot fails to recognize genuine knowledge in others' and, 'the robot acknowledges its own lack of skill, after being exposed' (Lee C, 2012).

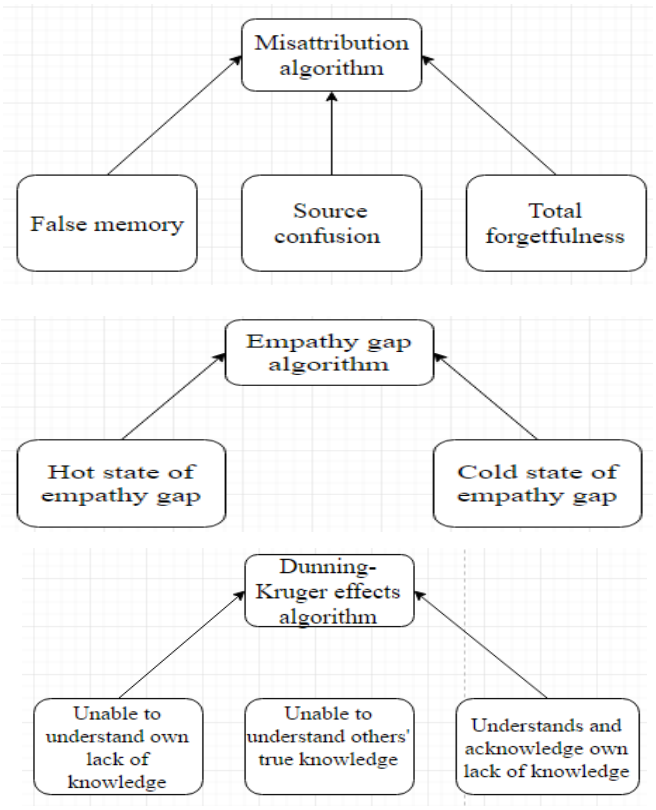


Figure. 2. Different biased algorithms in three interactions

Single Interaction Design

All the interactions were designed in three steps, such as, meeting and greeting, topic based conversation and farewell:

a.Meet and greet – this begins when participant enters in the room and goes up to the point when the robot finishes initial greetings.

b.Topics and conversation – this is the body of the interaction where the robot and participant discuss about various topics.

c.Farewell – this is the part where the robot says good bye to the participant and invites for the next interaction.

One of the main components of such interactions is the conversation. The conversation was designed based on question-answer. In the experiment, the dialogue design of the general conversation is based on four steps, such as:

a. Robot asks a question / says something

b. Participant responds

c. Robot states its own opinion

d. Robot waits for participant's responds / move to next dialogue

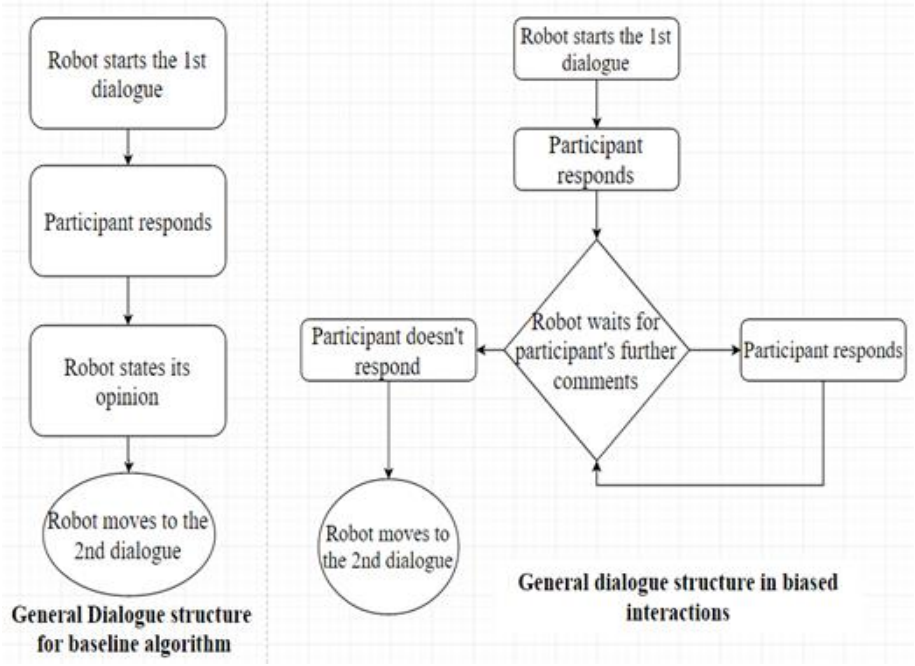


Figure 3. Differences in dialogue structures of baseline and biased interactions.

For example, MARC asks, "Do you like football?" The participant can respond as "yes" or "no", and also can extend their responses, but whatever participant's responses are, the robot would say something after that responses based on the algorithm developed. Then robot would wait for few moments to check if the participant wants to say something, otherwise it moves to the next dialogue. The differences in biased and baseline conversations are made in the step C, where the robot says something after the participant responds. In the baseline, the robot mainly

says ‘Okay’ or ‘That is great’ and move to the next topic, but in the biased interaction, robot’s dialogue reflects the bias effects, and the topic could continue further depends on the participant’s further responses. Figure 3 shows the dialogue structures and general differences between baseline and a biased algorithm.

The differences in biased and baseline algorithms were made in all the steps in the interactions. For example, in the 1st interaction’s meet and greet stage, there could be only three dialogues for the baseline algorithm, such as, (1) Hello, (2) My name is MARC, what is your name? and (3) Nice to meet you. But, in the case of the biased algorithm, the robot’s dialogues would be changed based on the bias, such as, for Empathy gap, the robot can be over joys or over sad to show the bias effects (hot-cold empathy). Therefore, the dialogues can be, (1) Hello my friend! I am very happy to see you today. It’s such a beautiful day. I hope you are feeling great today. (2) Hi. Today I am not feeling very good. Below we show two algorithms (a baseline algorithm and the 1st interaction misattributed algorithm) side by side as an example of differences in interactions. Figure 4 shows different stages of the interaction and the differences between baseline and biased algorithms in general. Such differences in dialogue structures and in algorithms developing made our algorithms different than each other and easy to understand their differences. Figure 4 represent basic differences between baseline and a biased algorithm. In that figure (Figure 4) false memory from the misattribution bias was used to develop in the robot. In this case, ‘false face’ represents, the robot misattributes participant’s face with someone others, and the robot does that for all the topics during the interaction.

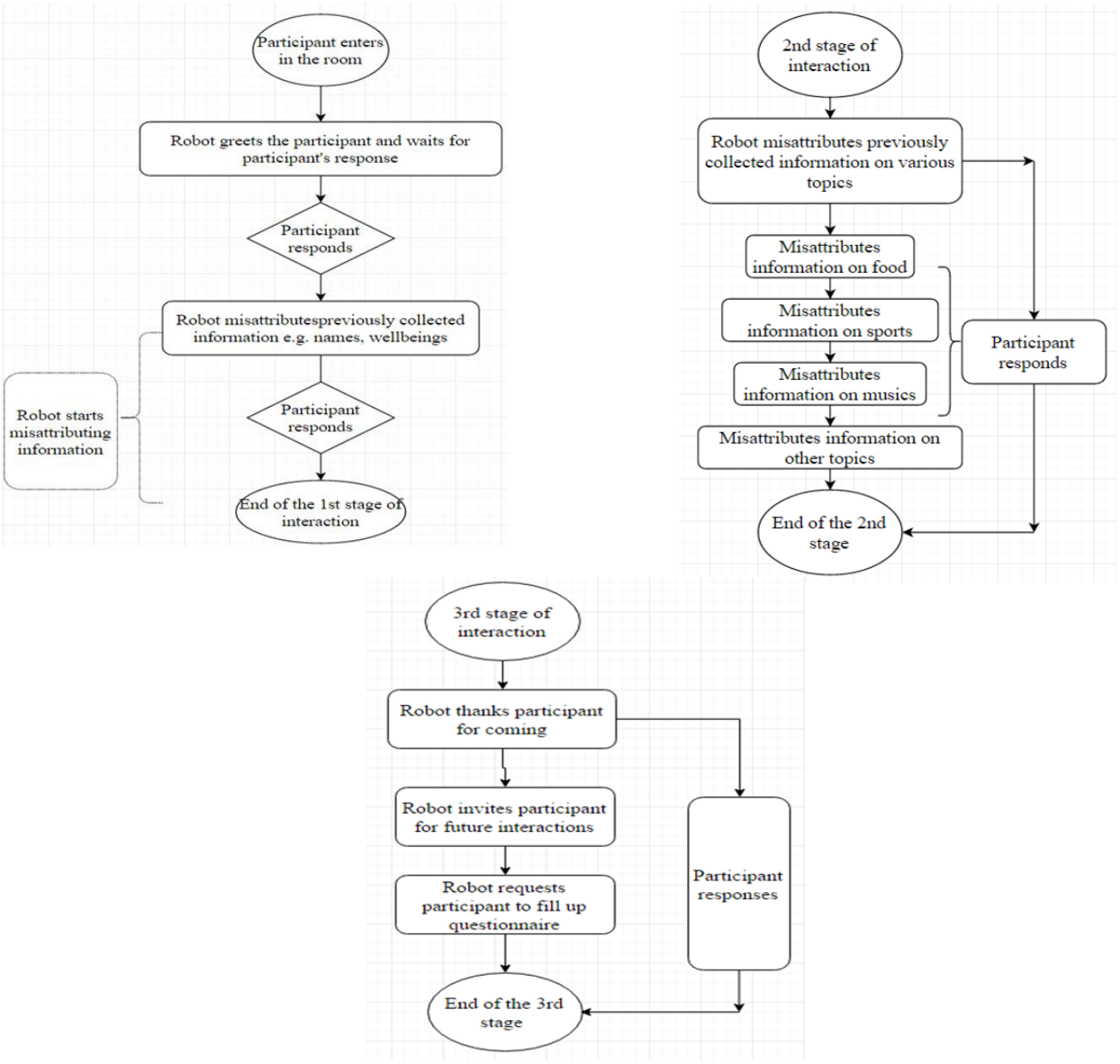


Figure 4. The diagram represents the three stages of the misattribution algorithms

Participants and grouping

Participants were invited by advertisements responses. From the all responses, we maintained a different age groups and 50-50 ratio of male and female. Participants were divided into two groups - in group A (Figure 5) there are 30 participants to interact with all four

algorithms in random order. In group B, 40 participants were selected to interact any of the individual algorithm. Therefore, in group B (Figure 6), for each algorithm there are 10 participants were selected. In each of the 10 participant's groups, the ratio of male and female was 50-50. The youngest participants were 15-year school girl and oldest was 55-year-old working male. In this case, the 10 participants are interacting with selected individual algorithm for three times. Such interactions should tell us the effects of each individual algorithm.

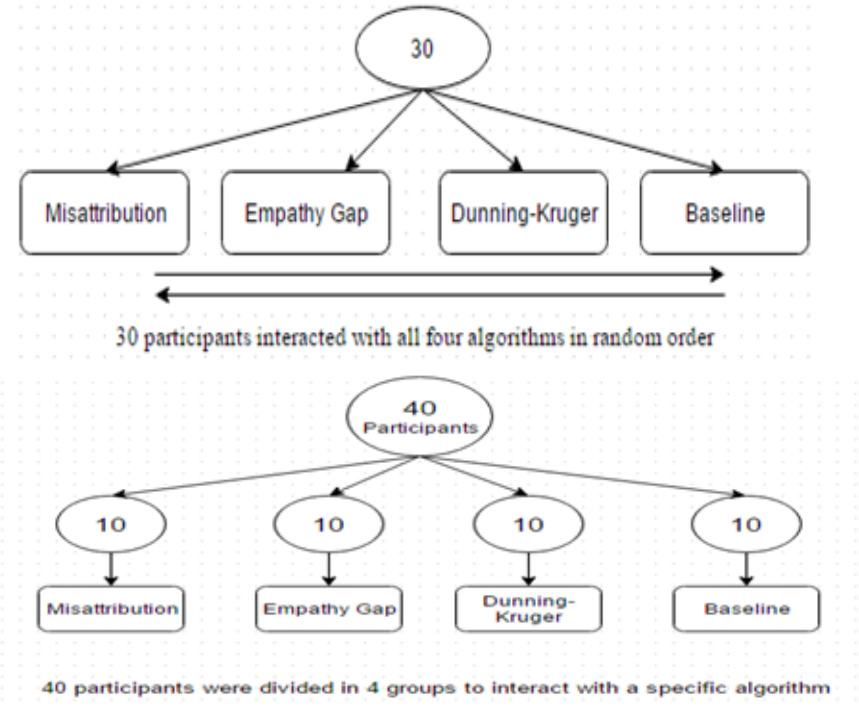


Figure 5. Group A - 30 participants interacted with all algorithms and Group B - 10 participants interacted with each of the algorithms

All the interactions were one to one basis, where each participant interacted with MARC individually for at least 8 to 10 minutes. There were total three interactions experiments between the participants and the robot, maintaining at least a week of gap between two interactions. The time interval was chosen to find out the effects of the algorithms in long period of time.

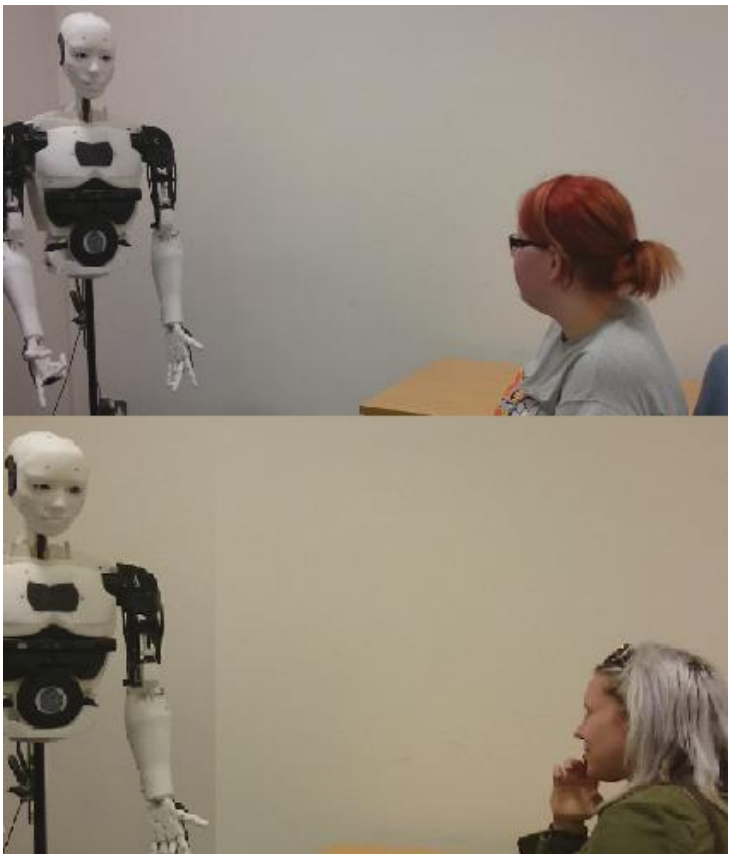


Figure 6. Participants are interacting with MARC

Data Collection

Participants were given a questionnaire after each of the interactions. The questionnaires were in 'Likert' method using a scale of '1' (least agreeablenesses) to '7' (most agreeableness). Such questionnaires were to find out the participant's likenesses and influences of a specific interaction algorithm. To do that, the questionnaires were designed based on several dimensions, such as, participant's experience likability (8 items) (Hone et al, 2000), comfort (6 items) (Hassenzahl, 2004) and rapport to the robot (15 items) (Multu B, 2006). Such dimensions were chosen to understand participant's closeness and involvements to the interactions, and also if they prefer biased algorithms over baseline. If the participants feel comfortable with the robot

and they like their experiences, then they should be involved in the interaction.The 3rd part of the questionnaires (Rapport) should tell us about their understanding to the algorithms. At the end of the final experiment, we took interview of each participants (Wyndol F, 2010).

Statistical Analysis

Data were analyzed in both groups based on the group formation. For the 1st group, as the 30 participants did all interactions, we ran one-way repeated ANOVA to compare and analyze the data. For the 2nd group, as each of the algorithms has 10 dedicated participants, we ran mixed ANOVA to analyze and compare data. The Cronbach's alpha (α) is calculated 0.916, which indicates high level of internal consistency for our scale.

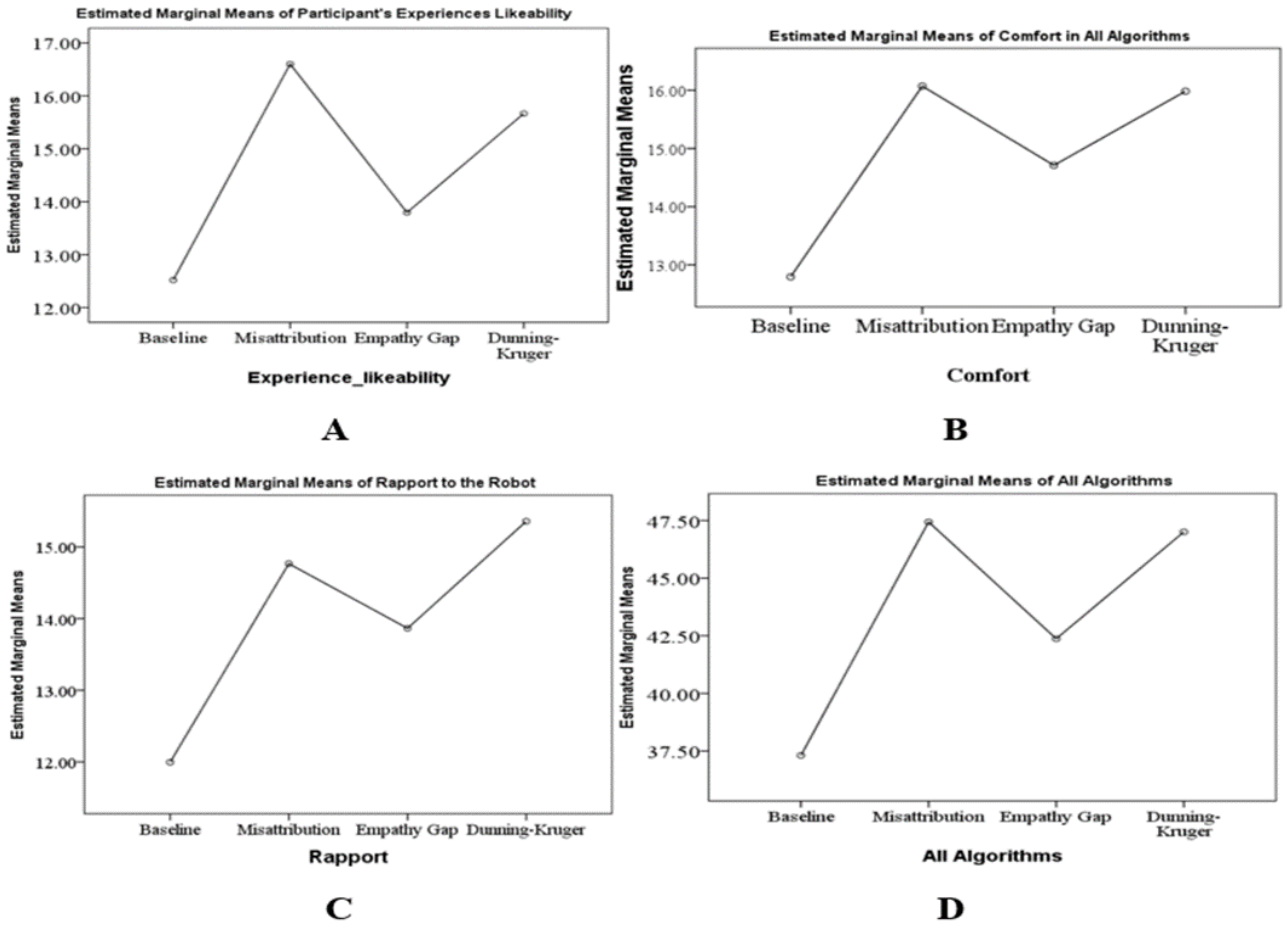
The 1st Group

Table 1 shows the Means from the algorithms in the 1st group. We calculated the total average ratings from all three interactions and compared using repeated measure ANOVA. The results are shown in the graph (Figure 7. A).

Table 1. Means of 3 dimensions

Algorithms		Means
Total Comfort ratings Average	Baseline	3.41
	Misattribution	5.42
	Empathy Gap	5.23
	Dunning-Kruger	5.58
Total Participant's experiences likeness ratings average	Baseline	4.10
	Misattribution	5.17
	Empathy Gap	5.12
	Dunning-Kruger	5.14
Total rapport to the robot ratings average	Baseline	3.75
	Misattribution	5.34
	Empathy Gap	4.57
	Dunning-Kruger	5.00

For the experience likability section, a repeated measure ANOVA were performed and the result shown in the graph (Figure 7. B). Similar to the other sections, for the 'rapport with robot' we calculated the total average ratings from all three interactions and compared using repeated measure ANOVA. The results are shown in the graph (Figure 7. C). Total Means graph has shown in Figure 7.D. The Means of each algorithms types were calculated by adding up all the ratings from participants. In the process Means were as, the Baseline the Mean is 37.31, where Misattribution approx. 47.43, Empathy gap approx. 42.37 and Dunning-Kruger approx. 47.0, - which means, in all the biased algorithms participants rated high in all three factors of the questionnaires. The lowest Mean difference is between Empathy Gap and baseline algorithms which is 5.06 (42.37 – 37.31) and the highest Mean difference is between Misattribution and baseline algorithms which is 10.12 (47.43 – 37.311). Such differences in Means indicate that the participants rated higher in biased algorithms (least 5 points to the highest 10 points) than baseline algorithms. However, there are differences in ratings in between the biased algorithms. In the graph (Figure 7. D), the Y axis is 'Estimated Marginal Means' and X axis shows the types of the algorithms. In all the pairwise comparisons, the Sig (p value) came out as <0.05 i.e. a very small probability of this result occurring by chance, under null hypothesis of no difference. So the null hypothesis is rejected, since $p<0.05$. So, there is strong evidence of participants preferring biased algorithms interactions over baseline interactions. Therefore, it can be said that the participants overall liked the biased algorithms interactions more than the baseline interaction. Based on the algorithms participants rated different in different dimensions. In the graph (Figure 7. D) it can be seen that each of the dimensions, participant's ratings were varied, but compared to baseline participants rated much higher in biased algorithms.



A – Shows the 'Comfort' dimension Means in different algorithms.
B - Shows the 'Participant's experiences likeability' dimension Means in different algorithms.
C - Shows the 'Rapport to the robot' dimension Means in different algorithms.
D - Shows Overall Means of the participant's ratings in all three dimensions in different algorithms.

Figure 7. The Mean graphs of the different dimensions and different algorithms for the 1st group

The 2nd Group

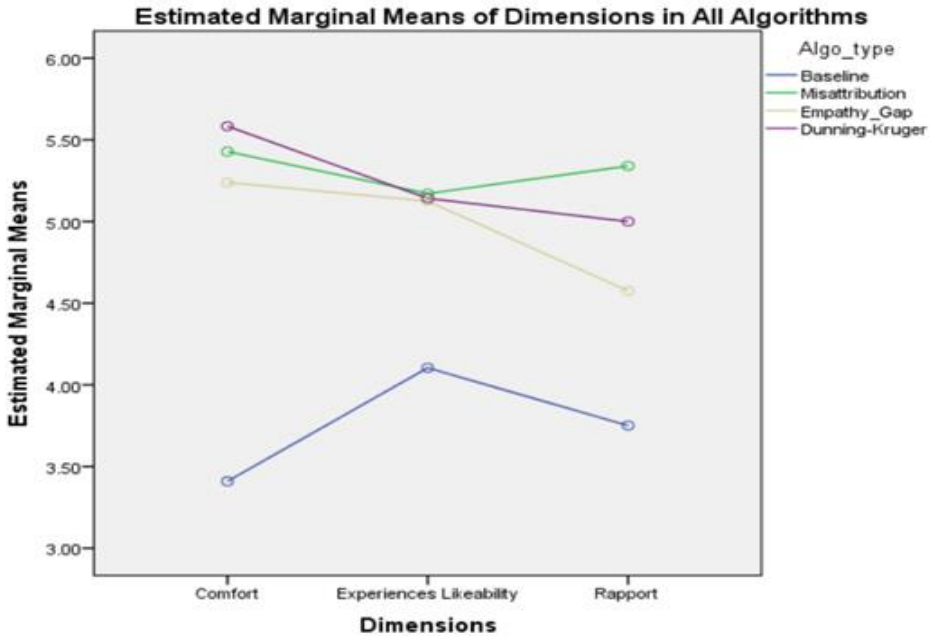
As stated earlier, a mixed (3x4) ANOVA was carried out on the dimensions (3) and algorithms (4). The Means from three dimensions were came as shown in the table (Figure 8).

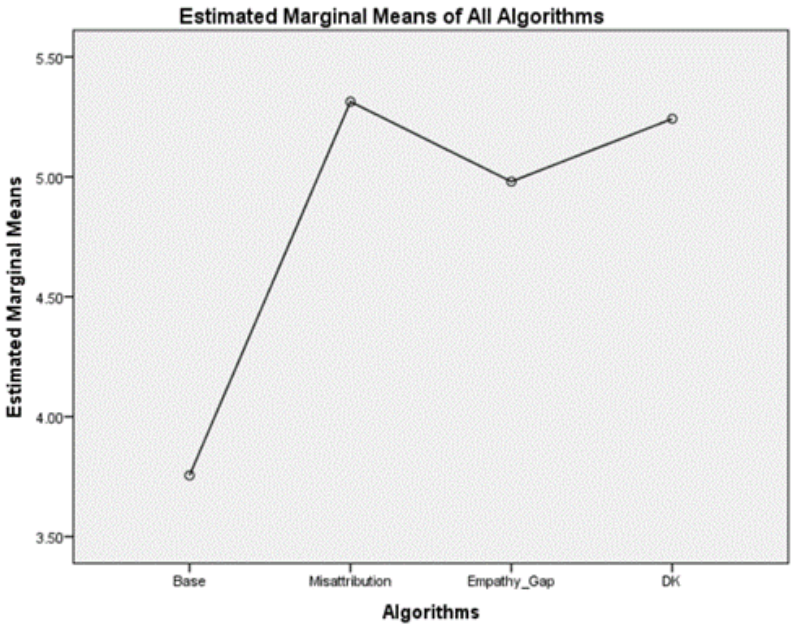
From the graphs it can be seen that the Means of each dimensions are higher for the biased algorithms than the baseline. There were stable positive increments in the ratings for each of the dimensions in all biased algorithms over the baseline algorithm. For example, the Comfort

Dimension Mean ratings has increased from 3.42 (baseline) to 5.58 (Dunning-Kruger), for Experience likeability dimension Mean ratings has increased from 4.1 (baseline) to 5.17 (misattribution) and, for Rapport dimension, Mean ratings has increased from 3.75 (baseline) to 5.34(misattribution), which are statically significant increases of 2.17, 1.07 and 1.59 (95% Confidence Interval, $p < 0.0005$). There was a statistically significant difference between means and therefore, we can reject the null hypothesis and accept the alternative hypothesis. The graph (Figure 8. A) shows plotting three dimensions in all algorithms. As clearly seen in the graph that the participants rated much higher for biased interactions than the baseline interaction.

The graph (Figure 8. B) shows the average Means plots from each algorithms in all the three experiments. The X-axis represents the algorithms and the Y-axis represents the marginal Means of each algorithm. The misattribution shows the highest point of the calculated Mean and baseline shows the lowest point of Mean. In fact, all biased algorithms Means are higher than the baseline. The graph was generated in the repeated measure test in SPSS using post-Hoc analysis. The graph (Figure 8) shows that the biased interactions were more popular than baseline interaction in this group.

Statistical analysis from the both group's data suggest that the biased algorithms were able to influence the participants to like the biased interactions more than the baseline.





Above: The above graph shows Means of the ratings based on 3 dimensions in all algorithms
Below: The above graph shows Means of the total ratings in all algorithms.
Figure 8. The Mean graphs of the different dimensions and different algorithms for 2nd Group

The 2nd Experiment – Game playing interactions based on humors and self-serving biases effects

Methodology

As same as the 1st experiment, the 2nd experiment compares robot's biased algorithms with 'baseline' behaviours. The 'baseline' algorithm was developed without the effects of the self-serving and humors effects cognitive biases. For example, in baseline behaviours, if the robot loses a game hand it simply says "You win" or "I lose", but in the self-serving algorithm robot tends to blame on the external factors and responses as "I was not ready" or "You are cheating". Such differences in dialogues are made in all conversational part of the interactions. On the other hand, in case of humors effects, robot makes fun of its own winning or losing.

Single Interaction Design

Theoretically the interaction was divided in five stages. Such stages were there for making clear differences between baseline and biased algorithms, so that, the baseline algorithm can be compared with the biased algorithm. The five stages were:

i. Meet and greet the participant – where participant meet with the robot and robot greets participant.

ii. Explaining the game rules – robot explains game rules

iii. Game playing – robot and participant start playing

iv. Game result – final results of the games

v. Farewell – where participant leaves

The robot may need to explain the rules, and there can be differences in dialogues based on the algorithms, therefore, we made additional ‘rule explain’ stage after initial greetings. Depending on the outcomes of single hand playing there could three cases, such as:

a. Robot wins - when robot wins a single hand.

b. Robot loses – when robot loses in a single hand play.

c. Draw – when both draw same hand.

The robot may need to explain the rules, and there can be differences in dialogues based on the algorithms, therefore, we made additional ‘rule explain’ stage after initial greetings. Depending on the outcomes of single hand playing there could three cases, such as:

a. Robot wins - when robot wins a single hand.

b. Robot loses – when robot loses in a single hand play.

c. Draw – when both draw same hand.

Based on such outcomes the robot response differently in both biased and baseline algorithms. The ‘game result’ is a state where the robot calculate and declare the winner. MARC’s dialogues would be different in this stage based on algorithms. For the self-serving algorithm, the robot praise itself, brags for winning, but blames others for losing. The robot motivates itself if it loses in all games of an interaction, and similarly, it influences it self-esteem if it wins all the game hands in an interaction. The game hands were drawing random, therefore, the outcomes could not be fixed. However, the experiments were designed to get the reactions from the participants in different situations of interactions. Therefore, the robot could lose in all games in all three interactions, or win it all, but finding out preferences of participants to an algorithm is the goal of the experiments. Figure 9 shows the differences of win/lose/draw situations in baseline and a biased interaction.

The core differences between the baseline and biased algorithms are in bias based conversation constructions, so that robot's responses could be biased. The baseline dialogues were brief, as the robot supposed not to say anything that reflect the bias effects. As seen in the diagram (Figure 3), the baseline conversation structure is straightforward which starts with the robot saying something or asking a question, then participant's response and, end with another statement by the robot. In between of two dialogues from the robot the participant can respond only one time. The 2nd dialogue from the robot usually comes as 'Okay', 'I understand' or a compliment, so that there is no open end for that particular conversation part, and the robot moves to the next dialogue. On the other hand, the biased dialogues are structured to take responses from the participant and to state the robot's own opinions. As discussed earlier in the self-serving bias, the robot blames external causes for losing a game hand. In our case, such external causes were such as the robot was not ready, the robot was looking other side, or something got into the robot's eyes. If the participant doesn't agree with the robot, it tries to convince the participant and challenges to play again. In such cases, the robot sometime blames the participants of cheating in games.

The differences between biased and baseline algorithms were made in all phases of the interactions. An example of 'game playing' phase has shown in the Figure 3. In this case, if the robot wins a game hand it says "I win" or "You win" for baseline interaction, but brags for win in self-serving interactions.

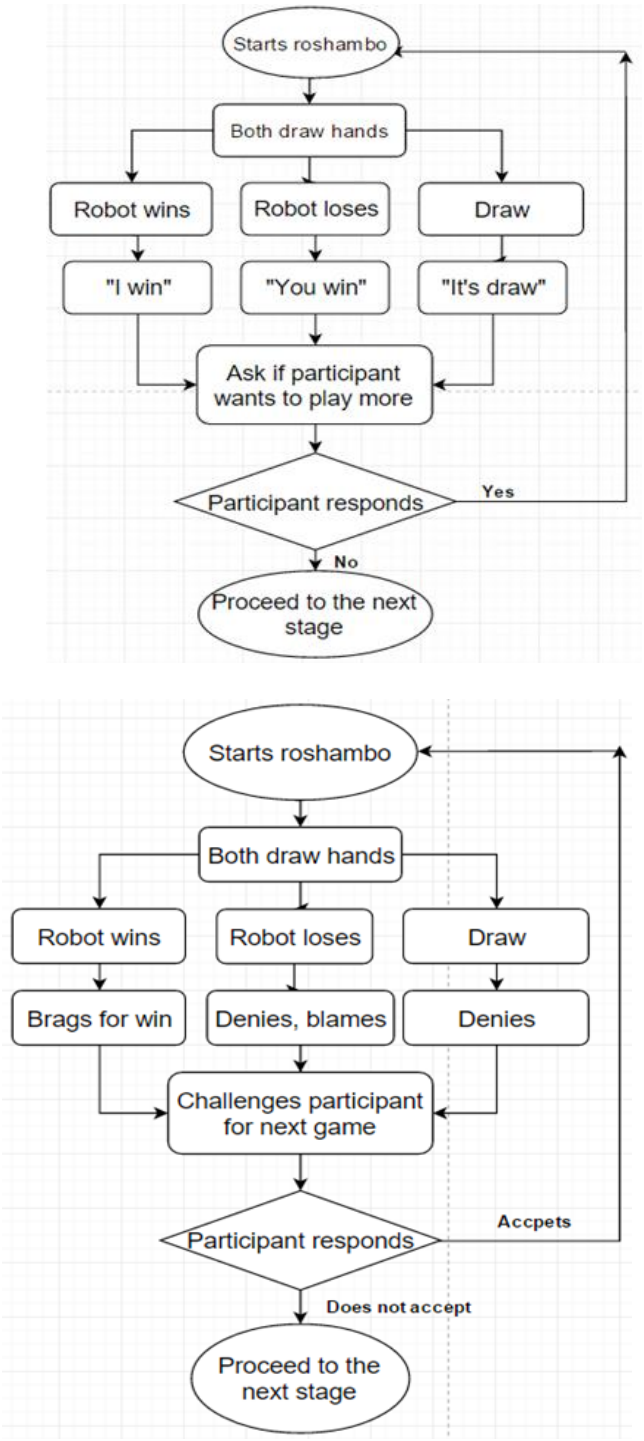


Figure 9. The game playing algorithm for both algorithms. Left side image represents the baseline, and the right side image represent the self-serving biased algorithm.

Participants and grouping

Participants were invited for three human-robot interactions by advertisements. 30 participants were selected to interact with any of the individual algorithms. Therefore, for each algorithm there are 15 participants (Figure 10). The gender and age groups ratio were balanced for both algorithms. There were three interactions in both algorithms maintaining at least a week interval between two interactions. Such interactions should tell us the effects of each individual algorithm in long period of time. Figure 4 shows the general experiment structure.

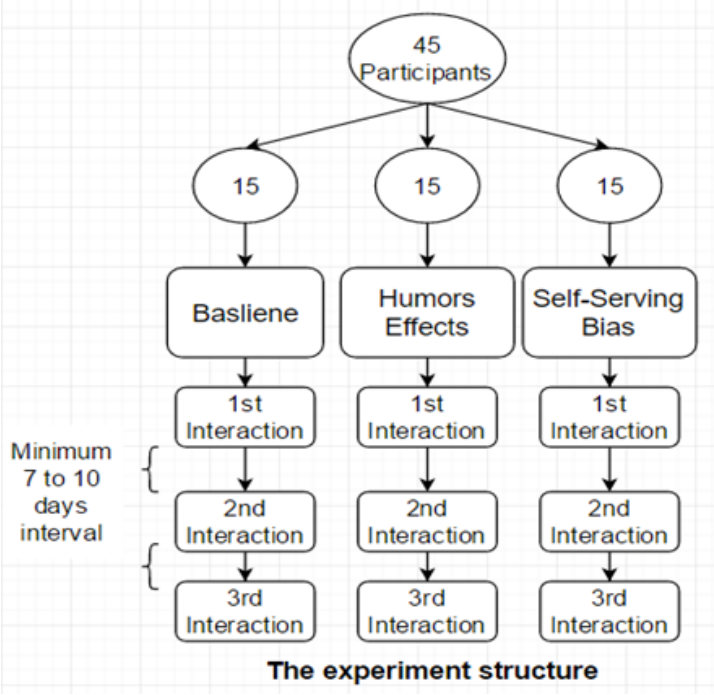


Figure 10. The experiment structure

All the interactions were one to one basis, where each participant interacted with MARC individually for at least 8 to 10 minutes.

Data Collection

In this experiment, we added another dimensions in our questionnaires, which is pleaser - how pleased participants were for the interaction. The other dimensions were the same as the 1st experiment.

Statistical Analysis

A mixed (4x2) ANOVA was carried out on the dimensions (4) and algorithms (2). Table 2 gives the Means from all algorithms in each of the dimensions. It shows that the Means of each dimensions are higher for the biased algorithms than the baseline. Among all the chosen factors, the self-serving biased algorithm scored higher than the baseline. There were stable positive increments in the ratings for each of the dimensions in all biased algorithms over the baseline algorithm. The Sig (p value) came out as <0.05 which indicate the significance our collected data over large population. There was a statistically significant difference between means and therefore, we can reject the null hypothesis and accept the alternative hypothesis. Figure 11 shows plotting four dimensions from two algorithms.

Table 2. Means of four dimensions

Algorithms		Means
Total Comfort ratings average	Baseline	4.35
	Humours effects	5.29
	Self-serving effects	5.73
Total Experiences Likeability ratings average	Baseline	4.82
	Humours effects	5.59
	Self-serving effects	5.58
Total Rapport ratings average	Baseline	3.95
	Humours effects	5.2
	Self-serving effects	4.9
Total Pleasure ratings average	Baseline	4.97
	Humours effects	5.37
	Self-serving effects	5.38

As clearly seen in the graph (Figure 11) that the participants rated much higher for biased interactions than the baseline interaction. Figure 12 shows the Average of Means ratings of the participants in the all 3 interactions. Figure 13 shows the overall Means plots from each algorithms in all the three experiments. The X-axis represents the algorithms and the Y-axis represents the marginal Means of two algorithms. As seen in the graph, the overall Means from baseline in much less than the self-serving. This graph can be called as the 'influence on participant' graph, as the graph represents the Mean ratings from all factors. The graph was generated in the repeated measure test in SPSS using post-Hoc analysis.

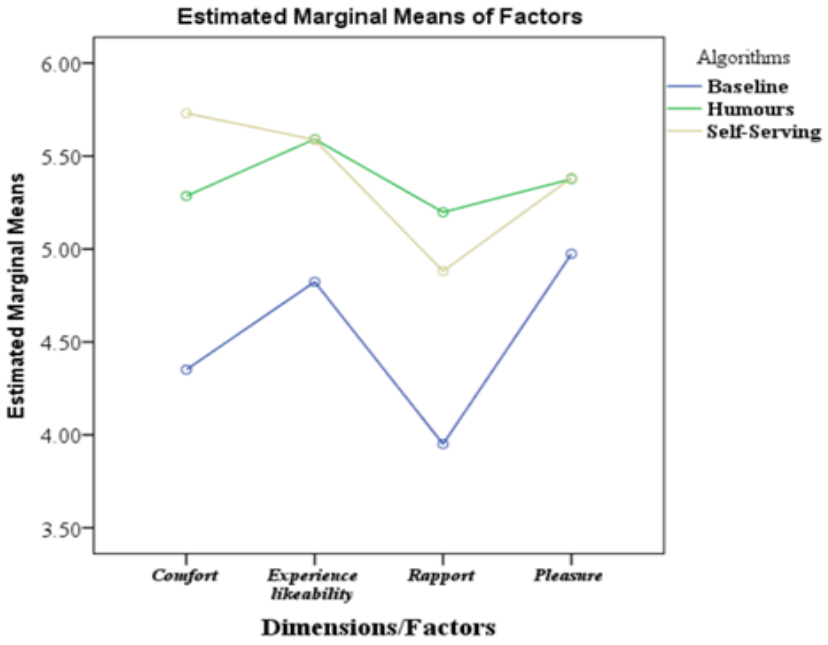


Figure11. The means of 4 dimensions in two algorithms

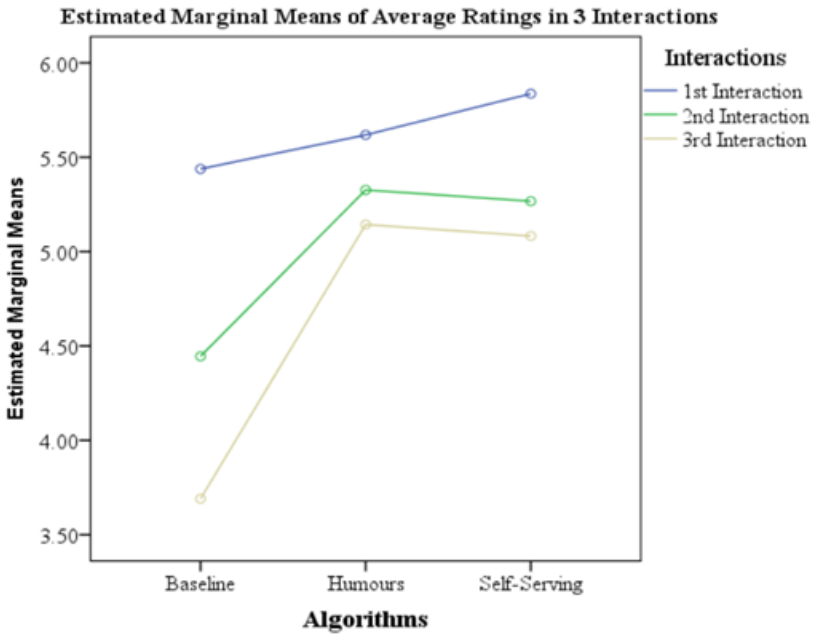


Figure 12. Means for 3 interactions in both algorithms

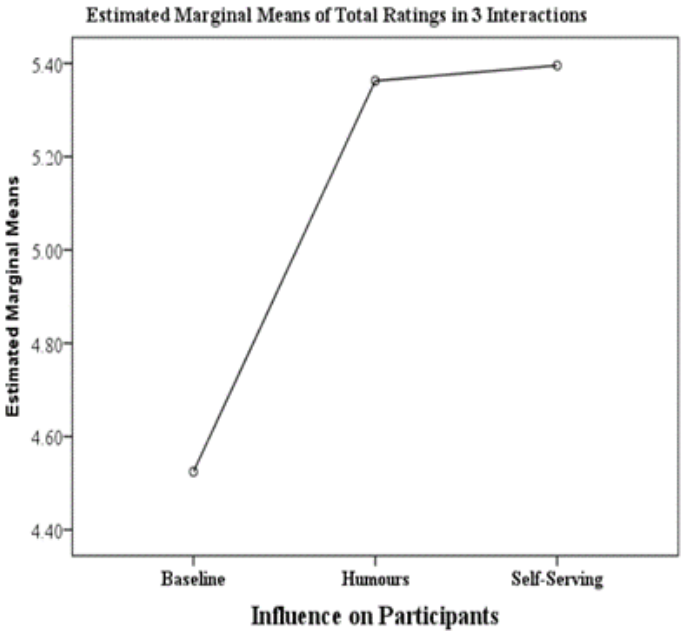


Figure 13. Influence on Participant graph – based on the total ratings in all interactions

Discussion

In the ‘comfort’ part of the questionnaires, there were six questions for the participants which were mainly to understand how easy and comfortable they feel with the robot in the different algorithms. For example, “Making conversation with the robot is comfortable for me”, “Making conversation with the robot is not difficult for me”, “Making conversation with the robot is not confusing for me” and similar. In the experiences likeability sections, questions were asked to find out how participant felt during the interactions. For example, “How much confident you felt during the interaction?”, “Will you visit for another conversation with the robot?” and others. The rapport with robot’s part of the questionnaires was asked to find out the overall likeness of the participant towards the robot, and how involvement the participant was in the interactions. For example, “Do you think that the robot and you feel very same about most things?”, “Would you choose to interact or communicate with the robot outside of this study?”, “Did you fell very close to the robot?” and others.

The statistics from the 1st experiment shows that MARC with bias algorithms is more likely to get participants attention therefore become more effective in influencing their likeability. Participants enjoyed their conversation and expressed their experiences and involvement in the questionnaires feedback. Participants in both groups rated much higher in all the biased interactions. Even for the different dimensions their ratings were higher in biased algorithms. Although, there are differences in ratings for different biases, but overall biased algorithms were most popular over the baseline algorithm. The reasons could be the differences in different biased robot’s behaviours. Participants found it very interesting that the robot could actually forget information, can make mistakes, brags about things which it doesn’t have any clue – all these behaviours are very common in humans. Participant’s reactions showed that they

were very surprised and enjoyed the fact that a robot could indeed have humanlike faulty characteristics where it sometime argues, doesn't understand what's going on and forgets. However, there were differences in likeness of humanlike behaviours and that's why the all ratings of biased interactions varied. For some of the participant's robot's forgetfulness behaviours was more enjoyable than its idiotic (Dunning-Kruger bias) behaviours, and some participants liked robot's empathy gap behaviours more than its forgetfulness, but overall, participants liked robot's humanlike fallible behaviours more than baseline robot's mechanical behaviours. In our experiments, we can see that the robot with baseline algorithms, developed a preliminary attachment with the participants, but robot with biased algorithms made the interactions more interesting to the participants. The participants felt more personable with the robot when the robot forgot and misattributed information, showed imperfect activities during interactions. At the end of the final interaction, participant was verbally interviewed by an examiner. To describe their experiences, participants told many moments of the interactions, few from those are "It was incredible to argue with a robot", "Robot has memory chip, but MARC forgot my name." and many more. Such type of biased behaviours participants liked in the biased algorithms. In case of the baseline algorithm, the robot did not argue or forgot anything, it responded as usual. Therefore, participant did not find the baseline algorithm as interesting as the biased. Even for the 2nd group, where a group of participants were dedicated to a single algorithm, biased algorithms participants rated higher than the baseline. The reasons behind differences of popularity in biased algorithms could be the differences in biases itself. In Empathy Gap, the robot responded overly in the 1st interaction, but remained less responsive in others. The data shows that such activities were not popular to some of the participants, rather they much enjoyed the robot being forgetful and robot bragging on wrong information.

In the 2nd experiment, overall statistical analysis shows very positive influence of the self-serving bias. In the graph (Figure 11) we can see that, in all the factors the biased robot scored higher than the baseline. The differences between Means of four factors (Table 2) are as, for comfort 1.38, for experience likeability 0.77, for rapport 0.97 and for pleasure it is 0.41. Self-serving bias scored high in all factors, but as seen, in the pleasure factor the difference in much less than others. To measure the 'pleasure' dimension, we added 8 items in questionnaire, some of those are, "Playing the game and having conversation with the robot is pleasurable to me", "Playing the game and having conversation with the robot is satisfying to me", "Playing the game and having conversation with the robot is enjoyable to me", "Playing the game and having conversation with the robot is entertaining to me" and similar. In their rating sheet, participants from the baseline interactions rated higher for first two questions (higher in 'pleased' and 'satisfaction') but lower for the other twos (lower in 'enjoyable' and 'entertainment'). On the other hand, the participants from self-serving interaction rated much higher for the last two questions (highly 'enjoyable' and 'entertainment'). In the comment section, some of the participants from commented that, it was very entertaining when the robot denies that it lost the game.

As it can be seen in the figure 12, in the 1st interaction there is very small difference in average Means of both algorithms. But in the 2nd and 3rd interactions, participant's ratings hugely dropped for baseline (21.75-14.76 = 6.99). On the other hand, self-serving ratings dropped in 2nd and 3rd interactions, but compared to baseline, such dropping was relatively smaller (23.34-20.33 = 3.01). In the interview, participants from self-serving group commented for the robot's excuses for losing a game as, at the beginning they thought MARC genuinely was not ready (or the Wizard), or they drew hand faster, but when MARC started making excuses

over and over then they found it very ‘interesting’ and also ‘entertaining’. In the 2nd and 3rd interactions, self-serving biased MARC accused them for cheating whenever it loses in games – in participant’s opinion they found it highly entertaining and liked it very much. To measure such bias effect, we added ‘comfort’ factor which had 6 questions in questionnaires, few of those are, “Playing the game and having conversation with the robot is uncomfortable for me”, “Playing the game and having conversation with the robot is uneasy to me”, “Playing the game and having conversation with the robot is difficult for me”, “Playing the game and having conversation with the robot is confusing to me” and similar. But surprisingly, participants in biased interactions did not find MARC’s such behaviours as uncomfortable, uneasy or very difficult for them. As they commented, they were surprised to be accused of ‘cheating’ from a robot. Participants also mentioned, it is very common human behaviour not to accept losing, and the robot acting same like their friends. Moreover, they found out MARC’s bragging behaviours after winning a game is hilarious. On the other side, the participants from baseline interactions did not find any of such humanlike behaviours from MARC, and to them its behaviours were ‘as common as a robot’. In their interviews and comments, they pointed out that, playing game with a robot was enjoyable but it went less enjoyable after few times. Even the robot was drawing random hands, but the participants found MARC’s responses are ‘stereotype’, ‘very mechanical’ and ‘common as robot-like’ in the baseline.

The figure 13 shows an overall Means difference between baseline and self-serving algorithms. As it can be seen that the baseline Mean is very smaller than the self-serving. From this graph it can be said that participants in biased interactions were much influenced by the robot’s biased and imperfect behaviours, and rated higher than the baseline. But, participants in the unbiased group did not find any of such human-like behaviours in their interactions.

Therefore, to them the robot's behaviours were mechanical and as usual like a robot. In the 1st interactions, both groups participants enjoyed the game and rated high, but in the later interactions, MARC continued to show biased humanlike behaviours in biased interactions, so that the participants found it interesting and rated very similar as the first interaction. But, the robot with baseline algorithm failed to show such humanlike behaviours in later interactions, and so participants found their interactions as mechanical and, the ratings dropped higher than the biased interactions.

Therefore, it can be concluded that cognitive biases and humanlike imperfectness are able to develop better interactions than a robot without humanlike imperfectness in its interactive behaviours. All three biased interactions received more popularity and gained more positive responses from the participants. The participants liked the robot's behaviours in different situations in games, such as, winning, losing and draw – that the robot brags about a win but blames on the participants or the external causes for losing and make draw, but despite of that the robot behaves very friendly – greets them, bid them farewell and requested for coming next times. Such kind of behaviours are very common in people, between close friends. In friendships, close friends could be very competitive in game playing and do not want to lose easily. Such types of behaviours are common human nature which we do and see in our daily life. When participants found out the same behaviours from our imperfect robot MARC they might found it easy to relate with it, and that might be the reason for biased and imperfect algorithm getting higher ratings than baseline. On the other hand, baseline MARC did not show any humanlike common behaviour rather than very generic impressions - which might be expected from a robot to our participants, and that could be the reason of the differences in ratings between biased and baseline algorithms. However, from the experiments and analysis of

collected data it can be concluded that, the humanlike biases and imperfectness in robot's interactive behaviours can enhance its abilities of companionship with its users over a robot without biases. In the figure 9 we see different interactions moments from the two experiments.

From the above discussion, the answer of the research questions can be found. All the experimental results suggest that statistically participants rated higher for their biased version of the interactions in all the experiments. Therefore, it can be said that the cognitive biases can play an important role to help the robot to engage in interactions with the participants for short-term and long-term interaction. Participants like this engagement with the biased version of the robot and therefore rated higher compared to the non-biased interactions. It can be said that cognitive biases in robots can help to develop long-term interactions significantly than the robot without biases. By introducing cognitive biases in robot MARC, we have seen significant difference in participant's likeness for all of the biased interactions, therefore, it can be said that the biases influence the human-robot interactions positively than the robot without such biases.

Conclusion

In general, there is a conflict between the people's perceptions from literature, science fiction movies and the goal of the HRI researches (Sandoval E, 2014). Our experiments show that, cognitive biases can be useful to reduce that conflict by making the robots cognitively imperfect (Biswas M, 2013). We expect that, these interaction experiments can be helpful to understand the necessity of using cognitive biases and humanlike imperfectness in robots for long-term companionship. Also, using of different biases and imperfectness can be helpful to

understand the difference in the effects of different biases in human-robot interactions. In our research, ‘cognitive imperfection’ refers to a robot which shows cognitive biased behaviours with humanlike fallible characteristics, such as, mistake making, wrong assumption, task imperfectness and others which are common in humans. In our understandings, cognitive imperfectness can make the robot humanlike cognitive which can help users to relate with the robot. By comparing data among all experiments, it can be said that humanlike imperfect fallible behaviours in robots helps to make robot-human interactions more enjoyable to the participants. As results, participant makes a preferred relation faster with a biased robot than unbiased robot. Our experiments show that long-term companionship can be possible between humans and robots with humanlike imperfect behaviours. In human psychological nature, it is easy to interact with another human-like personality that shows typical social characteristics (e.g. pet animals) (Meerbeek B, 2009). It is difficult to have a relationship with something that is too different or novel to us that pretends to be too perfect without having any mistakes or faults, unlike humans. The same is true even for humans. Some people with different emotional responses are sometime hard to understand. Robots in the other hand have abilities to perform human-like actions, can be designed to 'look' like humans and can appear to behave in a human like manner, but they lack human-like cognitive personalities. Aristotle (384-322BCE), Immanuel Kant (1724-1804) have argued that human characters and personality can be described as imperfectly perfect (Stanford Encyclopedia, 2008), where robots lack to present such type of imperfection in their cognitive behaviours, like, forgetfulness, unintentional mistakes, wrong assumptions, extreme presence of specific traits, task imperfectness and other human-like cognitive characteristics.

Interrelations grow from the attractions of differences in characters, unpredictability and cognitive difference and imperfectness of nature. We expect, if it’s possible to make the robot’s

cognitive behaviours human-like and fallible then it might be possible for robots to gain such type of attentions from humans that can create strong attachment for long-term companionship. In our understandings, imitation of humanlike cognitive actions does not just refer to programming a robot to tell a joke like humans, but we also want to find out, if the robot tells a joke poorly then what kind of impact that creates.

In our experiments, such human like behaviours using different cognitive biases were successful to create initial attachment bond with the participants. In further research, we want to include traits activities, emotions and mood with humanlike imperfect behaviours and different cognitive biases in robots to express various cognitive imperfectness, such as, mistakes, wrong assumptions, expressing tiredness, boredom or overexcitement amongst other humanlike common characteristics. We expect if robot can show in their behaviours as being similar to humans, then the robots could possibly be accepted to the majority of our society.

REFERENCES

1. Baroni, I. et al (2014). What a Robotic Companion Could Do for a Diabetic Child. In 23rd IEEE International Conference on Robot and Human Interactive Communication (RoMAN 2014).

2.Bless, H., Fiedler, K., & Strack, F. (2004). “Social cognition: How individuals construct social reality.” Hove and New York: Psychology Press

3.Bernt Mk, Martin S, Christoph B,“Iterative design process for robots with personality”, (2009). Iterative Design Process for Robots with Personality. Proceedings of the AISB2009 Symposium on New Frontiers in Human-Robot Interaction Edingburgh pp. 94-101.

4.Biswas M, Murray J, 2014, “Effect of cognitive biases on human-robot interaction: A case study of a robot's misattribution”, Robot and Human Interactive Communication, RO-MAN: The 23rd IEEE International Symposium, pp. 1024-1029.

5. Boston Dynamics, “Petman”, bostondynamics.com/ robot_petman.html, retrived 08/06/15

6. Bernt M et al, “Iterative design process for robots with personality”, AISB2009 Symposium on New Frontiers in Human-Robot Interaction, 2009, 94-101

7. Kahneman, D.; Tversky, A. (1972). "Subjective probability: A judgment of representativeness". Cognitive Psychology 3 (3): 430–454.

8.Kanda, T. et al, “Analysis of humanoid appearances in human-robot interaction”, ATR Intelligent Robotics & Commun. Labs, Kyoto, Japan, 2005

9. Kerstin D et al, KASPAR – A Minimally Expressive Humanoid Robot for HumanRobot Interaction Research, Applied Bionics and Biomechanics, 2009

10.Gary F. Rl, “Interactive Perceptual Psychology: The Human Psychology That Mirrors The Naturalness Of Human Behaviour”, Mid-Western Educational Research Association, Annual Conference, October 1994.

11. Gross, M. Elizabeth A. Crane, and Barbara L. Fredrickson. 2010. “Methodology for Assessing Bodily Expression of Emotion.” Journal of Nonverbal Behavior 34 (4) (July 31): 223–248. doi:10.1007/s10919-010-0094-x

12. Haselton, M. G., Nettle, D., & Andrews, P. W. (2005).The evolution of cognitive bias. In D. M. Buss (Ed.), The Handbook of Evolutionary Psychology: Hoboken, NJ, US: John Wiley & Sons Inc. pp. 724–746.

13. John R L, “Forgetful or Bad Memory?”, Proceedings of the 32nd Hawaii International Conference on System Sciences – 1999.

14. Kahneman, D. & Tversky, A. (1972). "Subjective probability: A judgment of representativeness". Cognitive Psychology 3 (3): 430–454.

15. Langer, E. J., 1997. The power of mindful learning, Addison-Wesley, Reading, MA.

16.Levinger, G. (1983). Development and change. In H.H. Kelley, et al. (Eds.), Close relationships. (pp. 315–359). New York: W.H. Freeman and Company.

17.Kiesler, D.J., Schmidt, J.A. & Wagner, C.C. (1997). A circumplex inventory of impact messages: An operational bridge between emotional and interpersonal behavior. In R. Plutchik & H.R. Conte (Eds.), Circumplex models of personality and emotions (pp. 221–244). Washington, DC: American Psychological Associatio

18.Lee K M et al, “Can Robots Manifest Personality? An Empirical Test of Personality Recognition”, Social Responses, and Social Presence in Human–Robot Interaction, Journal of Communication, 2006, ISSN 0021-9916

19. Lilia M et al, “Time-varying affective response for humanoid robots, Progress in Robotics”, Communications in Computer and Information Science Volume 44, 2009, pp 1-9

20. Larry L., Fergus I C, Effects of Elaborating of Processing at Encoding and Retrieval: Trace Distinctiveness and recovery of Initial Context, Effects of Elaboration and Processing. Manuscript, 1975

21.Melton, A.W.; Lackum, W. J. von (1941). "Retroactive and proactive inhibition in retention: evidence for a two-factor theory of retroactive inhibition". American Journal of Psychology 54: 157–173. JSTOR 1416789

22.Michael L. W et al, “Avoiding the Uncanny Valley – Robot Appearance, Personality and Consistency of Behavior in an Attention-Seeking Home Scenario for a Robot”, Autonomous Robots Journal, Volume 24 Issue 2, February 2008, Pages 159 – 178

23.Michael T, The Human Adaption of Culture, Vol. 28: 509-529, DOI: 10.1146/annurev.anthro.28.1.509

24.Mondesire, S., & Weigand, P. (n.d.). Forgetting Classification and Measurement for Decomposition-based Reinforcement Learning. Retrieved May 5, 2014, fromhttp://weblidi.info.unlp.edu.ar/WorldComp2013-Mirror/p2013/ICA3556.pdf

25.Murray, J, (2008) et al, MIRA: a learning multimodal interactive robot agent. In: Hybrid Intelligent Systems, 2008, 10-12 September 2008, Barcelona, Spain.

26. Paul B et al, “Long-Term Human-Robot Interaction with Young Users”, in Proceedings of the IEEE/ACM HRI-2011 workshop on Robots interacting with children, Lausanne, 2011.

27. Paul C. (2008) "Seven Types of Forgetting".pg. 59-71

28. Sadler, P., & Woody, E. (2003). Is who you are who you’re talking to? Interpersonal style and complementarity in mixed-sex interactions. Journal of Personality and Social Psychology, 84, 80-96

29.Sangha, S et al (2005). Impairing forgetting by preventing new learning and memory. Behavioral Neuroscience, 119(3), 787-796.

30.Nomura T, T. Kanda, and T. Suzuki, “Experimental investigation into influence of negative attitudes toward robots on human-robot interaction,” AI & SOCIETY, vol. 20, no. 2, pp. 138–150, 2006

31.Volkmann, Kelsey (28 April 2011). "Honda's ASIMO visits FIRST robotics event". St. Louis Business Journal. Retrieved 9 August 2011.

32.Wilke A. and Mata R., Cognitive Bias, The Encyclopedia of Human Behavior, vol. 1, pp. 531-535. Academic Press. 2012