The Effects of Cognitive Biases and Imperfectness in Long-term Robot-Human Interactions: Case Studies using Five Cognitive Biases on Three Robots

Biswas M, Murray J

University of Lincoln, School of Computer Science, Lincoln, UK.

Abstract:

The research presented in this paper demonstrates a model for aiding human-robot companionship based on the principle of 'human' cognitive biases applied to a robot. The aim of this work is to study how cognitive biases can affect human-robot companionship in long-time. In the current paper, we show comparative results of the experiments using five biased algorithms in three different robots such as ERWIN, MyKeepon and MARC. The results were analysed to determine what difference if any of biased vs unbiased interaction has on the interaction with the robot and if the participants were able to form any kind of 'preference' toward the different algorithms. The experimental presented show that the participants have more of a preference towards the biased algorithm interactions than the robot without the bias.

Keywords:

Human-Robot Interaction, Human-Robot Long-term Interactions, Humanoid robot, Cognitive Bias, Imperfect Robots

1. Introduction

It is evident that social behaviour is an important factor in human-human, then we can be safe to assume that such interactions are important in social cognition behaviours in social robots during robot-human interactions. Mahani and Eklundh (2009) suggest that, "If through long-term use these [service] robots gain social skills, they could be supportive of some social roles that people might assign to them". To develop such social intelligence, researchers have studied various methods for robots to adapt to human-like behaviour based social roles. Few of the most popular methods suggest developing human-like attributes in robots, such as, trait based personality attributes, gesture and emotions expressions and anthropomorphism.

Walters et al (2008) 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. Research suggests that developing cognitive personality trait attributes in robots can make them more acceptable to humans (Lee K, 2006). In addition to this, expressing emotions and mood changes in interactions can help to make the attachment bond stronger between a human and the robot. Meerbeek et al (2009) designed an interactive personality process in robots which was based on Duffy's (2003) anthropomorphism idea. Indeed, Duffy suggests 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) have shown that users will demonstrate certain biased driven personality traits to machines (e.g. Computers) and from that research they propose a 'user driven' mental model for domestic robots. Walters et al (2008) investigated people's perceptions of different robot appearances and associated attention-seeking features in video-based Human-Robot interaction trials. Their study revealed participant's preferences for various features of the robot's appearance and behaviour with their personality attributions towards the robots being comparatively similar to their own personalities. The above studies demonstrate approaches to making a robot more humanlike and thereby more intuitive for people to interact with. It is important to consider that humans have for millennia, interacted with other humans and as such our interactions and social norms are reflective of our own personalities and behaviours. It is therefore only natural that if we wish for humans to engage and interact with robots, that these robots not only understand human social constructs, but also display these traits. The research presented in this paper investigates an approach to developing socially interactive robots by applying selected cognitive biases with the aim to providing a more humanlike interaction.

Cognitive biases play a large part in influencing a human's characteristics and behaviours (A Wilke, 2012). Human personalities are considered unique but based on a set of different social behaviours, social norms and cultures (Haselton, 2005). Kahneman (1972) suggests that human thinking can be affected by a variety of biases which can influence a human into making wrong decisions, bad judgments and other fallible actions, after all we're only human!

Such differences in cognitive imperfectness among individuals hugely affects that individual's interactions, making them unique, natural and human-like. Making faults and misjudgments are common human characteristics. But in developing humanlike robots, we sometime ignore such facts and attempt to make robots as faultless as possible, with perfect memory recall and repeatable actions, that is, we make them less humanlike. Such cognitive imperfections (e.g. forgetfulness, making mistakes) have has yet not been fully explored in social robots for the purpose of developing a human-robot companionship. In the current research described in this paper we approach to find out the influences of cognitive biases in human-robot interactions by developing five cognitive biases (misattribution, empathy gap, Dunning-Kruger effects, self-serving and humours effects) in three different robots (ERWIN, MyKeepon and MARC see Figures 2,8 &13). The biases were developed individually and, based on the main attributes of such biases. To compare the biased interactions there was non-biased interactions were developed as well which was made free from the selected bias effects.

2. The Project: Cognitive Bias in Human-Robot Interaction

Cognitive biases are often a result of an attempt to simplify information processing which can help to make sense of the world and reach decisions with relative speed (Bless, 2004). Sometimes, these biases lead to poor decisions and bad judgments, but in other situations, those judgemental choices can be useful. Biases refer to a systematic pattern of deviation from rationality in judgement, whereby inference about other people and situations might be drawn in illogical fashion (Haselton, 2005). In a given situation however, biases can sometimes lead to a more effective set of actions (Gigerenzer, 1996). For example, if the given context demands immediate action over accuracy, heuristic biases enable the taking of decisions faster (Tversky and Kahneman, 1974). Cognitive biases can arise from various processes that are sometimes difficult to distinguish, such as, social influence (Wang, 2001), information processing shortcuts, mental noises (Hilbert, 2012), limited brain capacity of information processing (Simon, 1955; Marois, 2005) and emotional and moral motivation (Pfister, 2008).

Bless et al., (2004) suggested that cognitive biases can influence a human's behaviour towards positive or negative ways. Biases can effect individual's decision making (Tversky and Kahneman, 1974), behaviours (Brand, 1985) and social beliefs (Huijbregts, 2007). It is understood that such cognitive biases among other factors (e.g. mood, emotions, traits) effect on the individual's differences in characteristics behaviours. Society is an example of each person being different in behaviour and each has got their very own unique characteristics. In our understanding, such differences in cognitive characteristics among individuals are what make human interactions unique, natural and human-like. In existing social robotics, robots are now able to imitate different human behaviours, for example, eye-gazing, making gestures while talking, expressing emotions and others. But in human-human interactions, individual's own characteristics biases (e.g. forgetfulness, empathic gap, self-serving, humours effects etc.) are present which are 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 unable to show such human-like cognitively biased behaviours (e.g. forgetfulness, unable to understand correct emotions, bragging, blaming, remembering humours events etc.). Recent studies have focused humanlike faulty behaviours to develop in robot to find out their effects in human-robot interaction. Salem et al (2015) studied on how the perception of erroneous robot behaviour influences human interaction choices and the willingness to cooperate with the robot. Robinette et al (2016) studied of faulty behaviours in robots and 'over trust' of participants which shows that even in an emergency situation participants trusted a faulty robot. However, the effects of different cognitive biases are not explored in greater details in robots for human-robot interactions.

The research presented in this paper focuses on the main components of five cognitive biases, such as, misattribution, empathy gap, Dunning-Kruger effects, humours effects and self-serving effects. We hope that human-like cognitive imperfection can result a model which will allow for more attachment and companionship between humans and robots.

Figure 1 explains how we develop the biased algorithm in robots in general. Diagram 1 in Figure 1 shows how in case of the robot with non-biased algorithm interact with people and, In the 2nd diagram we apply our biased algorithm. In the above figures, we see that we apply biased algorithms in robot's functions and features, so that their functionality could be biased. The non-biased algorithm, however, does not change the robot's any of their functions and features, so that, non-biased interaction stays such biases free. In our experiments, participants interact with the both biased and unbiased robots to compare the differences in the robot's behaviours.

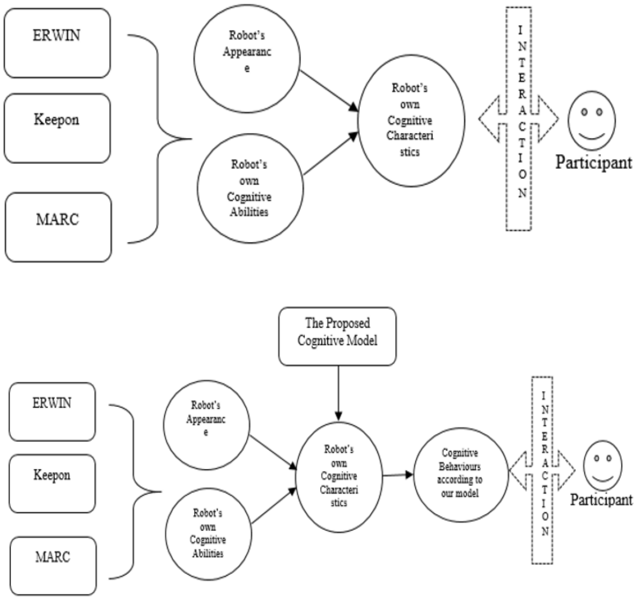

ERWIN
Keepon
MARC
Robot's Appearanc e
Robot's own Cognitive Abilities
Robot's own Cognitive Characteri stics
INTERACTION
Participant
The Proposed Cognitive Model
ERWIN
Keepon
MARC
Robot's Appearanc e
Robot's own Cognitive Abilities
Robot's own Cognitive Characteri stics
Cognitive Behaviours according to our model
INTERACTION
Participant

Fig. 1. Simplified diagrams of implementing proposed cognitive biases algorithm in robots: 1st diagram without biased algorithm and, 2nd diagram, applying biased algorithm

3. Hypothesis

The research described in this paper studies on developing a new approach in human-robot interactions which is based on cognitive bias introduced in robot. Thus the research seeks answer to the following hypothesis:

"*Can the introduction of cognitive biases in a robot influence Human-Robot Interaction and influence user preference?"*

The main hypothesis addresses three key challenges, such as:

- Development of cognitive biased behaviours in robots which could demonstrate the bias's behaviours properly,
- Study the cognitive biased behaviours in a robot to find out if that can influence human-robot interaction, and,
- Study the effects of such biased behaviours in human-robot interactions for long-period of time to find out the changes of the influences.

Based on such challenges, additionally we seek answers to the three research questions. The research questions are as follows:

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 cognitively biased behaviours?
2. Introducing cognitive biases to the robot's interactive behaviours is it possible to develop human-like biased behaviours in robots which can influence human-robot interaction?
3. Will cognitive biases help humans to relate to the robot in long-term interval interactions?

To get the answer for the above hypothesis and research questions, several different human-robot experiments were done (e.g. conversational and game playing). In such experiments we developed algorithms based on cognitive biased behaviours in robots and used in interactions with the participants. Parallel we experiment with another algorithm which is without the effects of the selected cognitive bias. At the end we compare participant's feedback from the experiments for biased and non-biased algorithms to find out which interaction participant preferred the most.

4. Selection of Cognitive biases

In our experiments, we have chosen five cognitive biases to test with three different robots in long-term scenarios. Biases were chosen based on few principles:

- The biases must need to be widely common in humans.
- The biases should have a minimum impact on interpersonal relations in our daily life.
- The biases must have clear impact on individuals so that others can recognise the effects.
- The biases can be developed on robots and can be experimented in HRI experiments.

Based on the above principles, we selected biases such as misattribution, empathy gap, Dunning-Kruger, humours effect and self-serving biases to develop in our robots.

Misattribution: Misattribution is the making of an incorrect attribution. This happens when people wrongly attribute an event to something else that truly does not have a connection or association to said event. Misattributions can be specified into two categories, such as Misattribution of Arousal and Misattribution of Memory.

Misattribution of Arousal is a psychological situation where people make a mistake in assuming what is causing them to feel in-text citations (White et al., 1981). In this study however, the focus is on the misattribution of memory which involves the source details retained in memory but to the wrong source (Schacter, 2001). In their study, participants with 'normal' memories regularly made the mistake of thinking they had acquired a trivial fact from a newspaper, when actually, the experimenters had supplied it (Schacter et al., 1984). This type of misattribution is fairly common and it can be tested in human-robot interactions.

Empathy Gap: Empathy gap is a cognitive bias which influences people to misunderstand the power of urges and feelings, such as, pain (Nordgren et al., 2006), hunger (Nordgren et al., 2007), sexual arousal (Ariely and Loewenstain, 2006), fatigue (Nordgren et al., 2009) and cravings (Sayette et al., 2008) on their behaviour. For example, when someone is angry, it is difficult to understand what it is like for one to be happy and vice versa; when someone is blindly in love with another, it is difficult to understand what it is like for one not to be, (as well as to imagine the possibility of not being blindly in love in the future).

In our one experiments, we used this bias to show the emotional differences between robots and the participants where the robots behave overly happy or sad and sometime unresponsive.

Dunning-Kruger effects: The Dunning-Kruger effect is a cognitive bias where relatively unskilled individuals suffer from illusory superiority, mistakenly assessing their ability to be much higher than is accurate. Dunning and Kruger (1999) described this bias effect as *"…incompetent people do not recognize—scratch that, cannot recognize—just how incompetent they are*". Their research also suggested that highly skilled individuals may underestimate their relative competence, they may erroneously assume that tasks which are easy for them are also easy for others and they may incorrectly suppose that their competence in a particular field extends to other fields in which they are less competent.

Self-serving bias effects: The self-serving bias relates to people's attribution for their personal outcomes. People make internal attributions for desired outcomes and external attributions for undesired outcomes (Shepperd, 2008). A classic example of self-serving bias is a student taking an exam. If the student does well on the test, he/she is more likely to believe that his or her own ability and effort (i.e. things under the student's control) were the reasons for success. However, if he/she receives a poor grade on the test, the blame will fall on the external factors such as luck, difficulty of the task, or lack of cooperation of others (Campbell and Sedikides, 1999). The student might claim that the professor made up an unfair test, or the lighting in the room was too dim so the student couldn't focus. In the workplace, the workers who attribute receiving promotions to their own hard work and exceptional skill, but they usually attribute denial of promotions to unfair bosses or other external causes. Athletes sometimes accredit themselves for performing well in the sports arena, but when they perform poorly, they blame external causes (Michele et al., 1998).

Humours effects: Humours effect bias is a cognitive bias of memory. It has been studied that humorous items are more easily remembered than non-humorous. This tendency might be explained by the distinctiveness of the humour, the increased cognitive processing time to understand the humour, or the emotional arousal caused by the humour. The beneficial effect of humour on experienced emotions could be based on the mechanism that humorous processing requires attentional resources so that people are distracted from negative stimuli (Strick et al., 2009). Humour is an integral part of everyday interactions. It is very common, whether people tries to navigate a bookstore, make conversation with the barista at coffee shop, or talk a police officer out of a ticket. Humans inherent their desire to laugh and that motivates various social actions, such as sharing funny YouTube videos, responding to text messages with a LOL and with many iconic faces. People even choose to get their daily news with a large side order of comedy from outlets like 'The Daily Show,' 'The Colbert Report' or "The Onion" (Jasheway, 2016).

These five biases were developed in three robots (ERWIN, MyKeepon and MARC) in four different experiments as such:

Table 1: Biases used in each experiments

Experiment No.	Cognitive biases	Robot used	Types of experiments
1st	Misattribution	ERWIN	Conversational
2nd	Empathy gap	MyKeepon	Game Playing
3rd	Misattribution, Empathy gap and, Dunning-Kruger effects	MARC	Conversational
4th	Humours effects and, Self-serving		Game Playing

This research was carried out over a length of three years of time. The four experiments were conducted individually as well as the biases. In the 1st and 2nd experiments, we tested two biases (misattribution and empathy gap) and, in the 3rd and 4th experiments we tested three new biases (Dunning-Kruger, Self-serving and humours effects) with previous two biases as well. All the biases and non-biases interactions were done individually in all four experiments. The next section, we describe each of the experiments based on their orders.

5. Experiment using ERWIN the expressive robot with Misattribution bias

5.1. ERWIN the expressive robot

ERWIN stands for Emotional Robot With Intelligent Networks. It's a robot head placed on a metal base. The robot is around 40cm tall included the height of the base. The robot is capable of expressing several emotion expressions and can be seen in Figure 2.

ERWIN was programmed using C. Its voice was made using a text to speech software called 'Speakonia'.

The robot can move its jaws and eyebrows which makes five basic emotions expressions. Such expressions are happy, sad, surprise, angry and shock or fear.

In this experiments, the emotions expressions were used as a tool of interacting with the participants. But the main goal was to find out the effects of the misattribution bias and forgetfulness in the robot's interactive behaviours and how that affects in the participant's likeness towards the robot.

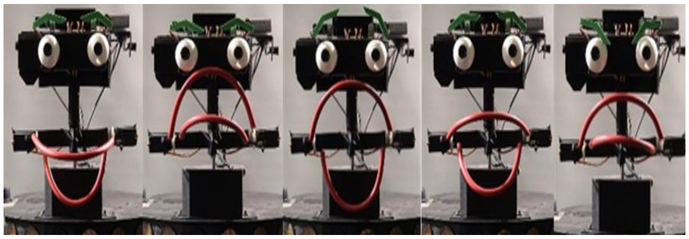


Fig. 2. ERWIN shows different emotions expressions

5.2. Interaction Design

Misattribution bias is associated with the memory. To express the biased effects, the robot needs to show it has forgotten about the previous information. In this case, the experiment needs at least two interactions; the first interaction is to collect the participant's information and the second interaction is for forgetting and misattributing the collected information.

The experiments require the comparison of the misattribution bias effect with another interaction without the misattribution bias effect, to determine the influences of the bias in participant's behaviours towards the robot. Therefore, the interactions were based on two algorithms, such as, misattributed algorithm and non-misattributed algorithm. based on two algorithms, such as, misattributed algorithm and non-misattributed algorithm.

5.3. Experiment Methodology

Each of the participants interacts with the robot three times in the entire experiment. The 1st interaction is the introductory where the robot collects participant's information. The 2nd and 3rd interactions are with misattributed and non-misattributed algorithms. The order of the 2nd and 3rd interactions is random. In general, half of the participants went through the misattributed algorithm as their second interaction, and the other half went through the non-misattributed as their second interaction. The reasons for this is to make the comparison fair between biased and non-biased interactions by not allocating any particular order for the interactions.

The three interactions were done by maintaining a time interval of at least a week to allow long-term affectivity in the participants.

The 1st introductory experiment was common for each participant to allow familiarization with the experimental environment and robot. The experiment was carried out in three steps: the first step was identification, the participants were asked to identify the different facial expressions of ERWIN from pictures to see if they could disambiguate the different expressions without meeting with the robot; the second step was the conversational session with ERWIN, where the robot started friendly conversation, greeting the participant, asking different questions and asking some general questions on various subjects such as hobby, favourite colour, sports and others. The conversations purpose was to allow the collection of basic information on the participants that would be used in the second and third experiments for ERWIN to misattribute. This initial conversation ends with a request from ERWIN to evaluate its performance. The participants were given a brief questionnaire on their experience with ERWIN.

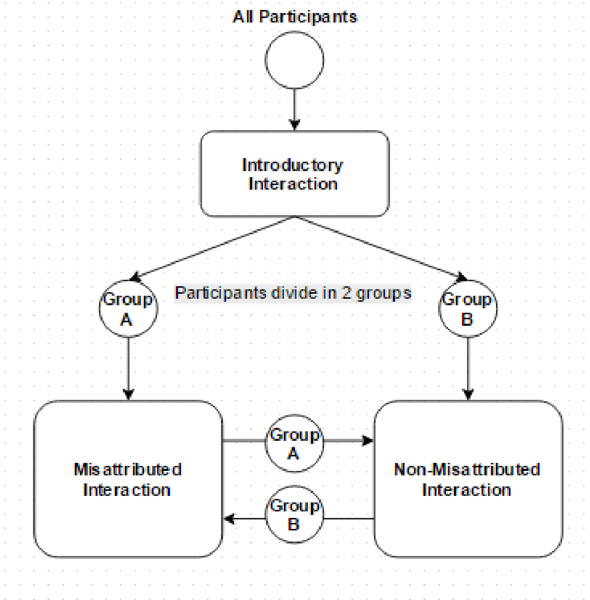

All Participants
Introductory Interaction
Group A
Participants divide in 2 groups
Group B
Misattributed Interaction
Group A
Group B
Non-Misattributed Interaction

Fig. 3. The group division in the experiments

In the 2nd experiment, the participants were categorised into two groups with ERWIN remembering and making general conversations with the first group and misattributing the collected information with the other group's participants. In both cases, the participants were asked to answer questionnaires at the end of the experiment to find out which group of the participants were happier and created satisfactory interrelations with ERWIN.

In the 3rd experiment, the participants from the previous experiment's 'non-misattributed group' experienced misattribute conversations and vice versa. Figure 3 shows the grouping process. At the end, all participants answered the same questionnaire as that given in the 2nd experiment to find out what type of characteristics in ERWIN, participants liked the most. All experiments were Wizard of Oz experiments, where the robot was controlled remotely and participants were watched through ERWIN's eye cameras.

5.4. Single Interaction Structure

The medium of expressing misattribution biased effects for ERWIN is mainly the conversation. This conversational interaction can be divided in three stages, such as, when the parties meet and start conversation, the middle of the conversation where parties discuss various topics and, at the end when the parties need to leave. In our interaction experiments, we make similar stages to develop various moments in conversation. To develop misattribution bias in robot's part of conversation, we divide the interaction in three stages, such as:

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

ii. Topics based conversation – where both parties make conversation in various topics.

iii. Farewell – where interaction ends and the participant has to leave.

Such three stages were developed in both misattributed and non-misattributed algorithms interactions. In the next section, we discussed the details of both misattributions biased and non-biased algorithms.

5.5. Algorithms Design

In this experiment, we needed three algorithms, one for collecting information – which was the introductory interaction, and other two are the misattribution biased and non-misattribution algorithms.

5.5.1. Introductory Algorithm:

This is the very initial stage of the interaction, as the participant's responds to ERWIN's question, it starts to ask about other information which is the next stage of the interaction with expressing corresponding emotions. This introductory algorithm was based on question-answer type conversation. ERWIN did not show any biased effects at all. This stage collects information such as participant's names, address, favourite food, sports and others in different stages of the interactions. For example, in the meet & greet stage robot asks participant's name and address, and topic based conversation stage it collects information about the participant's hobby, favourite things.

5.5.2. Non-misattribution algorithm:

In this interactions, robot s stated participant's previously collected information correctly. For example, in this interaction, robot called participants by their correct names.

Examples of Dialogues are shown below table 2:

Table 2: Examples of dialogues

Examples of dialogues		Differences with misattribution
1	Hello **. It is nice to see you again.	** states the name correctly
2	I correctly remember your name.	
3		*No need of the 3rd dialogue on name, robot moves to the 2nd topic.*

5.5.3. Misattribution algorithm:

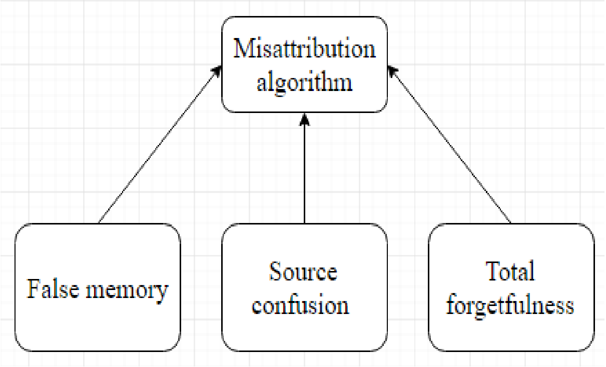


Figure 4. Misattribution algorithm was designed based on three components

The misattribution biased algorithm was designed based on the main characteristic of the bias, which is misattributing previously collected information. To reflect the effects of such biased characteristic in conversation, the algorithm was designed based on three main components of the bias such as false memory, source confusion and total forgetfulness (see figure 4). In the conversation, the robot expresses its misattribution bias effects in dialogues based on such three components.

In this interaction, robot forgot participant’s name and other information.

An examples of the differences in dialogues in the misattribution and non-misattribution algorithms can be shown below table 3:

Table 3: Examples of misattribution dialogues

Examples of dialogues		Misattribution components used
1	Hello Dave. It is nice to see you again.	Forgetfulness
2	I remember that last time you said your name is Dave.	False memory
3	It must be someone else who looks like you.	Source of confusion

5.6. Participants and grouping

Participants were invited by advertising. Total 30 participants were selected. Participants were mixed age groups, and genders. All participants were divided into two groups after the 1st interaction, one group to go for the misattribute interactions in the 2nd interaction and another group to go for the non-misattribute interaction. But in the 3rd interactions, the group reversed, so that the misattribute group from the 2nd interactions did the non-misattribute interaction and vice versa.

The first interaction with ERWIN was an introductory experiment. All participants must go through the introductory interactions. In this interaction, ERWIN collected data from the participants so that it could misattribute in the later interactions in the misattribute group.

5.7. Data collection and Measurements

Data were collected in forms of questionnaires. After each of the interactions, participants were given a set of questionnaires to answers. Such questionnaires are based on Likert method rating based, so that participants can rate their experiences. The rating options were between ‘1’ and ‘10’ where 1 represented ‘least agreeableness’ and 10 represented ‘most agreeableness’. Questionnaires were same for all the interactions. The measurements for this experiments were mainly the likeness of the participants to the algorithms There were total number of 17 Likert questionnaires, from them 4 questions were based on ‘yes/no’. The reliability scores from the 11 questionnaires for the biased and unbiased interactions are 0.94 and 0.756 (Cronbach's Alpha) which are very high. The examples of the questionnaires as such:

1. *Do you feel happy after speaking with ERWIN?*
2. *Would you like to chat with ERWIN again? If yes, then please rate how much.*
3. *How much were you pleased with ERWIN’s response?*
4. *How many times did Erwin make you chuckle? How good was that?*
5. *How happy were you when ERWIN was happy?*

Participants were given a set of ERWIN’s emotions pictures at the beginning of the introductory interactions with corresponded with names of various emotions. Participants had to choose the correct emotions name from the list. At that point, participant never seen ERWIN before, so such recognition could tell us about their skills of recognizing various emotions of ERWIN in the interactions.

The result from collected data analysis are shown in the next chapter. In the next section of the current chapter we discuss the further experiments with more biases and with different robots.

5.8. Experimental Results

At the beginning of the 1st experiment the participants were given a form with five different pictures of ERWIN’s emotions, and they had to identify the correct emotion from six corresponding options of choice. This identification shows participants ability to recognize various emotions. As the participants had never met ERWIN before, so they were actually identifying its emotions on the basis of human emotions knowledge. The pictures on the form showed the emotion expressions happy, sad, shocked, surprise and angry.

After evaluating the collected data for each emotion expressions picture it has found that most of the time majority of the participants 57% of the participants (i.e.8-10) had selected the correct emotion option for the corresponding emotion picture. 21% of the participants (2-3) had minor problems to identify the correct emotions expression and they were confused to differ the emotions between, shocked and surprise, angry and sad. Fig. 3 shows the full results of the identification test.

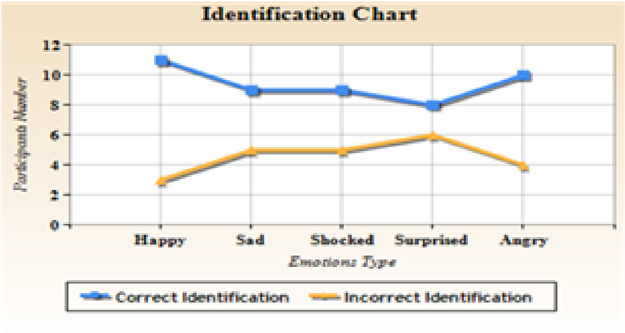


Figure 5: Identification of ERWIN's emotion expressions by the participants engaged in the experiments

To statistically measure of our experimental data, we run a paired sample T-test. To compute the collected data from the both experiments and merging them into graph, we analyzed based on each question and each participant.

Fig. 7, shows the histograms of responses from unbiased and biased interactions. The histogram shows (Fig.22) the average differences between the responses from participants in biased and unbiased interactions. In the questionnaires, the rating options were between 1 and 10, so in this case, the average of 40 points actually suggests that in each questions participants rated average of 4 points higher in biased interactions. Table 2 shows the average Means in both algorithms interactions. We discuss about the results in the next section.

Table 4: Paired Sample Statistics

Algorithm	N	Min.	Max.	Mean	Std. Deviation
Unbiased	30	14.00	71.00	47.67	14.98
Biased	30	61.00	107.00	87.93	10.97

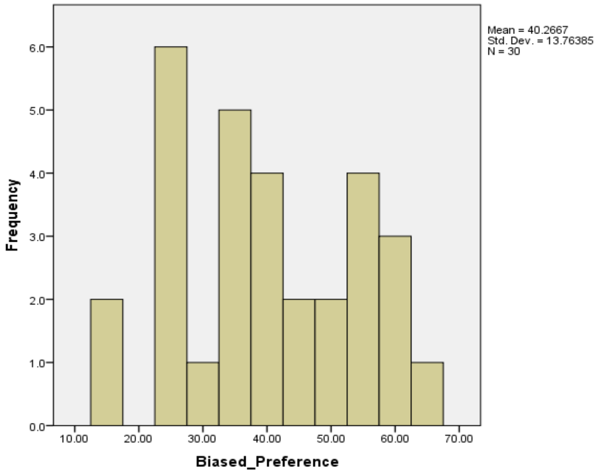
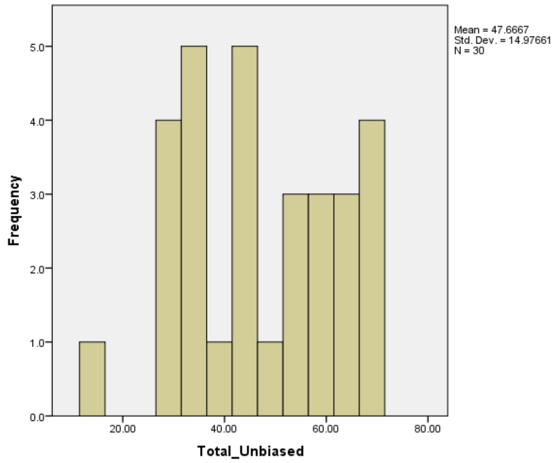


Figure 6: Biased preference from the total biased and total unbiased experiments.



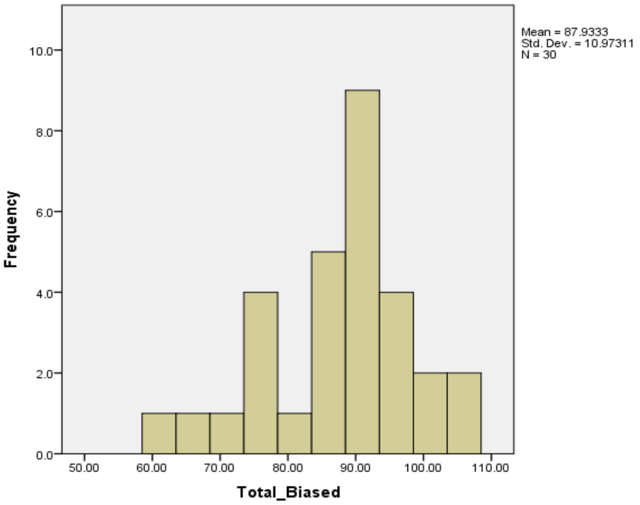


Figure 7: Means from the total unbiased and total biased. Graphs show that total biased Means are higher than the total unbiased.

The Our first set of experiments with the robot ERWIN show that robots with general 'misattributes' bias has got much preferences from participants. Participants enjoyed their first conversation and they expressed their experiences and involvement in the questionnaires feedback. In our case, the high Cronbach's Alpha actually supports to add all the ratings to get the score and compare between biased interaction and non-biased interaction. The histogram graphs (Fig. 6) for biased and unbiased responses are shown. As we can see (Fig.7) and compare the 2 graphs, the Mean for biased data is 87.93 which is approximately 40 point more from the unbiased Mean (47.67) which tells us that for each question participant's responses were average of 40 point (in our ratings 4) higher for biased than unbiased. It is clear in above graphs that the biased responses lied between 60 and 110, whereas the unbiased responses lied between 20 and 70. Now the bellow graph (Fig. 6) shows participant's preferences to biased conversation over unbiased:

The histogram shows (Fig.6) the average differences between the responses from participants in biased and unbiased conversation. The graph shows the number of participants preferred the biased interaction over the unbiased interaction. The mean calculated is 40.27 which tells that there is average of 40 point differences in ratings in prefer to biased interactions. In the questionnaires, the rating options were between 1 and 10, so in this case, the average of 40 points actually suggests that in each questions participants rated average of 4 points higher in biased interactions. From the calculations and graphs, it can be concluded that participants liked the biased robot interactions over the unbiased robot interactions.

The 2-tailed sig (p value) came out as <0.05 which indicate the significance our collected data over large population. From the above t-statistic, t = 16.024 and p < 0.001, 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.

The analysis suggests of participants (t = 16.024, p<0.0001) preferring biased robot interactions over non-biased interactions. In this data set, participants preferred biased ERWIN interactions, on average, by approximately 40.27 points (in our case 4 point). In 95% confident interval, we can see that lower and upper limits are 35.12 and 45.4, which means larger population can prefer the biased interaction by Mean 40 and in a range between 35 and 45 points for each question. Therefore, from the above statistical analyses, we can conclude that, misattribution biases affected our interactions experiments, and overall, participant's significantly liked the interactions using misattribution bias in ERWIN's speech. This experiment result confirms our hypothesis and motivate to examine more biases in robot's interactive behaviours in our future experiments.

6. Experimenting with MyKeepon with Empathy gap bias

6.1. MyKeepon the expressive robot

MyKeepon has two different interactive modes – one is dancing mode and another is touch mode. While it is in the dancing mode, a sensitive microphone in its nose allows it to hear the music been playing and dance to it. My Keepon listens for the tempo of the music and matches the beat with an uncanny sense of timing. My Keepon has his own non-verbal language, which he uses to express himself or try to grab your attention. The touch mode is very sensible for the robot. When in the touch mode, MyKeepon responds any touch, poke and taps on its body and move to the direction of it.

Figure 8: MyKeepon robot

We used an open source software interface called ViKeepon to control the robot remotely. ViKeepon allows to create custom movements and sounds which can be used remotely just by clicking on the buttons. As seen figure 9, each button represents a set of movements to perform a particular task, such as, 'Greetings1' has three movements and two sounds to perform a warm welcome to the participant. ViKeepon supports minimum 15 different sounds including wake up, yawn, sleep, chimp, sneeze and others. With custom moves and sounds we created different interaction moments, such as, greetings, dance movements, sad expressions and many others. Each participant was allowed up to 10 minutes to interact with the MyKeepon.

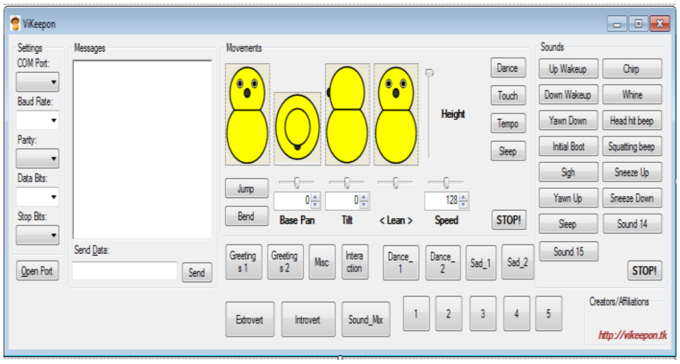


Figure 9: ViKeepon interface

6.2. Experiment methodology

As stated earlier, there were basically two interactions experiments – one is with the basic algorithm and another is with the empathic gap biased algorithm. Each participant must to do both of the interactions.

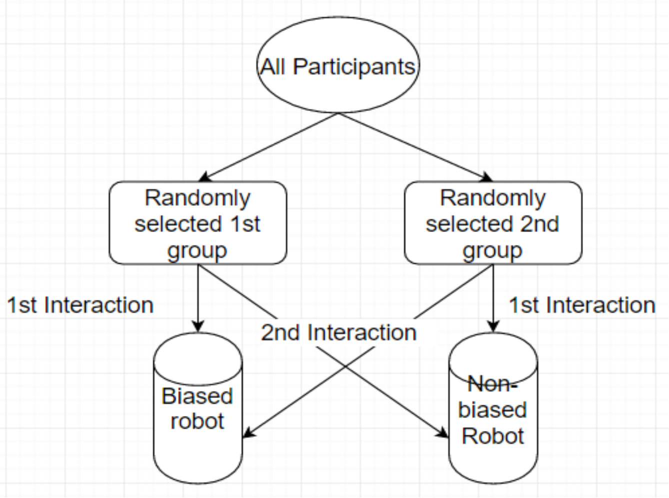


Figure 10: Random selections of interacting with biased and non-biased robot

This experiments were conducted in two interactions. In one interaction, participants interacted with the empathy gap biased version of the robot and in other interaction they interact with the non-biased version of the robot.

As stated earlier, ViKeepon (Figure 9) supports a different ranges of sounds including wake up, yawn, sleep, chimp, sneeze and others. With custom moves and sounds we created different interaction moments, such as, greetings for 2 experiments, different dance movements, sad expressions and many others. Each participant was allowed up to 10 minutes to interact with the MyKeepon. The interaction process included greetings, trying to establish eye contact with participant, responding to participant's actions, showing different movements, showing biased behaviours, showing sad expressions at the end of the interactions. Participants were allowed to touch and tap on the MyKeepon's head, also, they could clap to make Keepon dance. In general, number of times participant claps, the robot jumps same number of times. But, in biased interactions, we introduced empathy gap cognitive bias which allowed robot to become too happy if it jumped correct numbers of claps and too sad if it jumped wrong numbers of claps. In general, for the unbiased interactions the robot jumped correct numbers of the participant's claps and for the biased interactions it jumped wrong numbers of time and jumped fewer or more times. Also, MyKeepon became unresponsive during interactions to see the participant's reactions in biased interactions. These type of different biased behaviours made the interaction different compared to unbiased interaction. In unbiased interaction, MyKeepon did not made mistakes in counting claps, or showing different behaviours. MyKeepon interacting with the participants without being unresponsive and the interaction followed very specific script, like, greeting, make the eye contact, showing different behaviours, jump when claps, be sad when participant leaves.

6.3. Single interaction design

There can be different ways to express the empathic gap in behaviours. But in this experiment, we choose to show such biased behaviours based on the robot's own cognitive abilities. Keepon is known to be extremely emotive robot. MyKeepon's unique dance moves, interactive movements and noises can easily attract people. However, to show the empathic gap in MyKeepon there are only available behaviours are differences in noises and differences in movements. Therefore, in the basic algorithm MyKeepon interacts in general friendly manner, but in the biased algorithm it shows empathic gap in its behaviours. Below we show the different stages of the both basic and biased interactions, and discuss about the differences. In our experiments, participants interact with MyKeepon individually. The interaction has three stages:

i. Meet and greet

ii. Game playing

iii. Farewell

Each participant interacts with MyKeepon two times, the 1st interaction is without Empathy gap bias and the 2nd interaction is biased. In general, in the first stage of the interaction, i.e. meet and greet MyKeepon starts to make an attachment with the participant, in the second stage participant plays a short game and in the third stage participant leaves and MyKeepon becomes sad.

In the next section, we discuss the algorithms design for both biased and non-biased interactions.

6.4. Algorithm Design

The biased algorithm was created based on main principle of the Empathy gap bias, which is unable to understand other's emotional state. The non-biased interaction was created without such biased effects. Both biased and non-biased effects could be expressed by the robot MyKeepon using its own functions and features. As stated in the previous para, in this experiments, the interaction was divided in three theoretical stages, such as, meet and greet, playing game and, farewell. The picture (Fig. 11) show an example of differences between empathy gap biased and non-biased algorithms.

6.5. Participants and Grouping

Total of 30 participants were selected from advertising. Participants were from different backgrounds and random ages and genders.

In this experiments, each of the participants interacted two times with the robot. There was no particular grouping in this experiments. In one interaction, randomly selected 15 participants interacted with MyKeepon without Empathy gap behaviours and, other 15 interacted with the biased version of the robot. In the 2nd interaction, participants who interacted with non-biased robot in the last interaction, now interact with the biased MyKeepon, and others interact with the non-biased MyKeepon. The order of interaction was random as well.

6.6. Data collection and measurements

Participants were requested to complete a set of questionnaires after end of each interaction. The questionnaires were made followed by 'Likert' method rating based and were same for both experiments so that the differences of the participant's likability can be compared in the two interactions. There were total 14 questions in the questionnaires. In addition, participants can leave their own comment about the interaction as well. The questionnaires are based on Likert method rating based, so that participants can rate their experiences. The rating options were between '1' and '10' where 1 represented 'least agreeableness' and 10 represented 'most agreeableness'. Questionnaires were same for all the interactions. The reliability scores from the 14 questionnaires for the biased and unbiased interactions are **0.87** and **0.83** (Cronbach's Alpha) which are high. The measurements for this experiments were mainly the likeness of the participants to the algorithms. In this experiment, the questionnaire was similar to the previous ERWIN experiments.

6.7. Experimental results

Due to the similar type of this experiment with the previous, we run T-test to statistically analyze data. Fig 23.A shows the average ratings from both unbiased and biased interactions. Fig 23.B shows participant's preferences between two interactions. Table 3 represents the average Means from both interactions and standard deviations values.

Table 5: Paired Sample Statistics from MyKeepon experiments

Algorithm	**N**	**Mean**	**Std. Deviation**
Unbiased	28	52.3	4.99
Biased	28	55.37	5.91

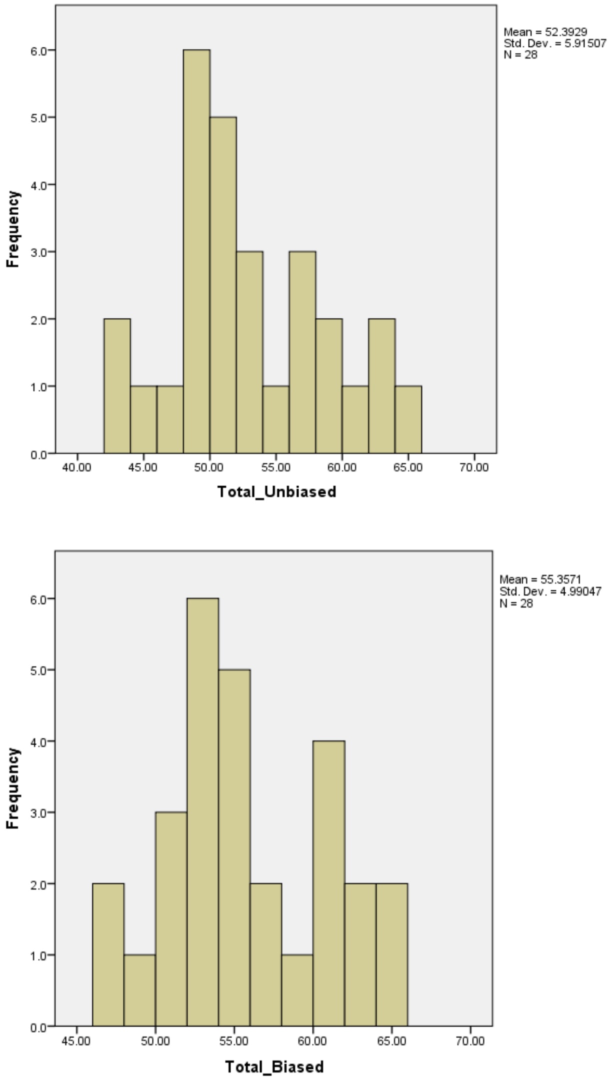


Figure 11: The Means graph of Total Unbiased and total biased

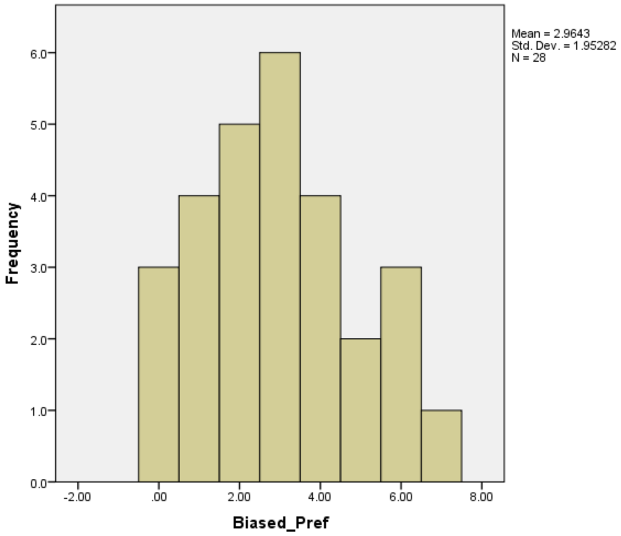


Figure 12: Bias's preference graph in MyKeepon experiment

In the MyKeepon experiments, the T-test are shown in the previous section. The histogram graphs for biased and unbiased responses are shown below. As we can see and compare the 2 graphs (Fig. 11), the Mean for biased data is 55.36 which is approximately 3.0 point more from the unbiased Means (52.39) which tells us that for each question participant's responses were average of 2 ratings higher for biased than unbiased.

From the graph (Fig. 12) we can see that the mean calculated is 2.97 which tells that there is average of 3 (approx.) point differences in ratings in prefer to biased interactions.

The correlation between the two sets of scores is **0.95**. It can be said that the pattern of change is consistent for each participant for each questions.

From the t-statistic, t = 8.032 and p < 0.001, 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. According to the above measurements, there is strong evidence (t = 8.032, p<0.0001) of preferring biased robot interactions over non-biased interactions. In this data set, participants preferred biased Keepon interactions, on average, by approximately 2.97 points. In 95% confident interval, we can see that lower and upper limits are 2.20706 and 3.72151, which means larger population can prefer the biased interaction by Mean 2.97 and in a range between 2 and 3 points for each question. Therefore, from the above statistical analyze, we can conclude that, developed cognitive biases actually affected the interaction between the robot and the participants, and overall, participant's liked the interactions using focusing effects and empathy gap biases in MyKeepon experiments.

7. MARC the Humanoid robot with Misattribution, Empathy gap and Dunning-Kruger affect biases

7.1. The robot MARC

We used a 3D printed humanoid robot MARC for this experiments. The MARC was built inspired by the open project InMoov (2015). 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. MARC can move its hands, arm and body, tilt its head and look around, also it can move jaws while speaking. In the experiments, MARC used common gestures and such gestures were designed from various studies (Wallbott et al, 1986), (Gross M, 2010). MARC's voice was created using text-to-speech software and then edited using Audacity to make it more robotic voice.

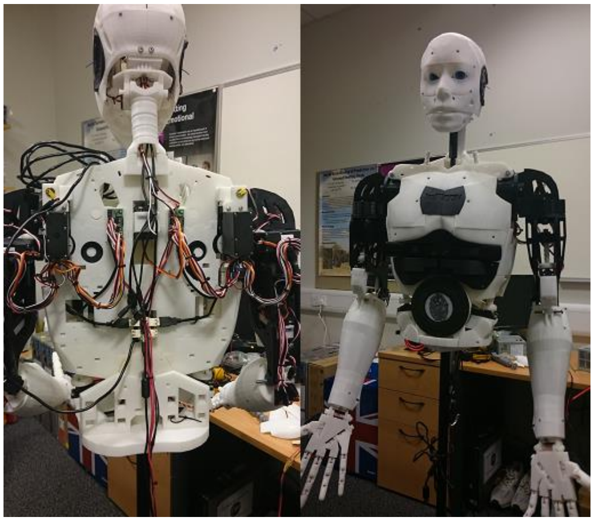


Figure 13: MARC the Humanoid Robot

7.2. Experiment algorithms

This was the 3rd experiment of the series of testing cognitive biases in the robot for human-robot interaction. In this experiment, we used previously tested misattribution and empathy gap biases and, also added a new bias called Dunning-Kruger effects to develop in a humanoid robot MARC. The reason for using previously tested biases was to find out if the positive responses received from the participants in the ERWIN and MyKeepon experiments (Biswas M, 2015) was for the algorithm or the robot itself. In this experiment we developed similar misattribution and similar empathy gap algorithms in MARC, also we try the Dunning-Kruger effects bias alongside. We compared robot's all biased behaviours with baseline behaviours through conversational interactions.

7.3. Methodology

In this experiment, three cognitive biased and a non-biased algorithm was used in the robot for interactions. The selected biases were misattribution, empathy gap and Dunning-Kruger effects. As described in the previous experiments, the non-biased algorithm was a simplistic version which does not show any of the chosen biased behaviours.

The misattribution bias was experiment using similar methodology of ERWIN experiment. Therefore, the 1st interaction was introductory where the robot collect information from participants to misattribute in later interactions. Followed by the introductory interaction there were three times of misattributed interactions maintaining at least a week or two time intervals.

The empathy gap bias tested using similar procedure of previous MyKeepon experiment. In this case, we developed two algorithms for empathy gap bias – hot state of empathy and cold state of empathy. Such algorithms were assigned randomly for the participants in three-time interval experiments.

Dunning-Kruger effects bias was tested in the three-time interval interactions.

Similar way, participants interacted with the baseline or non-biased algorithm three-time interval interactions. At the end, all data from biased interactions have compared with the data from the baseline interactions (Figure 14).

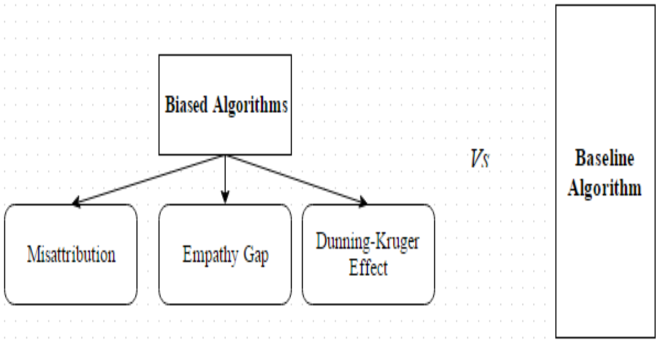


Figure 14: The experiment structure

There were two different methodologies applied for the interactions. experiments therefore, were performed in two separate groups (Figure 5.24.1, 5.24.2) where one group of participants interacted on all four algorithms and the other group interacted with only one algorithm at a time for three times. This was to aid in finding out the participant's reactions on two different occasions, such as participants who interacted all biased and baseline algorithms for three times and participants who interacted with only one algorithm for three times. The

The interactions were on a one to one basis, where each participant interacted with MARC individually for at least 8 to 10 minutes. These three interactions were based on conversations between the robot and the participants. The conversation ended by a request to fill in questionnaires from the robot.

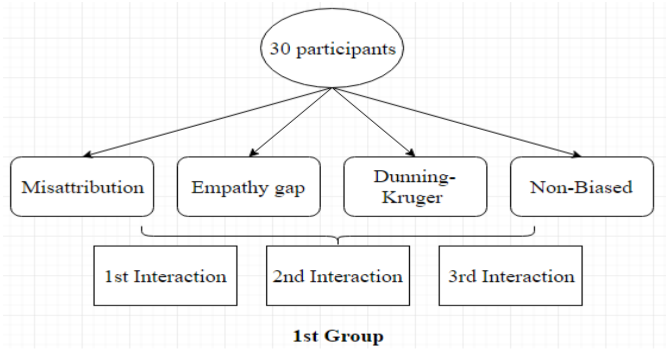


Figure 15.1: The 1st Group: 30 participants interacting with all algorithms

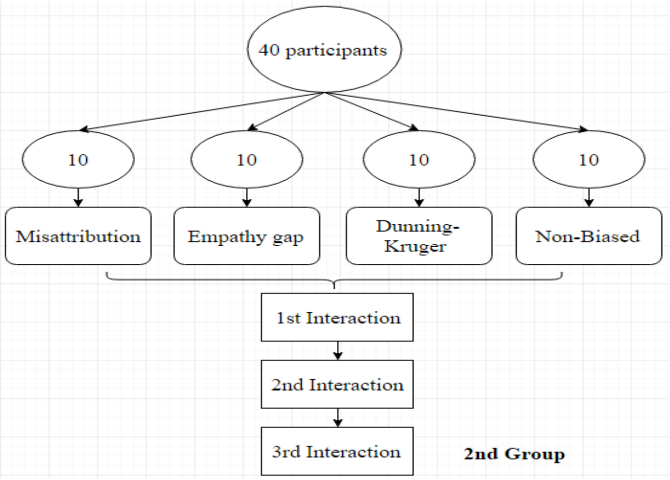


Figure 15.2: The 2nd Group: 40 participants interacting with individual algorithm

7.4. Participants and grouping

A total of 70 participants were randomly chosen from responses to advertisements. The number of different gender races and age groups were maintained equal for both groups.

For the first group (shown in figure15.1), 30 participants were selected where each participant interacted with all four algorithms (three biased and unbiased) of the robot. In the second group (shown in figure 15.2), 10 participants from total 40 were selected to interact with each of the individual algorithms (individual biased or unbiased) throughout the experiments. As with the first group, all 30 participants were interacting with all biased and unbiased algorithms, so their responses would be based on the comparisons between the biased and unbiased interactions. Such responses would reflect a comparable outcome between those who used cognitive bias as well as the unbiased algorithms in developing a long-term interaction. In the second group, each of the 10 participants interacted with their selected individual algorithm three times. Such interactions could tell us the effects of each individual algorithm in developing long-term interactions with the robot.

7.5. Single Interaction Design

All the interactions were designed in three steps, such as, meeting and greeting, topic based conversation and farewell (Figure 16):

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.

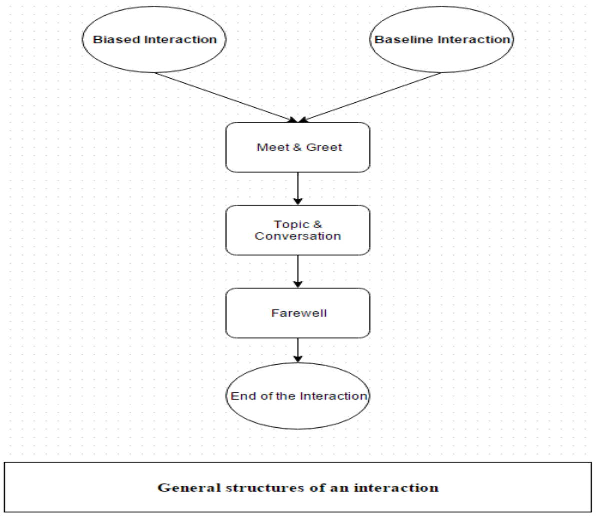


Figure 16: The interactions are divided into three stages

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

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 (Figure 17). 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 17 shows the dialogue structures and general differences between baseline and a biased algorithm.

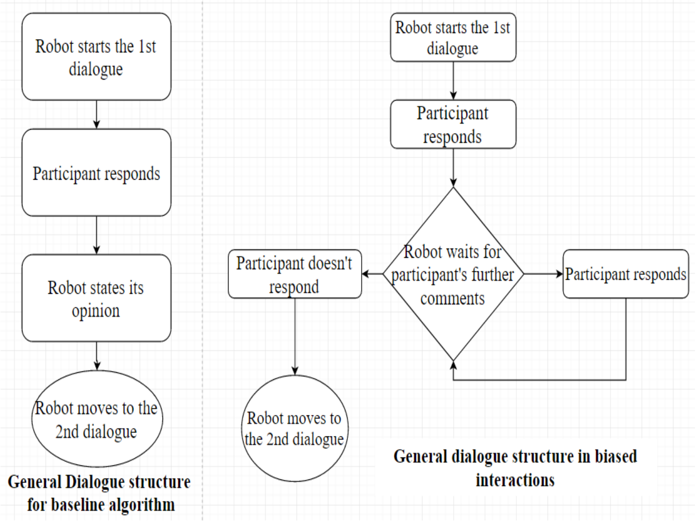

Robot starts the 1st dialogue
Participant responds
Robot states its opinion
Robot moves to the 2nd dialogue
General Dialogue structure for baseline algorithm
Robot starts the 1st dialogue
Participant responds
Robot waits for participant's further comments
Participant doesn't respond
Participant responds
Robot moves to the 2nd dialogue
General dialogue structure in biased interactions

Fig 17. Differences in dialogue structures of baseline and biased interactions.

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.

7.6. Data Collection and measurements

Participants were given a questionnaire after each of the interactions. The questionnaires were in 'Likert' method using a scale of '1' (least agreeableness) to '7' (most agreeableness). Such questionnaires were to find out the participant's likenesses 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. Moreover, 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).

7.7. Algorithm Design

In this experiment, proposed was to have three interactions for each of the selected biases for over a month period of time. The biased interactions were designed based on the components of the selected bias. For example, to study the influence of the misattribution bias, the algorithm was developed based on the three main components of the bias – false memory, source confusion and forgetfulness. Empathy gap biased algorithm was developed based on the 'hot' and 'cold' states of the empathy. Dunning-Kruger bias was developed based on the main three components of the bias, such as the robot failing to recognize its own lack of knowledge, the robot failing to recognize genuine skills or knowledge in others and, the robot recognizing and acknowledging its own lack of skill, after being exposed. Such components of each of the biases in interactions are shown in the figure 18.

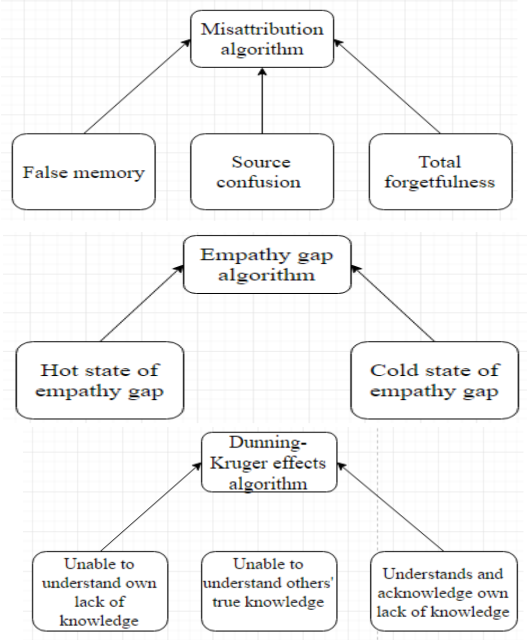

Misattribution algorithm
False memory
Source confusion
Total forgetfulness
Empathy gap algorithm
Hot state of empathy gap
Cold state of empathy gap
Dunning-Kruger effects algorithm
Unable to understand own lack of knowledge
Unable to understand others' true knowledge
Understands and acknowledge own lack of knowledge

Figure 18. Different biased algorithms in three interactions

7.7.1. Designing baseline algorithm

In this experiment, one of the most important tasks was to develop a baseline algorithm which could be compared to all the biased algorithms to get the differences in each interaction. Such a baseline algorithm should not reflect the effects of the chosen biases in the experiment, so that it can be compared. As stated earlier, in this experiment the baseline algorithm was developed to be basic. The dialogues in these interactions was brief. The robot was not supposed to say anything that may reflect any bias. Therefore, this interaction was mainly based on questions-answers type conversation. The dialogue structure for the baseline algorithm is on the figure 17:1st figure. The conversation is supposed to be straightforward for baseline and starts with the robot saying something or asking a question, then the participant's response and ending with another statement by the robot. In between the two dialogues from the robot, the participant can respond only at one time. The second 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.

7.7.2. Designing misattribution algorithm

In this experiment, the misattribution bias was developed similar to the earlier ERWIN experiment. The conversation structure (Figure17: 2nd figure) is different than baseline, as in this case the robot needs to show biased behaviours.

As seen in the above picture (Figure 17: 2nd figure), MARC states its 1st dialogue and wait for responses, then MARC states its opinion and wait for responses. If the participant replied, then MARC state something again otherwise moves to the next dialogue. For example, see table 3.

7.7.3. Designing empathy gap algorithm

As stated earlier, empathy gap happens when a person is unable to understand another person’s emotions properly. For example, if someone is in physical pain and needs rest, it is sometimes difficult for others to understand what that person is feeling. In the experiment with MARC, interactions dialogues are made using two main components of the empathy gap bias, which are:

a. ‘Hot’ state of the empathy gap
b. ‘Cold’ state of the empathy gap

For example, see the table 6

Table 6: Examples of dialogues of empathy gap

Examples of dialogues		Empathy gap components used	Actions
1	Hello! It is great to see you again. I am very happy that you have come to talk with me.	‘Hot’ state of the empathy gap	Wait for response
2	Hi.	‘Cold’ state of the empathy gap	Wait for response

The dialogue structure has shown in the picture 17 in which the participants had an option to reply to the robot’s statement and the robot could respond to that as well.

In the ‘hot’ state of the empathy gap, MARC remains happy and its responses are very cheerful despite the participant’s responses. If the participant asks the reason for that, the robot doesn’t specify any reason, which indicates that being cheerful and over joyous is normal for the robot.

In the cold state of empathy gap, MARC interacts with opposite behaviours to that of the hot state interaction. Through the entire interaction, it usually stays very sad, brief, unhappy and unwilling to talk much. Although the interaction structure is similar with other interactions, which means the robot usually talks about everything the same way as the other interactions but through the entire conversation, the robot does not give much in way of response.

7.7.4. Designing Dunning-Kruger algorithm

As was discussed earlier, the Dunning-Kruger bias was developed based on the three main components of the bias, which are the robot failing to recognize its own lack of knowledge, the robot failing to recognize genuine skill or knowledge in others and, the robot recognizing and acknowledging its own lack of skill, after being exposed. Such components were developed individually in each of the interactions as shown in the figure 18.

In these interactions, the dialogues structure is the same as the previous misattribution biased dialogues (Figure 5.38). In this interaction, the robot always tries to convince the participants that whatever it says is correct and the participant is wrong. If the participant doesn’t argue with the robot, then it moves to the next dialogue.

Table 7: Examples of dialogues of Dunning-Kruger based conversation

Examples of dialogues		Dunning-Kruger effects components used	Actions
1	What type of music is your favourite?		Waits for response
1	No, that is not good. You should listen to **	Unable to understand other’s true knowledge	Wait for response
2	No, you are wrong. I have listen to that and that is not good.	Unable to understand own lack of knowledge	Wait for response and participant reply then move to the 3rd
3.	Okay. May be I am wrong		Move to the next topic

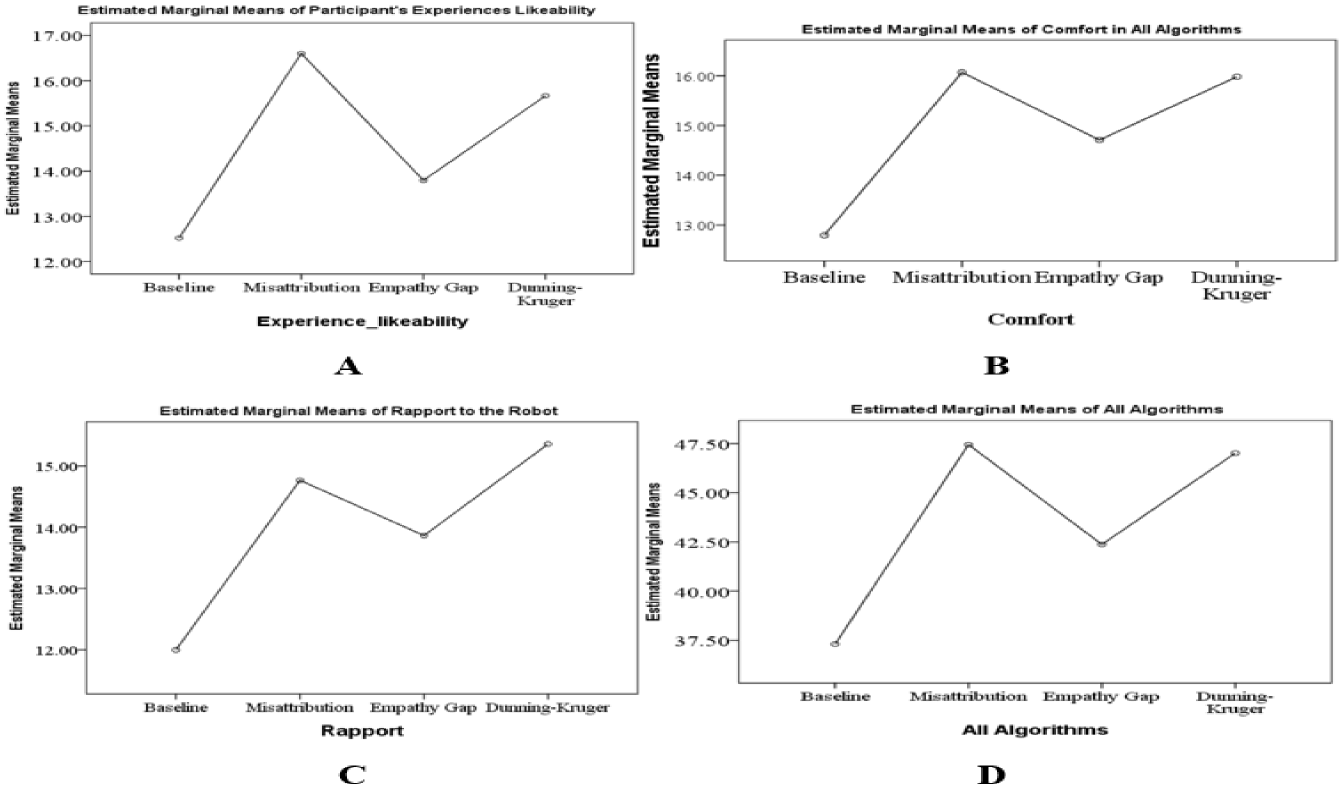
7.8. Experimental results

Data were analysed 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 analyse the data. For the 2nd group, as each of the algorithms has 10 dedicated participants, we ran mixed ANOVA to analyse and compare data. The Cronbach's alpha (α) is calculated 0.916, which indicates high level of internal consistency for our scale.

Graph (Figure 19.) shows the average ratings Means from four different dimensions of questionnaires from the 1st group of participants.

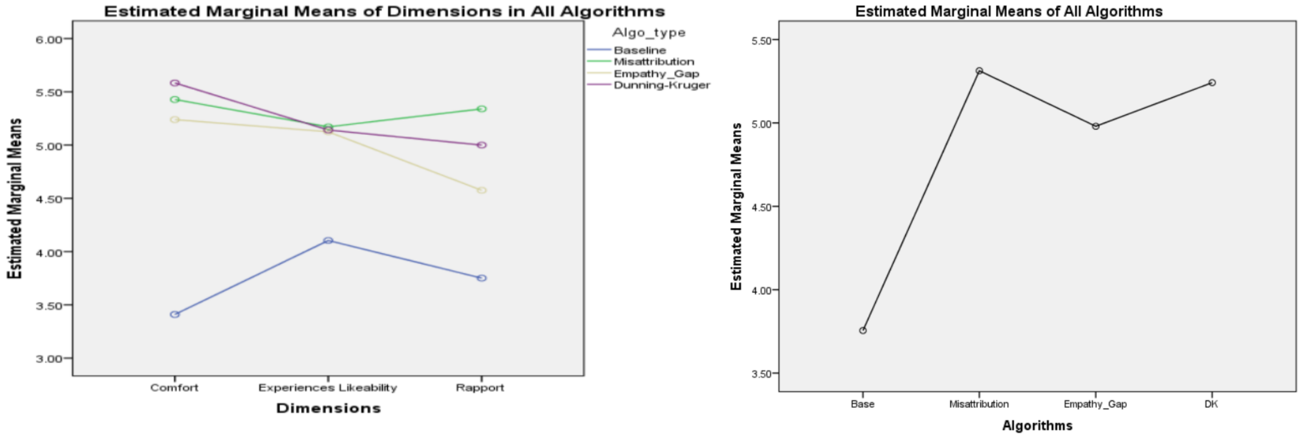
Table: 8: Means from all interactions in 2nd group

Algorithms		Means
Avg. of comfort	Baseline	3.41
	Misattribution	5.42
	Empathy gap	5.23
	Dunning-Kruger	5.58
Avg. of experience likeability	Baseline	4.10
	Misattribution	5.17
	Empathy gap	5.12
	Dunning-Kruger	5.14
Avg. of rapport to the robot	Baseline	3.75
	Misattribution	5.34
	Empathy gap	4.57
	Dunning-Kruger	5.00



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



Left side - The above graph shows Means of the ratings based on 3 dimensions in all algorithms
Right side - The above graph shows Means of the total ratings in all algorithms.

Figure 20. The Mean graphs of the different dimensions and different algorithms for 2nd Group

The 2nd groups interactions results are shown in the graph 20. In the 2nd group data set, the Means from three dimensions were came out as shown in the table (Table 8). The table 8 shows a descriptive analysis of each dimensions. 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 20-left side) shows plotting three dimensions in all algorithms. The graph (Figure 20-right side) 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 20) 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

In this experiments, questionnaire was based on four dimensions. 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. We calculated the total average ratings from all three interactions and compared using repeated measure ANOVA. The results are shown in the graph (Figure 19. A).

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. As previous, we ran repeated measure ANOVA and the result shown in the graph (Figure 19. B).

The rapport 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. Similar to the other dimensions, we calculated the total average ratings from all three interactions and compared using repeated measure ANOVA. The results are shown in the graph (Figure 19. C).

Total Means graph has shown in Figure 19.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 19. 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 19. 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.

8. MARC the humanoid robot with Self-serving and Humours effect bias

8.1. Methodology

As same as the 1st experiment of MARC, the 2nd experiment compares robot's biased algorithms with 'baseline' behaviours. The 'baseline' algorithm was developed without the effects of the self-serving and humours 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 humours effects, robot makes fun of its own winning or losing.

As the self-serving bias motivates an individual to attribute any credit for their success on themselves but any reason for failure on external sources; the most appropriate interaction for the self-serving bias to demonstrate these behaviours is through the application of a game. Humours effects was also can be shown at the various points of the game playing. In the experiments the robot plays the popular paper-rock-scissors game with the participants. This game was used as it is easy an easy to understand game that is played in many countries and familiar to all ages and genders. In addition to the ease of understanding, there are several other factors about this game which makes it appropriate for this cognitive bias. The timing of the game is particularly important and if a player is slower than the other, they can change their move or adapt their move to win by cheating, making it an important feature for the experiments.

8.2. 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.

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.

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 it was important to ensure the robot’s responses didn’t reflect the biased responses in any way. As seen in the diagram (Figure 21), the baseline conversation structure is short and starts with the robot saying something or asking a question, then the participant’s response is given and the robot ends with another statement. In between the 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.

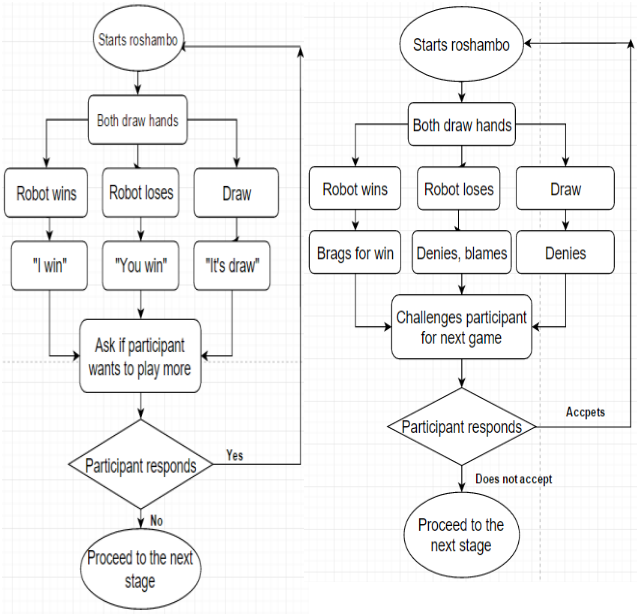


Fig. 21. The game playing algorithm for both algorithms. Left side image represents the baseline, and the right side image represent the self-serving biased algorithm.

The differences between self-serving 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.

8.2.1. Designing self-serving algorithm

For the self-serving bias, if the robot wins a hand, it should express its joy in such a way that it won the hand due to its own intelligence, so that it knew that the participant was going to play that particular option and it’s not just a matter of chance but the robot solved it by its own intelligence. As it is a friendly robot, it gives few tips to the participant for winning the next hand. Giving the tips and related actions expresses its over-confidence and self-serving intelligence and as such, it is going to win all the remaining hands against participants. In this case, the steps are:

i. Ask participant if he/she is ready to play
ii. Draws a hand
iii. Get excited for winning
iv. Brags for winning
v. Gives tips for winning to the participant

vi. Requests participant to play more

For the self-serving bias, losing a hand is not easy for the robot. It tries to find reasons to blame losing on, such as the surroundings and even the participant. Other than just admitting the fact it lost the hand, the robot's actions keep pointing to the excuses and false blames, such as, 'I was not ready' and 'You must have cheated'. As it's a friendly robot, such arguments are limited and mostly ends up with a challenge for winning the next hand. Such interaction steps confirm that despite the robot being the victim of self-serving bias, it still wants to keep playing with the participant. For this biased interaction, the steps would be as follows:

i. Ask participant if he/she is ready
ii. Draws a hand
iii. Shows sad expressions for losing
iv. Gives excuses
v. Blames on various factors
vi. Ultimately blames on the participant
vii. Challenges the participant to play more
viii. If losing continuously, give up the game by blaming others, showing various excuses

8.2.2. Designing humours effects algorithm

For humours effect, the winning is fun for the robot and it expresses it joy in a friendly way. It also gives encouragement to the participant for the next hand. If the participant continuously loses however and doesn't want to play anymore, the robot tells funny stories of encouragement so that the participant keeps playing.

In this case, the steps are:

i. Ask participant if he/she is ready to play
ii. Draws a hand
iii. Gets happy for winning
iv. Gives tips for winning to the participant in funny ways
v. Requests participant to play more

For the humours effect, the robot's speech is supposed to be funny, but it's also important to limit the funny elements, because it's a friendly companion robot.

For humours effect, the losing is not really very bad for the robot and it express its defeat in a friendly way. The reason is, the robot is trying to make the participants remember various moments of the interaction with the help of humour. As such, winning or losing doesn't matter in this interaction. Also, it keeps encouraging the participant to play the next hand but if the participant continuously wins and doesn't want to play anymore, the robot tells funny stories of encouragement so that the participant keeps playing.

The humour effect biased interaction is simple and similar to the baseline interaction, but with humour in conversation. The goal is to find out if the participant likes and prefers such a friendly humorous interaction with a robot in developing long-term interactions

.

8.3. Participants and grouping

Participants were invited for three human-robot interactions by advertisements. 45 participants were selected to interact with any of the individual algorithms (Figure 22). Therefore, for each algorithm there are 15 participants. 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.

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

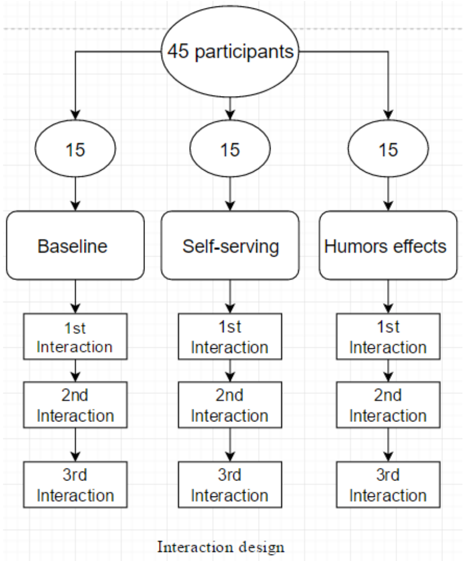


Figure 22: Groups and interaction design

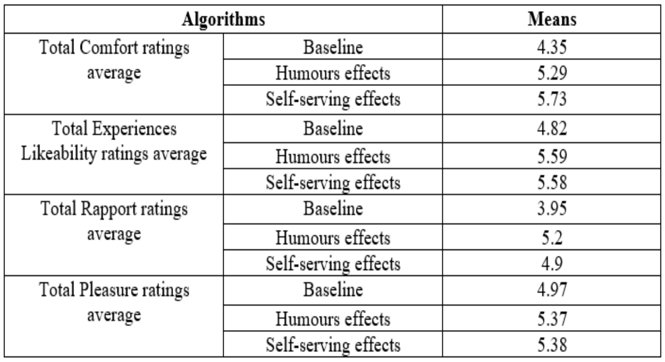
8.4. Data Collection

The goal of the experiments is to investigate the influence of the biased algorithm, therefore, we chose 4 factors to analyze the data, such as, **pleasure** – how pleased participants were for the interaction, **comfort** – how much comfortable participants felt during the interaction, **likeability** – how much they liked the interactions and, **rapport** – how involved they were in the interactions. Such factors should help to understand the influences of the biases. Participants were given a questionnaire after each of the interactions. At the end of the final experiment, we took interview of each participants to know their experiences (*winning, losing games etc.*).

8.5. Experimental results

A mixed (4x2) ANOVA was carried out on the dimensions (4) and algorithms (3).

Figure 23.A shows a descriptive table of each dimensions from all interactions. It shows that the Means of each dimensions are higher for the biased algorithms than the baseline. Among all the chosen factors, the humours effects and 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 23.B shows plotting four dimensions from two algorithms. Figure 23.C shows the Average of Means ratings of the participants in the all 3 interactions. Figure 8 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 23.D, 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.



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

Figure 23.A: Means of four dimensions

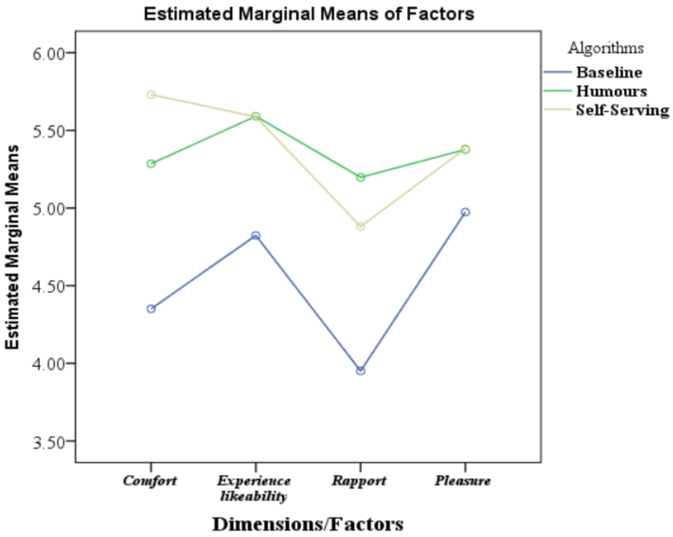


Figure 23.B: The means of 4 dimensions in three algorithms

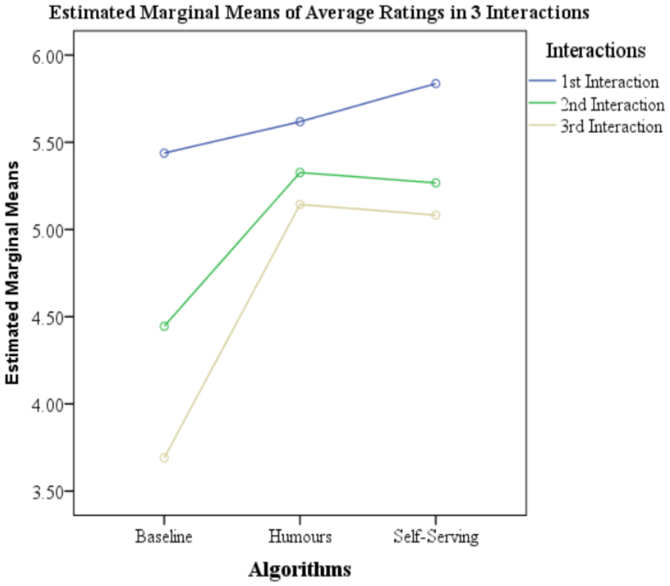


Figure 23.C: Average Means in 3 interactions for all algorithms

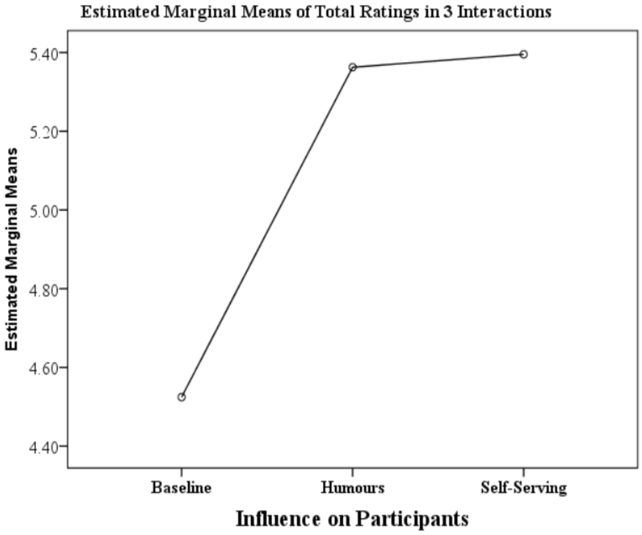


Figure 23.D: Influence on Participant graph – based on the total ratings in all interactions

The overall statistical analysis shows very positive influence of the self-serving and humours effects biases. In the graph (Figure 23.B) we can see that, in all the factors the biased robot scored higher than the baseline. Between the self-serving and baseline the differences of Means of four factors (Figure 23.A) 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 8items 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 23.C, 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 23.D 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.

9. Discussion

The 1st research question can be answered by summing up all the experimental results. The discussions of the previous section 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. Furthermore, as we chose different robots for our experiments, therefore it can be said that cognitive bias can improve the human-robot interaction and that is not affected by the shape, colour or abilities of robots. If we consider other biased interactions in three different robots, in all the cases participants found the biased interactions are much preferable.

The 2nd research question can also be answered from the statistical analysis of all data. Data analysis suggests that participants found it interesting and enjoyable to interact with the biased algorithms. In the previous experiments chapters, we have shown the differences of dialogue design in each algorithms in all experiments. Such differences were very clear in the practical interactions which affected in the participant's ratings. Their ratings suggest that such type of behaviours where the robots were showing humanlike behaviours such as, making mistakes, forgetting information, bragging or blaming were more popular than common 'robot-like' prototype behaviours. Therefore, it can be concluded that humanlike biases make biased behaviours of robot which helps to form of human-robot interaction better than the interactions without common mistakes and biases.

The 3rd research question can be answered by analysing the two experiments of MARC the humanoid robot. From the graphs it is clear that participants bonded with the biased algorithms interactions more the baseline/unbiased interactions. From all the three interactions from two different experiments it can be said that the cognitive biases played an important role to keep the ratings much higher in the biased interactions. The graphs show changes in overall means for different biases and, sometime for the biased algorithms ratings were lowered than previous interactions, but the for the baseline algorithm, the ratings always dropped since the 1st interaction. From the experiments with MARC, it can be said that cognitive bias can make better performance in long-term human robot interaction than interactions without biases.

The main Hypothesis question can be answered by combining all three answers of the research questions from above. By combining all answers of the related research questions above, it can be concluded that,

1. Cognitive biases can influence the human-robot interactions positively despite of the robot's shape, size or colour.
2. Cognitive biases can make the robot's interactive behaviours cognitively imperfect which helps human-robot interactions.
3. Cognitive biases can help to create better long-term human-robot interactions than an interaction without cognitive bias.

All three conclusions point out that the cognitive bias can help and improve the human-robot interaction for long period of time. Therefore, it can be said that developing cognitive biases in the social robot's interactive behaviours can help the robot to interact better than a robot without biases.

10. Lesson learned

From all the experiments and discussions, we learned that humanlike cognitive biases can play an important role in long-term robot human companionship. The statistical analysis suggests that biased algorithms can be used in social robot to enhance their abilities of developing companionship. Therefore, it can be said that social robots should have human-like faults, characteristics biases and prone to carry out common mistakes that humans make on a regular social basis – which will develop the robot's own characteristics and should lead to the acceptance of a robot for long-term relationships with human. The results show that, participants enjoyed and developed a preferred relationship faster with biased and imperfect robots than the robots without the bias, also it shows how one simple cognitive bias can develop a better interaction with participants than the interactions without such bias. Human characters and personality can be described as imperfectly perfect, where robots lack to present such type of cognitive characteristics like unintentional mistakes, wrong assumptions, extreme presence of specific traits, task imperfectness and other human-like cognitive characteristics. In our research, the cognitive biases in robot's behaviours suggest to express cognitive imperfectness, such as, judgemental mistakes, wrong assumptions, expressing tiredness, boredom or overexcitements, or scared of darkness and many other humanlike common characteristics. It is difficult to have a relationship with something that is too superior to us, and pretend to be too perfect without having any mistakes, faults which are unlike humans. We expect, if robot can show similar type of imperfections as humans in their behaviours, then the robots could be accepted to the majority of our society. The research described in this paper shows that cognitive biases and humanlike imperfectness could be the key for long-term robot-human companionship.

11. Conclusion

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 interactions. Also, using of different biases and imperfectness can be helpful to understand the difference in the effects of different biases in human-robot interactions. By comparing data among all experiments, it has 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 interactions can be possible between humans and robots with humanlike imperfect behaviours. 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 interactions. 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.

12. References

1. Baroni, I., Nalin, M., Baxter, P., et al (2014). What a Robotic Companion Could Do for a Diabetic Child. *In 23rd IEEE International Conference on Robot and Human Interactive Communication* 2014 RO-MAN.
2. Bless, H., Fiedler, K., and Strack, F. (2004). Social cognition: How individuals construct social reality. *Hove New York: Psychology Press*
3. Biswas M., and Murray J., Effect of cognitive biases on human-robot interaction: A case study of a robot's misattribution. *Robot and Human Interactive Communication*, 2014 RO-MAN, pp. 1024-1029.
4. Reeves, B., and Nass, C. (2000) Perceptual Bandwidth, *Communication of the ACM*. March 2000/Vol. 43, No. 3.
5. Breazeal., C. (2003) Social Interactions in HRI: The robot View. *IEEE Transactions On Systems, Man, and Cybernetics.* Vol. 34, No. 2.
6. Campbell, W.K., and Sedikides, C. (1999). Self-threat magnifies the self-serving bias: A meta-analytic integration. *Review of General Psychology*, vol. 3.
7. Campbell, W.K., Sedikides, C., Reeder, G.D., Elliot, A.J. (2000). Among friends? An examination of friendship and the self-serving bias. *British Journal of Social Psychology.* Vol. 39, Issue 2: pp. 229–239.
8. Aronson, E., Wilson, T., Akert. R. (2005) A Textbook of Social Psychology, 6th edition. Scarborough, Ontario: Prentice-Hall Canada.
9. Russell, G.F. (1994) Interactive Perceptual Psychology: The Human Psychology That Mirrors The Naturalness Of Human Behaviour. *Mid-Western Educational Research Association*, October 1994.
10. Cornelis, J., and Wynants, M. (2008) Brave New Interfaces: Individual, Social and Economic Impact of the Next Generation Interfaces, ‘Probo, a friend for life?’. ASP Vub Press. pp. 253-257.
11. Park, J.W. et al. (2010) Artificial Emotion Generation Based on Personality, Mood and Emotions for Life-Like Facial Expressions of Robots; *HCIS 2010*, IFIP AICT 332, pp. 223-233, 2010.
12. Kahneman, D. and Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology 3* (3): pp. 430–454.
13. Kanda, T. et al. (2005) Analysis of humanoid appearances in human-robot interaction. *ATR Intelligent Robotics & Commun. Labs*, Japan, 2005.
14. Dautenhahn, K. et al, (2009) KASPAR – A Minimally Expressive Humanoid Robot for HumanRobot Interaction Research. *Applied Bionics and Biomechanics*. 2009.
15. Lee, J.M., Seung-A Jin, W.P., Yan, C. (2006) Can Robots Manifest Personality? Social Responses, and Social Presence in Human–Robot Interaction. *Journal of Communication*. 2006. ISSN 0021-9916.
16. Moshkina, L. et al. (2009) Time-varying affective response for humanoid robots. Progress in Robotics. *Communications in Computer and Information Science.* Vol 44, 2009, pp 1-9.
17. Walters, M.L., Syrdal, D.S., Dautenhahn, K., Boekhorst, R., Koay, K.L. (2008) Avoiding the Uncanny Valley – Robot Appearance, Personality and Consistency of Behavior in an Attention-Seeking Home Scenario for a Robot. *Autonomous Robots Journal*. Vol 24, Issue 2, February 2008. pp 159 – 178.
18. Mahani, M., and Eklundh, K. S. (2009). A survey of the relation of the task assistance of a robot to its social role. *Communication KCSa* Royal Institute of Technology: Stockholm, Sweden.
19. Mandler G, (2002) Origins of the Cognitive (r)evolution. *The Journal of the History of the Behavioural Sciences*. Vol. 38, Pp. 339-353
20. Meerbeek, B., Saerbeck, M., Bartneck, C. (2009) Iterative design process for robots with personality, *AISB2009 Symposium on New Frontiers in Human-Robot Interaction*. 2009, pp. 94-101
21. Myers, D.G. (2015). Exploring Social Psychology. 7th Edition. New York: McGraw Hill Education.
22. Melissa, G.M., Crane, E.A., Fredrickson, B.L. (2010) Methodology for Assessing Bodily Expression of Emotion. *Journal of Nonverbal Behavior*. Volume 34 Issue 4. (July 31): pp. 223–248.
23. Haselton, M. G., Nettle, D., and 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.
24. Tomasello, M. The Human Adaption of Culture, Vol. 28: 509-529, DOI: 10.1146/annurev.anthro.28.1.509
25. Paul Baxter et al. (2011) Long-Term Human-Robot Interaction with Young Users, in Proceedings of the IEEE/ACM HRI-2011 workshop on Robots interacting with children, Lausanne, 2011.
26. McCrae, R.R., John, O.P. (1992) An Introduction to the Five-Factor Model and Its Application. *Journal of Personality*. Vol. 60. pp.175-215.
27. Wallbott, H.G., Klaus, S. (1986) Cues and Channels Emotion Recognition. *Journal of Personality and Social Psychology*: 690
28. Volkmann, K. (2011). Honda's ASIMO visits FIRST robotics event. St. Louis Business Journal. Retrieved 9 August 2011.
29. Shepperd, J., Malone, W., Sweeny, K. (2008). Exploring Causes of the Self-serving Bias. Social and Personality Psychology Compass 2 (2): pp. 895–908.
30. Wilke A. and Mata R. (2012) Cognitive Bias, The Encyclopedia of Human Behavior, vol. 1, pp. 531-535. Academic Press. 2012
31. Tapus, A., Mataric, M.J., Scassellati, B. (2007) Socially assistive robotics. *IEEE Robotics & Automation Magazine*. Vol. 14, Issue 1 2007. pp. 35-42.